

https://dspace.org/

Variation in the LBP-CD14-TLR4-LY96 gene complex and consequences of microbial translocation in HIV-1 infected black South Africans

Mncube, Sizanani

2024
https://hdl.handle.net/10539/37603

Downloaded from a DSpace repository. https://dspace.org/

Variation in the *LBP-CD14-TLR4-LY96* gene complex and consequences of microbial translocation in HIV-1 infected black South Africans

Sizanani Mncube

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, 2023

Supervisors:

Prof. Caroline T. Tiemessen

Dr. Maria Paximadis

Dr. Sharon Shalekoff

DECLARATION

I, **Sizanani Mncube** (Student number: **541681**) am a student registered for the degree of PhD (Virology) in the **academic year 2023.**

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: MNCUBE Date: 19 June 2023

ABSTRACT

Persistent immune activation and inflammation in people living with HIV-1 (PLWH) has been associated with higher morbidity and mortality, even in individuals on antiretroviral therapy (ART). Microbial translocation, among other factors, has been identified as a major driver of persistent immune activation. A subgroup of PLWH collectively known as HIV-1 controllers can naturally control the HIV-1 infection without the use of ART. Little is known about the extent and the role of microbial translocation/immune activation in African HIV-1 controllers. Translocated lipopolysaccharide (LPS), a component of gram-negative bacterial cell walls elicits innate immune responses through the activation of the toll-like receptor 4 (TLR4) in a complex pathway, which requires the use of cluster of differentiation 14 (CD14), LPS binding protein, and Lymphocyte antigen 96 (LY96) also known as Myeloid differentiation factor 2 (MD-2). Although numerous studies have reported associations of expression levels of the LPS recognition and signalling molecules as well as variants in the genes encoding for these molecules, with the risk and severity of various inflammatory, autoimmune, and infectious diseases, such studies are limited in African populations. Given the large genetic diversity in African populations, characterisation of both the constitutive expression levels and genetic variation in these molecules is essential to understanding HIV-1 infection in the populations most affected by the AIDS epidemic.

We quantified constitutive expression of cell surface TLR4 and CD14 (mCD14) on monocytes and neutrophils using flow cytometry and quantified plasma levels of soluble CD14, LBP, and MD-2 using commercially available ELISA kits in two ethnically divergent South African populations [healthy HIV-1 uninfected black (n=17) and white (n=21) individuals]. Furthermore, the influence of sex and age on the expression levels of these molecules was also investigated. We found higher LBP plasma levels in black South Africans compared to white South Africans ($p<0.0001$), however these two populations did not differ significantly in expression levels of CD14 (mCD14 and sCD14), TLR4, or MD-2. Sex differences in the TLR4 expression levels, with higher TLR4 on total monocytes ($p=0.016$) and $CD14^{+}1CD16^{-}$ ($p=0.009$) and $CD14^{+}CD16^{+}$ ($p=0.009$) subsets of monocytes in females compared to males were observed in the white South African population but not in the black South African population. Significant population and sex-specific negative correlations between age and CD14 expression on monocytes, monocyte subsets and neutrophils, and TLR4 expression on neutrophils were observed. In addition, we found that there is differential regulation of TLR4 expression on monocytes and neutrophils between black and white South Africans post

stimulation with lipopolysaccharide (LPS) and lipoteichoic acid (LTA). Together, these findings suggest that population differences in plasma levels of LBP, and population-specific sex differences in TLR4 expression, are likely to differentially impact TLR4 functionality.

Using whole genome sequencing data (WGS), we next sought to fully describe the genetic variation and linkage disequilibrium (LD) patterns in the *LBP*, *CD14*, *TLR4*, and *LY96* genes in HIV-1 uninfected black South Africans (n=87, SA controls), and compared the representation of the variants to select populations from the 1000 Genomes Project. Our results revealed that the representation of genetic variants and LD patterns across these genes in the SA black population more closely mirrored those of representative African subpopulations (Yoruba in Ibadan, Nigeria, and Luhya from Webuye, Kenya) than the European and Asian populations. These findings emphasize that there are vast genetic differences in African populations compared to non-African populations, which could differentially affect gene regulation and associations with various diseases. Several novel variants and putative haplotypes were identified in the SA black population which, upon verification in future studies, will serve to add to understanding the genetic diversity in this population group.

Using WGS data, we also assessed the representation of the *LBP*, *CD14*, *TLR4* and *LY96* gene variants in a cohort of black South African ART-naïve HIV-1 controllers (n=39) comprised of elite controllers (n=21), viraemic controllers (n=6), and high viral load long-term non-progressors (n=12), relative to the SA controls. Only one *CD14* 5' flanking region SNP (rs186291587) showed a significant difference in minor allele frequency (MAF) representation in elite controllers when compared to SA controls (p=0.024; OR=13.3, CI: 1.3 – 131.4). The representation of several *TLR4* variants showed significant differences when HIV-1 controllers were compared to SA controls and the most significant differences were predominantly found in comparison to the HVL LTNPs - the most significant difference observed was overrepresentation of two SNPs in complete LD (r^2=1), a newly identified intronic variant (*TLR4* NI-2), and a 3' flanking region SNP (rs113017335) in HVL LTNPs compared to SA controls (p=0.006; OR=24.71, CI: 2.46-248.51). The representation of several *LBP* variants also differed between HIV-1 controllers and SA controls, here predominantly when viraemic controllers were compared to SA controls. Minor allele frequency overrepresentation of the *LBP* intronic SNP (rs1250247980) in the total group of HIV-1 controllers (p=0.003), and viraemic controllers (p=0.0002), relative to the SA controls, was the most significant difference observed. Furthermore, differences in the representation of *LY96* variants were observed when the total group of HIV-1 controllers, elite controllers and HVL LTNPs were compared to SA

controls - the most significant difference observed was the MAF and heterozygosity overrepresentation of an intronic SNP (rs149605245) in elite controllers compared to SA controls (MAF: p=0.007; heterozygosity: p=0.007). These results suggest a potential role of the LPS recognition and signalling molecules in natural HIV-1 control.

Lastly, in ART-naïve black South African elite controllers (n=44), HVL LTNPs (n=12), progressors (24), and in HIV-1 uninfected controls (HUCs, n=17), we measured and compared plasma levels of select innate immune molecules that are considered markers of microbial translocation and gut damage (LBP, sCD14, REG3α), or are important in interacting with TLR4 (MD-2). We found no differences between groups in plasma levels of LBP and MD-2. However, sCD14 was significantly higher in progressors compared to all groups (HUCs, p=0.0001; ECs, p<0.0001; HVL LTNPs, p=0.0005), with no differences between HIV-1 uninfected controls, elite controllers and HVL LTNPs. Plasma levels of REG3α were unexpectedly significantly lower in progressors compared to elite controllers (p=0.007) and HVL LTNPs (p=0.018), however similar to HIV-1 uninfected controls (p>0.05). Marked sex-specific differences in REG3α levels were evident, with females having significantly higher levels compared to males in all groups (HUCs and ECs, p<0.0001; HVL LTNPs, p=0.036; progressors, p=0.005). Our data suggests that in black South Africans, REG3α plasma levels are not a reliable marker of gut damage, and that increased levels in elite controllers and HVL LTNPs might contribute to protection from excessive systemic activation in the presence of microbial translocation, consistent with reduced monocyte activation in these groups. Progressors, on the other hand, appear to have an inability to produce REG3α while having substantial monocyte activation. Our findings highlight the importance of sex differences, and that studies conducted in populations of different ethnic backgrounds are often not directly comparable.

Overall, findings presented in this thesis contribute to the understanding of the baseline expression levels and the genetic diversity in the *LBP*, *CD14*, *TLR4*, and *LY96* gene complex in the black South African population, and the representation of these variants in black South African HIV-1 controllers. This thesis also highlights the importance of taking ethnicity, sex, and age into consideration when exploring measures that quantify biological parameters. Understanding of the molecules important in the TLR4 signalling pathway can help elucidate approaches that could contribute to curbing immune activation in the context of HIV-1 infection, as well as other diseases.

PUBLICATIONS AND PRESENTATIONS

Publication from this thesis

Mncube S, Paximadis M, Shalekoff S, Reddy S, Kiepiela P, Ebrahim O, Waja Z, Martinson N, Tiemessen CT. Innate immune markers of microbial translocation in HIV-1 Infected black South Africans with extreme clinical phenotypes. (In preparation)

Presentation from this thesis

Mncube S, Paximadis M, Shalekoff S, Reddy S, Kiepiela P, Ebrahim O, Waya Z, Martinson N, 5, Tiemessen CT. Markers of microbial translocation and immune activation in different clinical phenotypes of HIV-1 infection [oral presentation]. Wits Faculty of Health Sciences Biennial Research Day. 15 September 2022, University of the Witwatersrand, Johannesburg, South Africa.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Prof. Caroline T. Tiemessen, Dr. Maria Paximadis and Dr. Sharon Shalekoff for their invaluable supervision, continuous support, and most importantly their patience through this journey. I consider myself extremely lucky to have had supervisors who are so knowledgeable and always provided insightful comments and suggestions.

I would also like to thank Kayley Reveley for the help she provided with analysing the whole genome sequencing data. I am also grateful to the study participants for their invaluable contribution to science.

In addition, I would like to thank everyone in the Cell biology laboratory for creating a positive work environment, encouraging a healthy lifestyle through lunch time walks, the therapeutic picking of pine cones, and for all the support and encouragement. To Gemma Koor, my fellow PhD student, taking this journey with you made me feel like I was not alone, thank you for all the “memes” that kept me going.

Last but not least, I would like to express my sincere gratitude to my family and friends, when I count my blessings, I count all of you 100 times. Thank you for holding my hand and supporting me through this journey.

This work would not be possible without funding received from the National Research Foundation of South Africa and the University of Witwatersrand Postgraduate merit award.

TABLE OF CONTENTS

DECLARATION ..... **ii**

ABSTRACT ..... **iii**

PUBLICATIONS AND PRESENTATIONS ..... **vi**

ACKNOWLEDGEMENTS ..... **vii**

LIST OF FIGURES ..... **xii**

LIST OF TABLES ..... **xiv**

LIST OF ABBREVIATIONS ..... **xv**

CHAPTER 1: Introduction **1**

1.1 Human Immunodeficiency Virus 2

1.1.1 HIV epidemiology 2

1.1.2 Antiretroviral treatment 3

1.1.3 Clinical course of HIV-1 infection 4

1.2 Natural control of HIV-1 6

1.3 Immune activation in HIV-1 infection 9

1.3.1 Causes of immune activation 10

1.4 Microbial translocation 11

1.4.1 Defences against microbial translocation 11

1.4.2 Causes of microbial translocation 12

1.4.3 LPS recognition and signalling 13

1.5 TLR4 signalling receptors and accessory molecules 14

1.5.1 Lipopolysaccharide binding protein (LBP) 14

1.5.2 Cluster of differentiation 14 (CD14) 17

1.5.3 Myeloid differentiation factor 2 (MD-2) 19

1.5.4 Toll-like receptor 4 (TLR4) 22

1.6 HIV and microbial translocation 24

1.6.1 Markers of microbial translocation 24

1.6.2 Evidence of microbial translocation in HIV-1 infection....24
1.6.3 Microbial translocation in ART treated people living with HIV-1....25
1.6.4 Microbial translocation in HIV-1 controllers....26
1.6.5 Intestinal damage in HIV-1 infection....27
1.7 Study rationale....28
1.8 Aims and objectives of the study....29
1.8.1 Aims of the study....29
1.9 Specific objectives....29
CHAPTER 2: Expression levels of LPS recognition and signalling molecules, LBP, CD14, TLR4, and MD-2 in two ethnically divergent South African populations....30
2.1 Introduction....31
2.2 Methods....32
2.2.1 Study participants....32
2.2.2 Quantification of whole blood CD14 and TLR4 expression levels using flow cytometry....33
2.2.3 Quantification of sCD14, LBP, and MD-2 plasma levels....35
2.2.4 Statistical analysis....36
2.3 Results....36
2.3.1 CD14 and TLR4 expression on monocytes and neutrophils....36
2.3.2 Plasma levels of LBP, sCD14, and MD-2....37
2.3.3 Effect of sex on the expression levels of CD14 and TLR4 and plasma levels of LBP, sCD14, and MD-2....39
2.3.4 Effect of age on the expression levels of CD14, TLR4 and plasma levels of LBP, sCD14, and MD-2....42
2.3.5 Correlation between constitutive CD14 and TLR4 expression, and LBP, sCD14, and MD-2 plasma levels....43
2.3.6 Evaluating the effect of LPS and LTA stimulation on TLR4 expression....45
2.4 Discussion....49

CHAPTER 3: Variation in the genes for LPS recognition receptors, CD14 and TLR4, in the black South African population: comparison with other populations and prevalence in HIV-1 controllers 56

3.1 Introduction 57

3.2 Methods 58

3.2.1 Study participants 58

3.2.2 Whole genome sequencing 59

3.2.3 Linkage disequilibrium analysis and Hardy-Weinberg equilibrium 60

3.2.4 Data analysis 60

3.3 Results 61

3.3.1 Genetic characterisation of the *CD14* and *TLR4* genes 61

3.3.2 Linkage disequilibrium and identification of putative haplotypes in the *CD14* and *TLR4* genes 68

3.3.3 Comparisons of the representation of the *CD14* and *TLR4* variants and haplotypes between South African controls and HIV-1 controllers 79

3.4 Discussion 82

CHAPTER 4: Variation in the genes for LPS binding proteins, LBP and MD-2, in the black South African population: comparison with other populations and prevalence in HIV-1 controllers 89

4.1 Introduction 90

4.2 Methods and materials 91

4.2.1 Study participants 91

4.2.2 Whole genome sequencing 91

4.2.3 Linkage disequilibrium analysis and Hardy-Weinberg equilibrium 92

4.2.4 Data analysis 92

4.3 Results 92

4.3.1 Genetic characterisation of the *LBP* and *LY96* genes 92

4.3.2 Linkage disequilibrium and identification of putative haplotypes in the *LBP* and *LY96* genes 98

4.3.3 Comparisons of the representation of the *LBP* and *LY96* variants between South African controls and HIV-1 controllers 103

4.4 Discussion 108

CHAPTER 5: Plasma levels of REG3α and sCD14 distinguish extreme clinical phenotypes of black South African persons living with HIV-1 114

5.1 Introduction 115

5.2 Methods 117

5.2.1 Study participants 117

5.2.2 Quantification of LBP, sCD14, MD-2, and REG3α plasma levels 119

5.2.3 Statistical analysis 119

5.3 Results 119

5.3.1 Characteristics of the study participants 119

5.3.2 Plasma levels of LBP, sCD14, MD-2, and REG3α 120

5.3.3 Sex differences and plasma levels of LBP, sCD14, MD-2, and REG3α 123

5.3.4 Plasma level differences (LBP and sCD14) among elite controllers according to site (JHB, KZN) 127

5.3.5 Correlations between LBP, sCD14, and REG3α plasma levels 127

5.3.6 Correlations between age, CD4+ T cell count, viral load, and LBP, sCD14, and REG3α plasma levels 130

5.4 Discussion 132

CHAPTER 6: Summarising discussion and conclusion 139

REFERENCES 151

APPENDIX 1: Supplementary Tables for Chapter 4 174

APPENDIX 2: Ethical clearance 183

LIST OF FIGURES

Figure 1.1: HIV life cycle. 4
Figure 1.2: Clinical course of HIV-1 infection.. 6
Figure 1.3: A Schematic representation of the TLR4 signalling pathway..... 14
Figure 1.4: A model showing the interaction between lipopolysaccharide (LPS), LPS binding protein (LBP), high-density lipoproteins (HDL) and 16
Figure 2.1: Gating strategy for identifying monocytes and neutrophils 35
Figure 2.2: Comparisons of CD14 and TLR4 density on monocytes and neutrophils from HIV-uninfected black (n=17) and white (n=19) South Africans. 38
Figure 2.3: Comparison of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 plasma levels between healthy black South Africans (n=17) and white South African (n=21). 39
Figure 2.4: Comparisons of (**A**) CD14 and (**B**) TLR4 expression on $CD14^+$ monocytes and $CD16^+$ neutrophils between females and males..... 41
Figure 2.5: Comparisons of plasma levels of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 between females and males among black and white South Africans..... 42
Figure 2.6: Heatmap showing correlations (Spearman's rank) between age and the expression of (**A**) CD14 and (**B**) TLR4 on $CD14^+$ monocytes, ... 44
Figure 2.7: Heatmap showing correlations between expression levels of TLR4, mCD14, sCD14, LBP, and MD-2 in black South Africans, white South Africans and the combined total group of our study participants. 46
Figure 2.8: Comparisons of TLR4 expression between unstimulated, LPS- and LTA-stimulated monocytes and neutrophils in whole blood.. 48
Figure 3.1: Linkage disequilibrium (LD) plot showing linkage between variants across the *CD14* gene in the black South African population.. 70
Figure 3.2: Linkage disequilibrium plot showing linkage between *CD14* variants in the (**A**) Yoruba in Ibadan, Nigeria (YRI, n=108), (**B**) Luhya from Webuye, Kenya (LWK, n=99), (**C**) Europeans (EUR, n=503), from the 1000 Genomes Project..... 72
Figure 3.3: Linkage disequilibrium plot showing linkage between variants across the *TLR4* gene in the black South African population..... 75
Figure 3.4: Linkage disequilibrium plot showing linkage between *TLR4* variants in the Yoruba in Ibadan, Nigeria (YRI, n=108), from the 1000 Genomes Project..... 77
Figure 3.5: Linkage disequilibrium plot showing linkage between *TLR4* variants in the Luhya from Webuye, Kenya (LWK, n=99), ... 78

Figure 3.6: Linkage disequilibrium plot showing linkage between *TLR4* variants in (**A**) Europeans (EUR, n=503), (**B**) East Asians (EAS, n=504), and (**C**) South Asians (SAS, n=489) from the 1000 Genomes Project. 78
Figure 5.1: Comparisons of plasma levels of **(A)** LBP, **(B)** sCD14, **(C)** MD-2, and **(D)** REG3α between HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long-term non-progressors (HVL LTNPs), and progressors.....122
Figure 5.2: MD-2 plasma levels.. 123
Figure 5.3: Comparisons of plasma levels of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 between males (M) and females (F) in HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long term non progressors (HVL LTNPs), and progressors.. 125
Figure 5.4: Comparisons of plasma levels of REG3α **(A)** between males (M) and females (F) in HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long-term non progressors (HVL LTNPs), and progressors.. 126
Figure 5.5: Comparisons of plasma levels of **(A)** LBP, **(B)** sCD14, **(C)** MD-2, and **(D)** REG3α between Johannesburg (JHB) elite controllers (ECs) and KwaZulu-Natal (KZN) ECs.....128
Figure 5.6: Relationships between plasma levels of LBP, sCD14, and REG3α, and with age and CD4+ T cell count.. 129
Figure 5.7: Correlations between **(A)** age and plasma levels of sCD14 in the total group of people living with HIV (PLWH).....131

LIST OF TABLES

Table 1.1: Potential Mechanisms for Virus Control in HIV-1 Controllers.....8

Table 3.1: Minor allele frequencies of *CD14* gene variants in the black South African population and comparison to select populations from the 1000 Genomes Project. 63

Table 3.2: Minor allele frequencies of *TLR4* gene variants in the black South African population and comparison to select populations from the 1000 Genomes Project. 66

Table 3.3: *CD14* 2-SNP haplotypes pairwise LD in the black South African control population and 1000 Genomes Project reference populations. 70

Table 3.4: *TLR4* 2-SNP haplotypes pairwise LD in the black South African control population and 1000 Genomes Project reference populations. 76

Table 3.5: Comparison of allelic and genotypic frequencies of *TLR4* variants between South African controls and HIV-1 controllers. 82

Table 4.1: Minor allele frequencies of *LBP* genetic variants (regulatory and coding regions) in the South African population compared to select populations 96

Table 4.2: Minor allele frequencies of *LY96* genetic variants (regulatory and coding regions) in the South African population compared to select populations Project. 98

Table 4.3: Putative *LBP* gene haplotypes (r^2=1) in the South African population. 101

Table 4.4: Putative *LY96* gene haplotypes (r^2=1) in the South African population.....102

Table 4.5: Comparison of allelic and genotypic frequencies of *LBP* variants between SA controls and HIV-1 controllers. 105

Table 4.6: Comparison of allelic and genotypic frequencies of *LY96* variants between SA controls and HIV-1 controllers. 107

Table 5.1: Characteristics of study participants. 120

Supplementary Tables

Supplementary Table 1: Minor allele frequencies of *LBP* genetic variants (introns) in the South African population compared to select populations ……175

Supplementary Table 2: Minor allele frequencies of *LY96* genetic variants (introns) in the South African population compared to select populations ……...180

LIST OF ABBREVIATIONS

°C	Degrees Celsius
µg	Microgram
µl	Microlitre
AIDS	Acquired immune deficiency syndrome
ART	Antiretroviral therapy (ART
BMI	Body mass index
BPI	Bactericidal/permeability-increasing protein
BV510	Brilliant Violet 510
CCR5	C-C chemokine receptor type 5
CD	Cluster of differentiation
CD14	Cluster of differentiation 14
CI	Confidence intervals
CMV	Cytomegalovirus
DAMPs	Damage-associated molecular pattern molecules
Der p 2	*Dermatophagoides pteronyssinus* group 2
DNA	Deoxyribonucleic acid
EAS	East Asian population
EBV	Epstein-Barr virus
ECs	Elite controllers
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EndoCAb	Endotoxin core antibodies
EUR	European
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
GM	geometric mean
GTEx	The Genotype Tissue Expression project
HAART	Highly active antiretroviral therapy
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HDL	High-density lipoproteins
HF	Haplotype frequency
HICs	HIV-1 controllers
HIV-1	Human Immunodeficiency Virus 1
HIV-2	Human Immunodeficiency Virus 2
HLA-DR	Human Leukocyte Antigen - DR isotype
HMGB1	High-mobility group box 1
HUCs	HIV-uninfected controls
HVL LTNPs	High viral load long-term non-progressors
i.e	That is
IBD	Inflammatory bowel diseases
IEC	Intestinal epithelial cells
I-FABP	Intestinal fatty acid binding protein
IFN-α	Interferon-α

Ig	immunoglobulin
IL	Interleukin
IQR	interquartile range
IRF3	Interferon regulatory factor-3
JHB	Johannesburg
kb	Kilobase
kDa	Kilo Daltons
KZN	KwaZulu-Natal
LBP	Lipopolysaccharide binding protein
LD	linkage disequilibrium
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
LWK	Luhya from Webuye, Kenya
LY96	Lymphocyte antigen 96
MAF	Minor allele frequency
mCD14	Membrane cluster of differentiation 14
MD-2	Myeloid differentiation factor 2
ml	Millilitre
mRNA	messenger RNA
MyD88	Myeloid differentiation factor 88
ND	Not done
NF-kβ	Nuclear factor-kappa beta
ng	Nanogram
NGS	next generation sequencing
NI	Newly identified
OD	Optical density
OR	Odds ratio
PAMPs	Pathogen-associated molecular patterns
PBMCs	Peripheral blood mononuclear cells (PBMCs)
PDCs	Plasmacytoid dendritic cells
PE	Phycoerythrin
pg	picogram
PLWH	People living with HIV-1
REG3α	Regenerating islet-derived 3 α
RNA	Ribonucleic acid
Rrna	Ribosomal RNA
SA	South Africa
SAS	South Asian population
Scd14	Soluble cluster of differentiation 14
SIFT	Sorting Intolerant from Tolerant
SIV	simian immunodeficiency virus
SNP	Single nucleotide polymorphism
Th17	T helper 17
TIRAP	TIR domain-containing adaptor protein
TLR	Toll-like receptor

TLR4	Toll-like receptor4
TNF-α	Tumour necrosis factor-α
TRAM	TRIF-related adaptor molecule
TRIF	TIR-domain containing adapter inducing IFN-β
UTR	Untranslated region
VCs	Viraemic controllers
VEP	Variant Effect Predictor
VL	Viral load
WGS	Whole genome sequencing
WHO	World Health Organisation
YRI	Yoruba in Ibadan, Nigeria

Amino acid abbreviations

Amino Acid	**Three-letter code**
Alanine	Ala
Arginine	Arg
Asparagine	Asn
Aspartate	Asp
Aspartic acid	Asx
Cysteine	Cys
Glutamate	Glu
Glutamine	Gln
Glutamic acid	Glx
Glycine	Gly
Histidine	His
Isoleucine	Ile
Leucine	Leu
Lysine	Lys
Methionine	Met
Phenylalanine	Phe
Proline	Pro
Serine	Ser
Threonine	Thr
Tryptophan	Trp
Tyrosine	Tyr
Valine	Val

CHAPTER 1

Introduction

1.1 Human Immunodeficiency Virus

Human Immunodeficiency Virus (HIV), a positive-sense single stranded RNA Lentivirus belonging to the *Retroviridae* family, causes Acquired Immune Deficiency Syndrome (AIDS). AIDS was first described over 40 years ago in homosexual males presenting with Kaposi sarcoma and other opportunistic infections such as *Pneumocystis carinii* pneumonia (Durack, 1981, Gottlieb et al., 1981).

Two types of HIV have been described, HIV-1 and HIV-2. HIV-1 originates from simian immunodeficiency virus (SIV) of the Central African chimpanzee (Gao et al., 1999), whereas HIV-2 originates from the SIV of the West African sooty mangabey (Santiago et al., 2005). HIV-1 and HIV-2 are morphologically indistinguishable and are differentiated based on their genetic characteristics and viral antigen differences (Nyamweya et al., 2013). Compared to HIV-2, HIV-1 has higher transmissibility and is more pathogenic (Nyamweya et al., 2013). HIV-1 is divided into four different groups, namely groups M (Major), O (Outlier), N (non-M, non-O) and P (Putative). The HIV-1 group M is the major circulating group, causing the global HIV/AIDS epidemic. It is divided into nine subtypes, subtypes A to D, F to H, J and K (Taylor et al., 2008). HIV-2 is comprised of subtypes A to H, with subtypes A and B being the most prevalent, and causing the HIV-2 pandemic (Santiago et al., 2005). HIV-1 and HIV-2 subtypes exhibit differences in their distribution in different geographical locations. HIV-2 is predominantly found in West African populations (Visseaux et al., 2016), whereas HIV-1 accounts for 95% of the global HIV infections.

1.1.1 HIV epidemiology

Globally, approximately 84 million people have been infected with HIV since the beginning of the epidemic and over 40 million people have died from HIV/AIDS related illnesses (WHO, 2022). In 2021, 650 000 people died from HIV/AIDS related illnesses, compared to the 1.7 million and 1.2 million people who died in 2005 and 2010, respectively (UNAIDS, 2022). The World Health Organisation (WHO) estimates that over 38.4 million people are currently living with HIV worldwide (WHO, 2022). An estimate of 1.5 million new infections were reported in 2021, indicating a 54% decline in new infections since 1996 (3.2 million), and a 32% decline since 2010 (2.2 million) (UNAIDS, 2022). The sub-Saharan African region is most affected by the HIV epidemic, accounting for over 67% of the HIV global infections, with 25.6 million people living with HIV (UNAIDS, 2022). South Africa has the highest number of people living with HIV worldwide, currently estimated at 7.5 million people. However, Eswatini (27.9%),

Lesotho (20.9%), and Botswana (18.6%) have a higher HIV prevalence in adults compared to South Africa (18.3%) (UNAIDS, 2022).

1.1.2 Antiretroviral treatment

There is currently no cure for HIV/AIDS. Antiretroviral therapy (ART) is a lifelong treatment to manage the disease. Highly active antiretroviral therapy (HAART) is an ART regimen which involves the use of three or more antiretrovirals (ARVs), which target different stages of the HIV-1 replication cycle. Currently, there are over 25 ARV drugs available, belonging to six different classes targeting different stages of the HIV-1 replication cycle (Eggleton and Nagalli, 2020) as shown in **Figure 1.1**. The six classes of ARV drugs are chemokine receptor antagonists (CCR5 antagonists), fusion inhibitors, nucleoside/nucleotide reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, integrase inhibitors, and protease inhibitors (Arts and Hazuda, 2012). ART suppresses HIV-1 replication thereby reducing HIV viral load, attenuating disease progression and transmission, and improving immune function (Eggleton and Nagalli, 2020).

Worldwide access to ART has increased from 2 million people in 2005 to approximately 28.7 million people in 2021. This corresponds to an increase in treatment coverage from 7% to 75% (UNAIDS, 2022). The increased ART coverage has successfully reduced morbidity and mortality in people living with HIV-1 (Montaner et al., 2014). The use of ART has reduced HIV/AIDS related mortalities by 68% since 2004 (UNAIDS, 2022). South Africa has the biggest ART programme in the world with over 5.5 million people currently on treatment (UNAIDS, 2022).

Although ART is lifesaving, it has several limitations. HIV drug resistance is common and occurs due to several factors including the high mutation rate during HIV replication, ART defaulting, poor adherence, and incomplete viral suppression (Clavel and Hance, 2004). The use of ART is also associated with high costs, severe side effects in some individuals (Rudorf and Krikorian, 2005), rapid rebound of viral replication upon treatment interruption (Davey et al., 1999) and most importantly, incomplete restoration of the immune system (Valdez et al., 2002). In addition, although ART has reduced HIV/AIDS related mortalities, the mortality rate is still higher in people living with HIV-1 on ART than the general population (Sandler and Douek, 2012, de Coninck et al., 2018). The higher mortality rate has been attributed to, among other things, inflammation and immune activation, which persists even in individuals on ART with undetectable viral loads (Lichtfuss et al., 2011). People living with HIV-1 on ART are

also at higher risk of developing non-AIDS defining illnesses such as cardiovascular disease, kidney disease, cancer and liver disease (Deeks et al., 2013). This illustrates the need for continued research to identify new strategies for an HIV vaccine and cure.

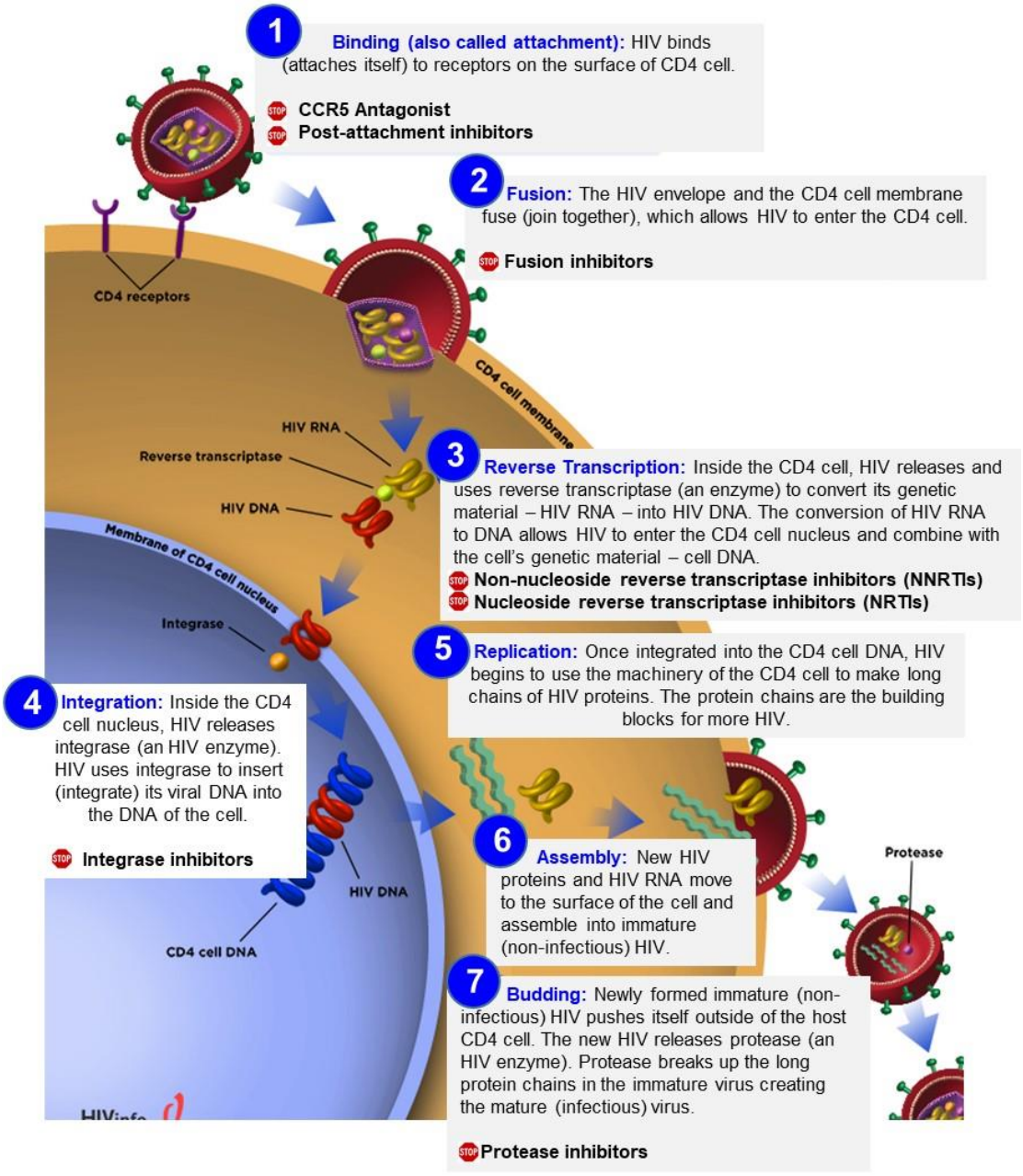


Figure 1.1: HIV life cycle. Reproduced from (https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/73/the-hiv-life-cycle).

1.1.3 Clinical course of HIV-1 infection

The clinical course of HIV-1 infection is generally divided into 3 stages, primary infection (acute phase), clinical latency (chronic or asymptomatic phase) and AIDS (**Figure 1.2**).

Primary HIV-1 infection is from initial HIV-1 entry to the completion of seroconversion and is characterized by the development of HIV-1 specific antibodies (Kassutto and Rosenberg, 2004). Some patients with primary HIV-1 infection experience acute retroviral syndrome with flu-like symptoms occurring 2 to 6 weeks post HIV-1 exposure (Schacker et al., 1996). The acute retroviral syndrome usually coincides with high levels of viraemia and the initiation of the immune responses in the form of HIV-1 specific antibodies and CD8+ cytotoxic T lymphocytes and CD4+ T helper HIV-specific responses (Kassutto and Rosenberg, 2004, Coffin and Swanstrom, 2013). There is a transient decline in CD4+ T cells in peripheral blood during the primary HIV-1 infection as a result of high viral loads (Weber, 2001). The partial control of viraemia by the immune system, and the exhaustion of target cells, both result in a decline in the level of viraemia at the end of the primary HIV-1 infection (Coffin and Swanstrom, 2013). Although this results in CD4+ T cell recovery, the CD4+ T cell levels are not restored to their pre-infection baseline levels (Weber, 2001).

The clinical latency or chronic stage is an asymptomatic phase characterised by a steady increase in viraemia and CD4+ T cell decline (Coffin and Swanstrom, 2013). The viral load set point is the average viral load achieved by an individual in the period directly after primary infection and is an early predictor of the rate of disease progression and mortality (Mellors et al., 1996, Lavreys et al., 2006). The set point is established at the end of primary infection and the time to reach the set point after infection is variable between individuals and population groups (Huang et al., 2012). The balance between the virus turnover and immune responses maintains the viral load set point in the absence of ART (McMichael et al., 2010). Early initiation of ART during primary infection has been associated with a lower immune activation set point (Jain et al., 2013), and enhanced recovery of the CD4+ T cell count (Le et al., 2013). The progressive depletion of CD4+ T cells to less than 200 cells/µl marks the onset of AIDS (Coffin and Swanstrom, 2013). This stage of HIV-1 infection is characterised by susceptibility to opportunistic infections such as pulmonary tuberculosis, pneumococcal pneumonia, cryptococcal meningitis, oropharyngeal candidiasis and Kaposi's sarcoma (Jung and Paauw, 1998).

In the absence of ART, the clinical course of HIV-1 infection is highly variable. The mechanisms underlying the variability are complex and multifactorial. In untreated individuals, different HIV-1 disease phenotypes have been described. These include typical progressors, rapid progressors and slow progressors (also broadly referred to as HIV-1 controllers). Typical progressors, comprising the majority of people living with HIV-1 (70-80%), progress to AIDS

within 6 to 8 years (Pantaleo and Fauci, 1996); rapid progressors (10-15%) progress to AIDS within 2 to 3 years (Pantaleo and Fauci, 1996); whereas slow progressors (5%) can remain healthy for more than 10 years.

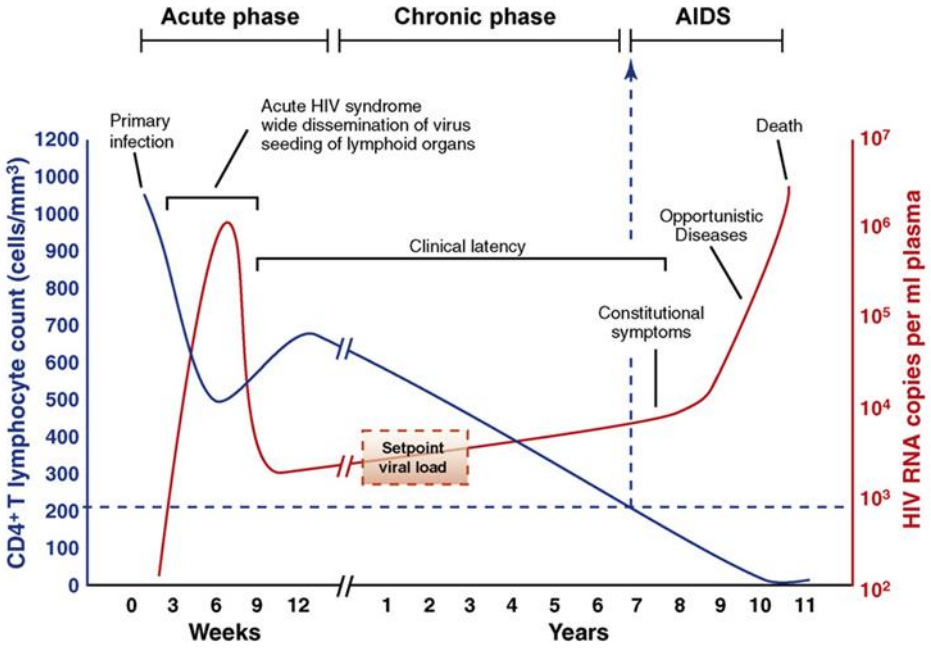


Figure 1.2: Clinical course of HIV-1 infection. Image source (O'Brien and Hendrickson, 2013).

1.2 Natural control of HIV-1

In the absence of ART, a small group (5-15%) of people living with HIV-1 can naturally control the HIV-1 infection. These individuals are collectively known as HIV-1 controllers (HICs). HIV-1 controllers are a heterogeneous group of individuals exhibiting different levels of HIV-1 control and hence they are divided into different groups: elite controllers (ECs), viraemic controllers (VCs) and high viral load long-term non-progressors (HVL LTNPs). There is, however, variation in the nomenclature and definitions of these groups between different studies (Gurdasani et al., 2014).

Elite controllers, making up less than 1% of people living with HIV-1, are characterised by their ability to maintain undetectable viral loads (<50 RNA copies per ml of plasma) and normal CD4+ T cell counts (>500 cells/µl) (Okulicz and Lambotte, 2011). Viraemic controllers maintain viral loads greater than 50 RNA copies but less than 2000 RNA copies and normal

CD4+ T cell counts (Gurdasani et al., 2014). This group of HIV-1 controllers is more common than elite controllers, with a prevalence of 3.3%.

Unlike elite controllers and viraemic controllers who maintain undetectable/low viral loads, HVL LTNPs are people living with HIV-1 who have the ability to maintain normal CD4+ T cell counts in the presence of high viral loads (>10 000 RNA copies per ml) without ART (Hunt, 2009) for an extended period (typically >7 years). This extreme HIV-1 phenotype is rare in adults living with HIV-1 (Rotger et al., 2011). In ART naïve children, paediatric non-progressors account for 5-10% of children living with HIV-1. Paediatric non-progressors resemble HVL LTNPs in that they maintain high CD4+ T cell counts in the presence of high viral loads (Adland et al., 2015, Muenchhoff et al., 2016). HVL LTNPs and paediatric non-progressors are characteristically similar to SIV infected sooty mangabeys, which are natural SIV hosts. SIV infected sooty mangabeys maintain high CD4+ T cell counts and do not develop simian AIDS despite high levels of viral replication (Silvestri et al., 2003).

Mechanisms underlying the natural control of HIV-1 infection are not clearly understood, and they include both viral and host factors. **Table 1.1** shows some of the mechanisms involved in natural control of HIV-1 infection (Deeks and Walker, 2007). Some HIV-1 controllers are not in possession of known protective mechanisms, suggesting that additional mechanisms are still to be identified.

Over time, a proportion of HIV-1 controllers experience virological, immunological and clinical progression due to factors not clearly understood (Groves et al., 2012, Leon et al., 2016). A number of studies have implicated immune activation as a driving force in disease progression. In some elite controllers, with high levels of immune activation, a decline in CD4+ T cell counts has been observed, despite undetectable viral loads. (Andrade et al., 2008). Higher levels of chronic immune activation and low-grade inflammation have been observed in elite controllers compared to virally suppressed individuals on ART (Hunt et al., 2008, Krishnan et al., 2014). HIV-1 controllers who experience disease progression have higher levels of immune activation and inflammation before progressing (Noel et al., 2015). Additionally, Pernas et al. (2018) found that elite controllers with low levels of Gag-specific T cell responses, high pro-inflammatory cytokines, and high viral diversity, were more likely to experience loss of virologic control. Elite controllers have also been shown to be at higher risk for non-AIDS defining illnesses (Pereyra et al., 2012).

Table 1.1: Potential Mechanisms for Virus Control in HIV-1 Controllers [Reproduced from (Deeks and Walker, 2007)].

Mechanism	Evidence
Adaptive Immune Responses	
HIV-specific CD8+ T cells	Controllers are enriched in certain class I HLA alleles and often have CD8+ T cells that produce multiple cytokines and/or proliferate in response to HIV peptides
HIV-specific CD4+ T cells	Controllers often have CD4+ T cells that express high amounts of HIV specific IL-2 and interferon-γ in response to HIV peptides
Innate Immune Responses	
Natural killer cells	Controllers are strongly enriched for certain NK cell receptors that are involved in regulating the function of these cells
Plasmacytoid dendritic cells (PDCs)	The number and/or function of PDCs are high in some controllers
Reduced cellular entry	Controllers are enriched for genetic polymorphisms in the CCR5 pathway associated with reduced amounts of this viral co-receptor and/or have high amounts of CCR5 ligands that may compete with HIV for cell entry
Immunoregulation	
T regulatory cells	T regulatory cells—which blunt antigen-specific T cell responses—are low in some HIV controllers, particularly in lymphoid tissues
Attenuated viruses	Some controllers harbour viruses containing mutations and/or deletions in key regulatory or accessory genes, particularly *nef*
	Mutations induced by $CD8^+$ T cell pressure leading to reduced viral fitness

Some elite controllers experience viral load blips which are defined as temporary, low level increases in viral load to detectable levels followed by suppression to undetectable levels without any intervention or change in ART regimen (Farmer et al., 2016). The use of ART in elite controllers is still controversial. A number of studies have shown that treating elite controllers with ART results in increased CD4+ T cell counts, but the gains in CD4+ T cells are lower than that the gains in non-controllers on ART (Okulicz et al., 2010, Boufassa et al.,

2014). Importantly, treating elite controllers with ART has been shown to result in decreased immune activation (Sedaghat et al., 2009).

A subgroup of elite controllers are considered exceptional elite controllers. Exceptional elite controllers achieve a high degree of viral control without disease progression for more than 25 years (Casado et al., 2020). These individuals have been shown to have distinct proviral reservoirs (Jiang et al., 2020), low levels of proviral DNA (Mendoza et al., 2012), and low levels of T cell activation than elite controllers who experience viral load blips (Canouï et al., 2017). Jiang et al. (2020) also described an exceptional elite controller with no detectable intact proviral sequences, suggesting that this elite controllers is cured.

1.3 Immune activation in HIV-1 infection

Chronic immune activation, inflammation and progressive depletion of CD4+ T cells from blood, mucosal tissues and lymphoid organs are hallmarks of HIV-1 infection (Marchetti et al., 2013). In SIV infected sooty mangabeys, non-human primates who do not progress to simian AIDS, chronic immune activation is absent despite high levels of SIV replication (Silvestri et al., 2003). This illustrates the importance of immune activation in progression to AIDS and suggests that viral replication is not solely responsible for the progressive depletion of CD4+ T cells that results in AIDS (Silvestri et al., 2003).

HIV-1 infection activates both the innate and adaptive arms of the immune system. Although CD4+ T cells are the lymphocytes primarily affected by the HIV-1 infection, other leukocyte subsets are also affected (Klatt et al., 2013). Chronic HIV-1 infection is associated with high levels CD4+ and CD8+ T cell activation and turnover (Hellerstein et al., 1999, Hazenberg et al., 2000). Although ART reduces levels of T cell activation, in virally suppressed people living with HIV-1, the levels of activation remain elevated compared to HIV-1 uninfected individuals (Valdez et al., 2002). Elevated CD4+ and CD8+ T cell immune activation levels associate with shorter survival time independent of viral load (Giorgi et al., 1999) and with poor ART mediated CD4+ T cell recovery (Hunt et al., 2003). Additionally, CD8+ T cell activation independently predicts disease progression and mortality in people living with HIV-1 on ART (Hunt et al., 2011). Low CD4+ T cell count and high levels of CD4+ T cell activation prior to HIV seroconversion associate with higher risk of developing AIDS after seroconversion (Hazenberg et al., 2003). In people living with HIV-1 on ART, low CD4/CD8 ratio is associated with morbidity and mortality from serious non-AIDS events (Serrano-Villar et al., 2014).

HIV-1 infection is also characterised by high levels of polyclonal B cell activation and dysfunction, which impacts antibody production (Lane et al., 1983). HIV-1 infection also affects the phenotype and function of natural killer cells (NK cells), leading to reduced cytokine production, dysfunctional homing and increased cytotoxicity (Alter et al., 2004, Reeves et al., 2010). High levels of pro-inflammatory cytokines such as interferon-α (IFN-α), Interleukin-1 (IL-1), IL-6, IL-18 and tumour necrosis factor-α (TNF-α) are also present in PLWH (Connolly et al., 2005). Plasma levels of IL-6 and the coagulation biomarker D-dimer have been shown to be predictors of death and serious non-AIDS defining illnesses in ART treated people living with HIV-1 (Grund et al., 2016).

1.3.1 Causes of immune activation

The causes of underlying immune activation during HIV-1 infection are complex and multifactorial. Several factors which directly or indirectly cause persistent immune activation and inflammation have been identified including HIV-1 replication, viral and other co-infections and microbial translocation.

The continuous HIV-1 replication during the HIV-1 infection course in itself causes immune activation. HIV-1 antigens can cause immune activation directly or indirectly (Appay and Sauce, 2008). Direct immune activation results from the infection of T cells (recognition and T cell activation) (Hunt, 2007). Indirect immune activation has been shown by *in vitro* studies where the binding of gp120 to CD4 or co-receptors is able to activate cells even in the absence of a direct infection (Merrill et al., 1989). The HIV-1 Nef protein induces the production of the chemokines, macrophage inflammatory proteins 1α (MIP-1α; CCL3) and MIP-1β (CCL4), and soluble factors, sICAM and sCD23, by HIV-1 infected macrophages (Swingler et al., 1999). This results in the recruitment and activation of resting T cells, making them more susceptible to infection by HIV-1 (Swingler et al., 1999). HIV-1 replication can also indirectly cause immune activation through the depletion of regulatory T cells (Tregs) (Eggena et al., 2005). Tregs are $CD4^{+}CD25^{bright}CD62L^{high}$ cells involved in down-regulation of T-cell activation (Eggena et al., 2005).

Viral infections such as Epstein-Barr virus (EBV), cytomegalovirus (CMV), hepatitis B virus (HBV), and hepatitis C virus (HCV) are common in individuals living with HIV-1. These co-infections contribute to immune activation, including in individuals with HIV-1 on ART (Naeger et al., 2010). With HIV-1/CMV coinfection, approximately 10% of the circulating memory T cells were found to be CMV-specific (Sylwester et al., 2005). Valganciclovir

treatment of asymptomatic CMV, in ART treated HIV-1 infection, results in decreased frequencies of activated CD8+ T cells, showing that CMV significantly contributes to immune activation in these individuals (Naeger et al., 2010). In addition to coinfections with viruses, bacterial and fungal coinfections also contribute to the immune activation in PLWH (Vinhaes et al., 2021).

Microbial translocation has also been shown to cause persistent immune activation in people living with HIV-1 as well as in SIV infected rhesus macaques (Brenchley et al., 2006).

1.4 Microbial translocation

Microbial translocation is defined as the passage of microbes (whole bacteria or fungal cells) or microbial products from the gastrointestinal tract into the circulatory system (Alexander et al., 1990). Translocated microbial products include ribosomal DNA, lipoteichoic acid (LTA), peptidoglycan, lipopolysaccharide (LPS), flagellin and unmethylated CpG-containing DNA (Sandler and Douek, 2012). The lumen of the gastrointestinal tract harbours a large quantity of different bacterial species, which play a central role in human health by influencing human physiology, nutrition, metabolism and immune function (Guinane and Cotter, 2013). The interactions between the gut microbiota and the host are complex and can be either beneficial or detrimental.

1.4.1 Defences against microbial translocation

In healthy individuals, there are several physical and immunological factors that prevent the translocation of symbiotic bacteria from the gastrointestinal tract to the systemic circulation. Firstly, the mucous layer of the gastrointestinal tract protects the epithelial surfaces, primarily through the gel-forming properties of glycoprotein mucins (Brenchley and Douek, 2012). Glycoprotein mucins are constituents of the mucus layer and their role is to protect the epithelial cells from dehydration, infection, and chemical or physical damage (Perez-Vilar and Hill, 1999). A tight barrier of epithelial cells in the gastrointestinal tract also prevents microbial translocation. The gastrointestinal tract epithelium is predominantly comprised of enterocytes, enteroendocrine cells, goblet cells, and Paneth cells (Xu and Ghishan, 2018). While enterocytes and enteroendocrine cells are involved in nutrient absorption and secretion of peptide hormones, respectively, goblet cells produce the protective mucins and Paneth cells secrete growth factors, digestive enzymes and antimicrobial defensins (Xu and Ghishan, 2018).

Microbial products traversing the mucous layer and the epithelial barrier can be cleared by specialised intestinal macrophages before they enter the circulatory system (Smith et al., 2005).

These macrophages phagocytose microbial products without resulting in the production of pro-inflammatory cytokines and this has been attributed to the fact that they do not express innate immune response receptors (Smith et al., 2005, Smythies et al., 2005). Microbes or microbial products, which bypass the specialised intestinal macrophages are transported to the liver via the portal vein, and they are cleared through phagocytosis by the Kupffer cells (Broadley et al., 2016, Zeng et al., 2016). High levels of immunoglobulin **(**Ig) M, IgG, and IgA specific for the LPS core antigen, termed endotoxin core antibodies (EndoCAb), are present in healthy individuals (Cohen and Norins, 1966, Strutz et al., 1999). These antibodies bind and neutralise LPS in the circulatory system. Additionally, LPS can be cleared by high-density lipoproteins in the presence of high concentrations of soluble cluster of differentiation 14 (sCD14) and LPS-binding protein (LBP) without initiating inflammatory immune responses (Kitchens et al., 2001, Zweigner et al., 2001).

1.4.2 Causes of microbial translocation

Several factors contributing to microbial translocation have been described. The disruption of the tight epithelial junctions of the epithelial layer is a major cause of microbial translocation. The production of pro-inflammatory cytokines due to the HIV-1 infection induces abnormal enterocyte differentiation and apoptosis of the enterocytes causing the disruption of tight epithelial junctions (Epple et al., 2010, Sandler and Douek, 2012). Exposure to HIV-1 directly impairs the integrity of the epithelial barrier (Nazli et al., 2010). The preferential depletion of CD4+ T cells from the gut-associated lymphoid tissue (GALT) in the acute stage of HIV-1 infection results in the loss of T helper 17 cells (Th17) cells (Brenchley et al., 2008). Th17 cells are involved in neutrophil recruitment, antimicrobial peptide production, and epithelial regeneration through IL-17 and IL-22 production (Liu et al., 2009).

In people living with HIV-1 coinfected with Hepatitis C, the loss of Kupffer cells results in reduced clearance of microbial products, promoting microbial translocation (Balagopal et al., 2009). In addition, HIV-1 infection leads to the depletion of macrophages and defective macrophage phagocytosis, which contributes to the impaired clearance of microbial products and allows them to enter the circulatory system (Kedzierska et al., 2003). During HIV-1 infection, there is B cell dysregulation resulting in low levels of IgA, which might lead to bacterial overgrowth contributing to increased microbial translocation (Kulkarni and Ruprecht, 2017). IgA is important in microbial clearance. It binds to bacteria and transports them back to the lumen from the lamina propria or binds to the bacteria to prevent them from binding to

epithelial cells, and also binds and neutralizes microbial products (Macpherson et al., 2012, Klatt et al., 2013).

1.4.3 LPS recognition and signalling

Translocated microbial products can directly elicit innate immune responses through activation of toll-like receptors (TLR1 to TLR9) and nucleotide-binding oligomerization domain 1 and 2, resulting in production of cytokines (Sandler and Douek, 2012). LPS, a major component of the gram-negative bacterial outer cell wall, initiates signal transduction though the TLR4 receptor (Ulevitch and Tobias, 1995). Structurally, LPS consists of a hydrophobic domain which is a lipid A molecule, a non-repeating "core" oligosaccharide (core sugar), and a distal polysaccharide (O-antigen) (Raetz and Whitfield, 2002). The Lipid A component of LPS is the endotoxin and the important component for TLR4 recognition (Kawai and Akira, 2010).

The recognition and binding of LPS to TLR4 for signalling requires accessory molecules and receptors which include LPS-binding protein (LBP), CD14 and Lymphocyte antigen 96 (LY96), also known as Myeloid differentiation factor 2 (MD-2) (Lu et al., 2008). LBP binds to LPS and catalytically transfers it to membrane bound CD14 (mCD14) or soluble CD14 (sCD14), which subsequently transfers it to MD-2, in the TLR4-MD-2 complex. The binding of LPS to MD-2 in the TLR4-MD-2 complex, results in conformational changes in TLR4 that initiate downstream signal transduction resulting in Nuclear factor-kappa beta (NF-kβ) activation and production of cytokines (Park et al., 2009, Widera et al., 2019). The TLR4 signalling pathways are shown in **Figure 1.3** and will be discussed in detail in **Section 1.5.4.**

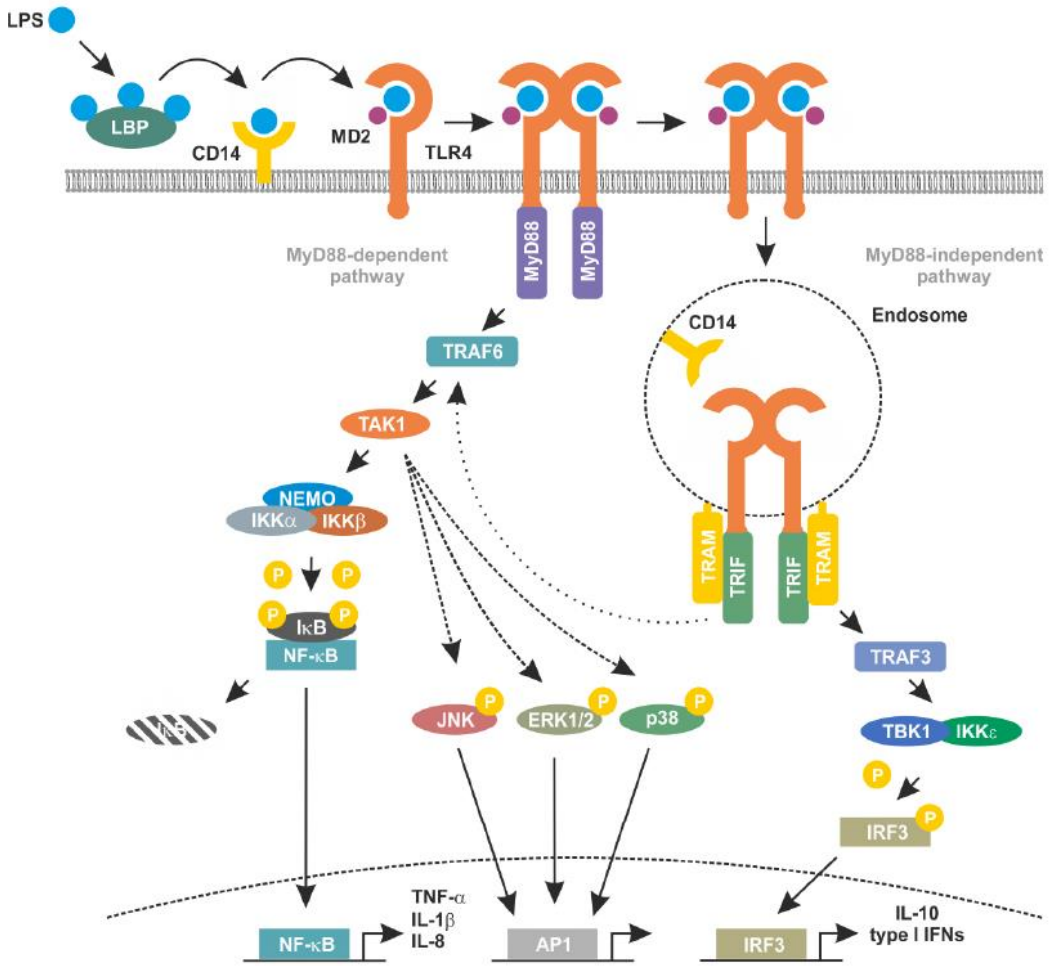


Figure 1.3: A Schematic representation of the TLR4 signalling pathway. LPS binds to LBP and transfers it to mCD14 or sCD14 (not shown) depending on the cell type. CD14 then transfers the LPS to TLR4 bound MD-2 causing conformational changes, which enables the recruitment of adapter molecules and TLR4 signal transduction through MyD88-dependent and MyD88-independent pathways. The MyD88-independent pathway occurs post receptor complex internalisation by endocytosis. Image reproduced from (Widera et al., 2019).

1.5 TLR4 signalling receptors and accessory molecules

1.5.1 Lipopolysaccharide binding protein (LBP)

Lipopolysaccharide binding protein is a 58–60 kDa glycosylated protein constitutively produced by the hepatic cells of the liver and is produced at higher levels during acute phase responses (Schumann et al., 1996, Hubacek et al., 1997). LBP is structurally and functionally related to bactericidal/permeability-increasing protein (BPI), which also binds to LPS, but binding of BPI to LPS diminishes the immune response through neutralisation of LPS (Elsbach and Weiss, 1995, Beamer et al., 1998).

LBP binds to the lipid A moiety of the LPS and catalytically transfers LPS to mCD14 on cells or sCD14 in plasma to initiate the LPS signal transduction. The binding of LBP to mCD14 is not solely dependent on the LPS binding to LBP; it is also dependent on the relative amount of LPS to LBP (Gegner et al., 1995). In low concentrations, LBP catalytically transfers LPS to CD14; however in high concentrations LBP can clear the LPS or LPS/CD14 complexes without the activation of the TLR4 receptor, thereby inhibiting an immune response to LPS (Zweigner et al., 2001).

In animal studies, it has been shown that approximately half of the LPS introduced into the circulatory system is cleared though binding to leukocytes which can initiate immune responses, while the remaining LPS binds to circulating lipoproteins and gets cleared by the liver (Mathison and Ulevitch, 1979). High-density lipoproteins have been identified as the major lipoproteins involved in LPS neutralisation (Vesy et al., 2000). The multiple roles LBP plays in LPS response and LPS clearance are illustrated in **Figure 1.4**. LBP is the main protein involved in the extraction of LPS from bacterial membranes with the subsequent transfer to CD14 or directly to high-density lipoproteins (HDL) for clearance (Vesy et al., 2000). LPS-LBP complexes can also be cleared through internalisation, and this process occurs after the binding to mCD14 and is independent of the LPS signal transduction pathway (Gegner et al., 1995). Another study also showed that LBP attenuates LPS responses even after the transfer of LPS to mCD14, by rapidly removing LPS bound to mCD14 and MD-2 (Thompson et al., 2003).

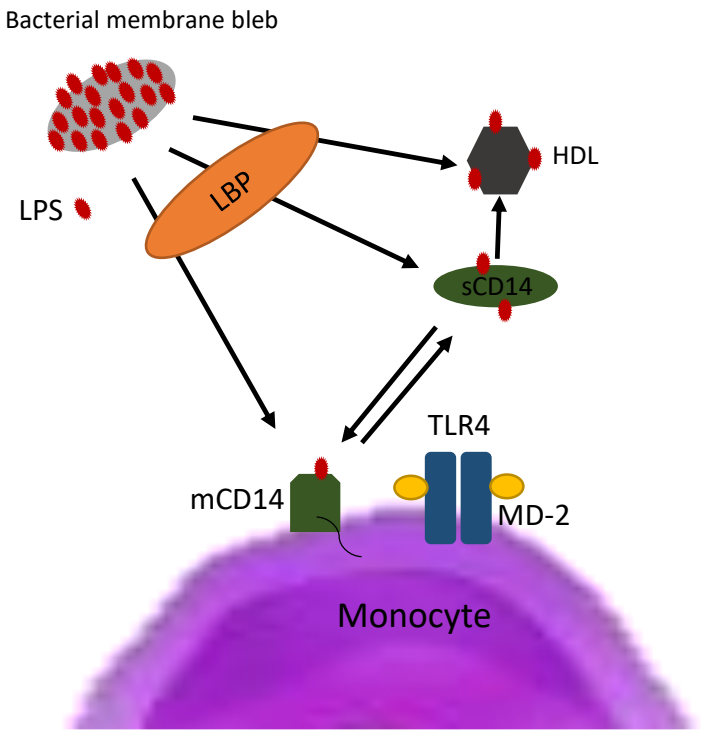


Figure 1.4: A model showing the interaction between lipopolysaccharide (LPS), LPS binding protein (LBP), high-density lipoproteins (HDL) and membrane (m) CD14-expressing cells, showing that LBP can transfer LPS to either mCD14 expressing cells, soluble CD14 or HDL. In turn, sCD14 transfers the LPS to mCD14 or to HDL; sCD14 can also transfer LPS from mCD14 to HDL. Image redrawn from (Vesy et al., 2000).

1.5.1.1 *LBP* gene

The *LBP* gene is located on the long arm of chromosome 20 (20q11.2), in a region that also encodes the *BPI* gene (Gray et al., 1993). It is comprised of 14 exons and is approximately 28.5 kb in length (Hubacek et al., 1997).

1.5.1.2 *LBP* genetic variants

Given that the different functions of LBP are dependent on the circulating levels, variants that affect LBP structure and expression levels can affect immune response to LPS, as well as the pathogenicity of various bacterial infections and diseases. In the coding region of the *LBP* gene, a +988C>T (rs2232613) single nucleotide polymorphism (SNP) has been associated with reduced cytokine production after LPS stimulation and an elevated risk of mortality from pneumonia and sepsis (Eckert et al., 2013). It has been suggested that the +988C>T mutation

introduces a proteinase cleavage site with the resulting mutant protein having reduced LPS and lipopeptide binding capabilities (Eckert et al., 2013).

Chien et al. (2008) found an association between the rs2232582 SNP (a synonymous Proline mutation) and gram-negative bacteraemia. They used the rs2232582 SNP as a tagSNP for nine *LBP* SNPs in linkage disequilibrium (LD) and three of those SNPs, namely, rs2232571, rs2232575, and rs2232578 were located in 5' *LBP* promoter region, with the rs2232571 SNP being associated with higher LBP plasma levels (Chien et al., 2008). The rs2232571 SNP is a C to T substitution (position: -788) in the CAAT box (transcription factor binding site) and the location of this SNP suggests that this variant might be enhancing the efficiency of the *LBP* promoter (Schumann et al., 1996, Chien et al., 2008).

Another variant, the rs2232618 SNP in the *LBP* gene, has been associated with the risk of sepsis and multiple organ dysfunction after trauma (Zeng et al., 2012a, Lu et al., 2018). This rs2232618 SNP results in a phenylalanine to leucine substitution in the C–terminal domain of LBP, a region essential for CD14 binding, and causes spatial conformational changes to the C–terminal domain active centre (Zeng et al., 2012a). *In vitro* studies showed that the presence of the rs2232618 SNP resulted in enhanced LPS transfer from CD14 and LPS induced cell activation (Zeng et al., 2012a).

A four SNP haplotype comprised of the rs1780616, rs5741812, rs2232571 and rs1780617 SNPs in the 5' *LBP* promoter region has also been associated with the risk of severe sepsis (Flores et al., 2009). Homozygosity for the four SNP haplotype was associated with higher LBP serum levels over time and at specific time points in individuals with sepsis when the data was controlled for demographic and clinical factors (Flores et al., 2009). It is important to note that the rs2232571 SNP, part of this four SNP haplotype, has been previously associated with higher plasma levels of LBP in healthy individuals as mentioned above (Chien et al., 2008).

1.5.2 Cluster of differentiation 14 (CD14)

Cluster of differentiation 14 exists as a glycosylphosphatidylinositol-anchored protein (mCD14) mostly expressed on the surfaces of monocytes, macrophages and neutrophils and as a soluble form (sCD14) derived from cleavage of the mCD14 or direct secretion from the liver (Ulevitch and Tobias, 1995, Su et al., 1999). The CD14 protein is approximately 50- to 55kD in size.

In the TLR4 signalling pathway, mCD14 plays a major role as a receptor for LPS-LBP complexes (Wright et al., 1990). The soluble form of CD14 (sCD14) also binds LPS-LBP complexes in cells which do not express mCD14 conferring responsiveness to LPS (Frey et al., 1992, Pugin et al., 1993). Soluble CD14 transfers the LPS from the LBP to either mCD14 or directly to TLR4-MD-2 complexes (Pugin et al., 1993). LPS binds to the NH_2-terminal hydrophobic pocket in the CD14 structure (Viriyakosol and Kirkland, 1995, Kim et al., 2005), with its binding to CD14 being enhanced by the LBP (Schumann et al., 1990). Upon the transfer of LPS from LBP, CD14 subsequently transfers the LPS to MD-2, in the TLR4-MD-2 complex in order to initiate LPS signalling. The region near the NH_2-terminal pocket of the CD14 has been implicated in the transfer of LPS from CD14 to MD-2 (Kim et al., 2005).

Like LBP, sCD14 is also involved in neutralisation of LPS. Subsequent to LBP transferring LPS to sCD14, LPS is transferred to either mCD14 to initiate the immune response or to HDL for neutralisation (Vesy et al., 2000). However, unlike LBP, sCD14 also has the ability to transfer LPS from the surface of mCD14 expressing monocytes to HDL for neutralisation as illustrated in **Figure 1.4** (Kitchens et al., 1999). The transfer of LPS to HDL is concentration dependent. When present in high concentrations, sCD14 can down-regulate immune responses to LPS, through the binding and transfer of LPS to plasma lipoproteins (Kitchens et al., 2001).

In addition to its role in LPS recognition through the TLR4, CD14 is also involved in the TLR2 signalling pathway triggered by a range of bacterial cell wall components such as peptidoglycan, lipoteichoic acid (LTA) and lipoarabinomannan (Pugin et al., 1994, Means et al., 1999, Schroder et al., 2003, Dziarski and Gupta, 2005). LTA is a major component of the cell wall of gram-positive bacteria, and a major virulence factor (Jang et al., 2015).

1.5.2.1 *CD14* gene

The *CD14* gene is located on the long arm of chromosome 5 (5q23-q31) in a region which encodes for receptors and growth factors (Goyert et al., 1988). The *CD14* gene only has two exons and is approximately 1.9 kb in length.

1.5.2.2 *CD14* genetic variation

Polymorphisms in the *CD14* gene have been shown to affect expression levels of both mCD14 and sCD14. A *CD14* promoter region SNP, -159C>T also known as -260C>T (rs2569190) (Eng et al., 2004) has been extensively studied. Homozygosity (TT) for the -159C>T/-260C>T mutation in the promotor region of *CD14* has been associated with higher levels of circulating sCD14 and higher mCD14 density on monocytes (Baldini et al., 1999, Hubacek et al., 1999,

Eng et al., 2004, Lin et al., 2007, Levan et al., 2008). The homozygous mutant genotype for the -159C>T/-260C>T SNP has also been associated with higher production levels of TNF-α and IL-6, and lower levels of IL-10 upon LPS stimulation (Lin et al., 2007) and increased TNF-α production after *Chlamydiae* stimulation (Eng et al., 2004). The -159C>T/-260C>T mutation enhances transcriptional activity of CD14 by decreasing the binding affinity of the GC box for Sp3 transcription factor (LeVan et al., 2001). The Sp3 transcription factor is known for inhibiting promotor activity in various genes. In addition, mutant allele homozygosity for the *CD14* SNPs, rs3138078 and rs2915863 has been associated with sCD14 levels (Levan et al., 2008). These SNPs are located in the 5' untranslated region (UTR) of the *CD14* gene. A recent study conducted in individuals of African ancestry reported associations between lower plasma levels of sCD14 and a *CD14* promotor SNP (rs75652866) as well as a rare frameshift deletion (rs770147646) (Stanislawski et al., 2021).

In a study conducted on asthmatic black South African children, -159C>T/-260C>T genotypes were not observed to influence sCD14 levels (Makamure et al., 2017). In children, it has been shown that differences in DNA methylation affects the association of *CD14* polymorphisms with levels of sCD14. A longitudinal study investigated the association of four *CD14* SNPs (rs2569191, rs5744455, rs2569190 and rs4914) with sCD14 levels at different ages (birth, two years and 10 years) (Munthe-Kaas et al., 2010). They found that the rs2569191, rs5744455, and rs2569190 SNPs were associated with sCD14 levels at birth and at two years, however, only the rs5744455 SNP associated with sCD14 at 10 years of age (Munthe-Kaas et al., 2010). Importantly, *CD14* methylation was significantly increased from two to 10 years of age, and was inversely associated with levels of sCD14 at the age of 10 years (Munthe-Kaas et al., 2010).

Regulation of *CD14* gene expression is important in a range of different diseases, and consequently, polymorphisms in the *CD14* gene have been associated with a number of diseases. These include myocardial infarction in different populations (Hubacek et al., 1999, Unkelbach et al., 1999, Shimada et al., 2000), asthma (Nabih et al., 2016), and sepsis (Wang et al., 2014).

1.5.3 Myeloid differentiation factor 2 (MD-2)

Myeloid differentiation factor 2, also known as Lymphocyte antigen 96 (LY96), is a 20-30-kDa glycoprotein that binds to the extracellular domain of TLR4 and confers responsiveness to LPS (Shimazu et al., 1999, Gray et al., 2010). MD-2 binds to the amino-terminal region of

the TLR4 extracellular domain to form the TLR4/MD-2 complex (Fugimotti et al 2004). MD-2 is an endotoxin-binding molecule; it binds to LPS with high affinity to facilitate its transfer from CD14. A 14 amino acid sequence in the MD-2 protein, which has Lys^{128} and Lys^{132} residues, is required for LPS recognition (Manček et al., 2002, Visintin et al., 2003). Resman et al. (2009) showed that hydrophobic residues located at position 82, 85 and 87 of the MD-2 protein are essential for TLR4 activation and the transfer of LPS from CD14 to MD-2.

Post synthesis, MD-2 is either directly secreted as a soluble active protein by primary cells or binds directly to TLR4 in the endoplasmic reticulum before the complex migrates to the cell surface (Visintin et al., 2001). *In vitro* studies have shown that MD-2 plays a crucial role in the surface distribution of TLR4 and the translocation of TLR4 from the Golgi apparatus to the cell surface (Nagai et al., 2002). The presence of MD-2 is required for the specific glycosylation which is essential for the cell surface expression of TLR4 (Ohnishi et al., 2003).

In cells which express TLR4 but not MD-2, the soluble form of MD-2 has the ability to bind to the TLR4 on the surface and confer LPS responsiveness (Abreu et al., 2001, Visintin et al., 2001, Kennedy et al., 2004). Intestinal epithelial cells (IEC) do not elicit an immune response to commensal bacteria found in lumen (Abreu et al., 2001). Studies using IEC cell lines have shown that these cells prevent LPS signalling by down-regulating the expression of both TLR4 and MD-2 (Abreu et al., 2001). Transfection of IEC cell lines with MD-2 but not TLR4, increases LPS responsiveness by more than 100-fold (Lenoir et al., 2008).

Interestingly, although MD-2 is primarily involved in TLR4 signalling, it has been shown that MD-2 weakly associates with TLR2 and enhances the responsiveness of TLR2 to gram-negative and gram-positive bacteria and their LPS and LTA cell wall components (Dziarski et al., 2001). This further demonstrates the multiple roles or the extent of the importance of MD-2 in the response to different microbial products.

An alternatively spliced MD-2 isoform (lacking exon 2) termed MD-2 short (MD-2s) which, like MD-2 is glycosylated and secreted, has been identified (Gray et al., 2010). MD-2s does not result in the production of pro-inflammatory cytokines and acts as an inhibitor of LPS-mediated TLR4 activation (Gray et al., 2010). Another alternatively spliced MD-2 isoform lacking the first 54 bases of exon 3, termed MD-2B, has also been described (Ohta et al., 2004). Although, MD-2B efficiently binds to TLR4, the TLR4/MD-2B complexes are not expressed on the cell surface thereby reducing LPS responsiveness (Ohta et al., 2004).

1.5.3.1 *LY96* gene

The *LY96* gene encoding for MD-2 is located in long arm of chromosome 8 (8q21.11). It is comprised of five exons and is approximately 38 kb in length.

1.5.3.2 *LY96* genetic variation

Although the *LY96* gene is highly variable, little is known about the functional significance of the variants in this gene. Given that MD-2 has an indispensable role in LPS signalling, genetic variants in the *LY96* gene have the potential to result in hypo- or hyper-responsiveness to LPS.

Of the few characterised variants in the *LY96* gene, a C to G change (rs11465996) at position -1625 of the *MD-2* gene promotor region has been shown to alter the promotor activity (Gu et al., 2007). Upon LPS stimulation in whole blood, individuals with the mutant allele of the −1625C>G SNP (G allele) were found to have increased expression of *LY96* mRNA and TNF-α (Gu et al., 2007). In addition, heterozygosity for the −1625C>G SNP has been associated with reduced risk of ulcerative colitis (Bank et al., 2014).

A rare +103A>G SNP (CM042737) in the first exon of *LY96*, which results in a threonine to alanine substitution, was found to result in reduced NF-kβ-activation and TNF-α production upon LPS stimulation (Hamann et al., 2004). A study investigating the functional effects of two missense *LY96* variants namely, Arg56Ter and Pro157Ser (rs6472812 and rs11466004, respectively) found that the rs6472812 but not the rs11466004 mutation resulted in reduced LPS responsiveness due to reduced transfer of LPS from CD14 to the mutant MD-2 (Vasl et al., 2008).

In a Taiwanese population study, two SNPs in the promoter region of *LY96* namely, rs1809441 and rs1809442 were found be associated with *Dermatophagoides pteronyssinus* group 2 (Der p 2) allergy (Liao et al., 2015a). Der p 2 is a dust mite allergen which has structural homology to MD-2. Presence of the rs1809441 or rs1809442 SNPs in the promoter region is associated with increased MD-2 expression. Additionally, stimulation with Der p 2 and LPS resulted in higher levels of TNF-α, IL-1β, IL-6, IL-8, and IL-10 expression (Liao et al., 2015b).

Given that variants in the *CD14* and *TLR4* genes have been associated with Tuberculosis (TB) (Ferwerda et al., 2007, Rosas-Taraco et al., 2007), a Chinese study investigated the role of six *LY96* promoter variants (rs11465996, rs1809442, rs1809441, rs1809440, rs16938754, and rs7842342) with susceptibility to TB, however no significant associations were found (Xue et al., 2010).

1.5.4 Toll-like receptor 4 (TLR4)

Toll-like receptor 4 (TLR4) is a type 1 transmembrane protein consisting of an extracellular domain with leucine rich repeats, a transmembrane domain and a cytoplasmic domain with high homology to the interleukin-1 receptor (Kang and Lee, 2011). TLR4 is mainly expressed by cells of myeloid origin, such as monocytes, macrophages and a subset of dendritic cells (Muzio et al., 2000).

Binding of LPS to TLR4 bound MD-2, forming a TLR4-LPS-MD-2 complex, induces dimerization of the complex. The dimerization results in the formation of an m–shaped receptor multimer consisting of two copies of symmetrically arranged TLR4-LPS-MD-2 complexes (Park et al., 2009). The TLR4 conformational changes enable the recruitment of Toll-Interleukin receptor (TIR) domain-containing adapter molecules. There are four types of TIR domain-containing adapter molecules involved in TLR4 signalling, namely: (i) Myeloid differentiation factor 88 (MyD88), (ii) TIR domain-containing adaptor protein (TIRAP) also known as Mal, MyD88-adapter-like, (iii) TIR-domain containing adapter inducing IFN-β (TRIF), and (iv) TRIF-related adaptor molecule (TRAM) (Maeshima and Fernandez, 2013, Vaure and Liu, 2014). An overview of the TLR4 signalling is shown in **Figure 1.3**.

TLR4 initiates intracellular signalling though the MyD88-dependent and TRIF-dependent (MyD88-independent) pathways (Vaure and Liu, 2014). The My88-dependent pathway involves the recruitment of TIRAP and My88 adapter molecules post the TLR4-MD-2 receptor dimerization (Kagan and Medzhitov, 2006). This results in the early activation of the NF-κβ transcription factor and the production of pro-inflammatory cytokines such as IL-6, TNFα and IL-12 (Maeshima and Fernandez, 2013).

The TRIF-dependent pathway involves the recruitment of TRAM and TRIF adapter molecules (Maeshima and Fernandez, 2013). The recruitment of these adapter molecules occurs after the TLR4-MD-2 receptor complex dimer internalisation into endosomes (Vaure and Liu, 2014). This pathway activates the interferon regulatory factor-3 (IRF3) transcription factor and results in upregulation of type 1 INFs and the production and secretion of TNF-α, which subsequently leads to the late activation of the NF-κβ. The majority of the LPS response is through the TRIF-dependent pathways (Vaure and Liu, 2014). Another TIR-domain-containing adapter molecule known as SARM (sterile α and HEAT-Armadillo motifs-containing protein) inhibits the activation of NF-κβ mediated by the TRIF-dependent pathway (Carty et al., 2006). The LPS

mediated TLR4 activation induces the expression of the SARM encoding gene (Carty et al., 2006).

1.5.4.1 *TLR4* gene

The *TLR4* gene is located on the long arm of chromosome 9 (99q32-q33) (Rock et al., 1998). It is comprised of three exons and is approximately 19 kb in length.

1.5.4.2 *TLR4* genetic variants

Two *TLR4* polymorphisms, namely Asp299Gly (rs4986790) and Thr399Ile (rs4986791), have been extensively studied. These polymorphisms are co-segregating in Caucasians (Arbour et al., 2000), but not in West African populations (Newport et al., 2004), illustrating the importance of population specific studies. These variants result in amino acid changes which are located within the leucine rich repeats in the TLR4 extracellular domain (Rallabhandi et al., 2006). The Asp299Gly and Thr399Ile TLR4 variants have been shown to result in hypo-responsiveness to inhaled LPS in humans (Arbour et al., 2000). *In vitro*, it has been shown that the Asp299Gly and Thr399Ile mutant TLR4 results in reduced NF-κβ reporter gene expression in response to LPS stimulation (Rallabhandi et al., 2006). Ohto et al. (2012) reported that the Asp299Gly mutation, rather than the Thr399Ile, resulted in local structural changes which might affect the binding sites. Additionally, *in vitro* studies have shown that it is the Asp299Gly mutation, rather than the Thr399Ile mutation, which is responsible for the reduced LPS signalling (Arbour et al., 2000, Long et al., 2014). The Asp299Gly and Thr399Ile variants have been associated with increased susceptibility to gram-negative bacterial infections (Agnese et al., 2002, Lorenz et al., 2002).

In addition, the Asp299Gly and Thr399Ile variants have been associated with susceptibility towards or disease severity of a range of infectious and non-infectious diseases (Noreen et al., 2012). With respect to HIV-1 infection, the Asp299Gly and Thr399Ile SNPs have been found in higher frequencies in individuals with high viral loads compared to those with low viral loads (Pine et al., 2009). Heterozygosity for these SNPs, however, has been associated with late $CD4^+$ T-cell recovery after ART (Yong et al., 2016).

Another SNP in the 3'UTR of the *TLR4* gene, namely rs11536889, plays a role in the upregulation of TLR4 expression (Sato et al., 2012). The homozygous mutant genotype for this SNP correlated with significantly higher expression levels of TLR4 on the surface of monocytes (Sato et al., 2012). Peripheral blood mononuclear cells (PBMCs) from individuals

with heterozygous or homozygous genotypes for the rs11536889 SNP were shown to produce higher levels of IL-8 in response to LPS stimulation (Sato et al., 2012). This variant has also been associated with various diseases including prostate cancer (Zheng et al., 2004), gastric atrophy (Hishida et al., 2009), recurrence of hepatitis B after liver transplant (Zhou et al., 2011), and intracranial aneurysm (Liu et al., 2018). A -2242T>C *TLR4* SNP in the promoter region of the *TLR4* gene was also found to enhance the activity of the *TLR4* promotor (Yongtang et al., 2009). Furthermore, a seven SNP haplotype, and two polymorphisms, rs10759932 (in Caucasians) and rs7873784 (in African Americans) in the *TLR4* gene, were found to be more frequent in people living with HIV-1 compared to HIV-1 uninfected individuals (Willie et al., 2014). This suggests a role for *TLR4* in the context of susceptibility to acquisition of HIV-1 infection (Willie et al., 2014).

1.6 HIV and microbial translocation

1.6.1 Markers of microbial translocation

Direct or indirect markers can be used to measure microbial translocation in the circulatory system. Direct markers include plasma or serum levels of LPS and bacterial 16S ribosomal RNA (16S rRNA) gene copy numbers (Marchetti et al., 2013). As these markers are not normally present in the circulatory system, their presence indicates that microbial translocation has taken place. Indirect markers of microbial translocation include the measurement of sCD14, EndoCAb, and LBP levels in plasma or serum (Marchetti et al., 2013). Soluble CD14 is a marker of immune activation and elevated levels indicate monocyte activation in response to LPS and other ligands (Landmann et al., 1996, Shive et al., 2015). Low levels of EndoCAb are indicative of LPS mediated chronic immune stimulation (Brenchley et al., 2006), while LBP levels are elevated in the presence of LPS.

1.6.2 Evidence of microbial translocation in HIV-1 infection

Brenchley et al. (2006) were the first to demonstrate an association between microbial translocation and persistent immune activation in people living with HIV-1. They found plasma levels of LPS to be significantly higher in people living with chronic HIV-1 infection as well as in SIV-infected rhesus macaques. It was subsequently shown that the gastrointestinal tract was the source of the translocated LPS in SIV-infected rhesus macaques. Furthermore, compared to SIV-infected rhesus macaques, which progress to AIDS, no evidence of microbial translocation was found in SIV-infected sooty mangabeys, which are non-human primates that do not develop AIDS despite high levels of viral replication (Brenchley et al., 2006).

Several subsequent studies have also investigated the role of microbial translocation in HIV-1 infection, immune activation and disease progression in different populations. Elevated plasma levels of sCD14, reduced EndoCAb, high LBP levels, high LPS levels and high levels of bacterial 16S rRNA gene copies were found to be associated with chronic HIV-1 infection, circulating $CD8^+$ $CD38^+$ HLA-DR^+ T cells, high viral load, low $CD4^+$ T cell count and disease progression (Lien et al., 1998, Jiang et al., 2009, Nowroozalizadeh et al., 2010, Marchetti et al., 2011, Reus et al., 2013, Pilakka-Kanthikeel et al., 2014). Another study found that sCD14 but not LPS levels associate with markers of disease progression (Romero-Sanchez et al., 2012).

High levels of sCD14 have been shown to correlate with high levels of LPS and LBP and low levels of EndoCAb (Ancuta et al., 2008). However, the correlation between sCD14 and LPS has been shown to be dependent on $CD4^+$ T cell levels in individuals with chronic HIV-1 infection (Romero-Sanchez et al., 2012). A weaker association between LPS and sCD14 in black individuals has been observed, suggesting that other factors (biological and environmental) affect sCD14 expression levels, further demonstrating that sCD14 is not a direct marker of microbial translocation but is a marker of immune activation (De Voeght et al., 2016). Plasma levels of sCD14 have been found to be independent predictors of mortality and disease progression in HIV-1-infected individuals (Marchetti et al., 2011, Sandler et al., 2011, Leon et al., 2015).

In black African children and adults living with HIV-1, microbial translocation studies have yielded conflicting results. In Uganda, no associations were found between microbial translocation and HIV-1 disease progression in adults (Redd et al., 2009) and immune activation in children (Fitzgerald et al., 2018). However, associations between microbial translocation and HIV-1 infection in adults and children were reported in South Africa and Guinea-Bissau (Cassol et al., 2010, Nowroozalizadeh et al., 2010, Papasavvas et al., 2011). The inconsistencies between the microbial translocation studies may be attributed to the poor reproducibility of the LPS Limulus lysate assay and background contamination in 16S rRNA quantitative assays (Marchetti et al., 2013).

1.6.3 Microbial translocation in ART treated people living with HIV-1

Several studies found that although ART decreases microbial translocation, the extent of microbial translocation remains higher in people living with HIV-1 on ART compared to healthy controls (Brenchley et al., 2006, Cassol et al., 2010). Microbial translocation associates

with poor CD4+ T cell reconstitution in people living with HIV-1 on ART (Marchetti et al., 2008, Jiang et al., 2009).

Other studies, however, found that while partially suppressed (VL=2.5–50 RNA copies/ml) individuals had plasma LPS levels higher than HIV-1 uninfected individuals, fully suppressed individuals (VL<2.5 RNA copies/ml) had LPS levels comparable to those of healthy controls, suggesting that to some extent successful ART can normalise microbial translocation (Baroncelli et al., 2009), but this remains debatable and is likely to be partially dependent on the individual.

1.6.4 Microbial translocation in HIV-1 controllers

Elite controllers have higher levels of microbial translocation (measured by LPS, sCD14 and LBP) than HIV-1 uninfected individuals, but lower levels than HIV-1 progressors (Brenchley et al., 2006). High levels of LPS in elite controllers were found not to result in reduced EndoCAb levels, suggesting that elite controllers are able to neutralise LPS more effectively by maintaining high levels of EndoCAb (Brenchley et al., 2006). Elevated plasma levels of LPS have also been associated with high levels of $CD8^+$ T activation and $CD4^+$ T cell depletion in HIV-1 controllers (Hunt et al., 2008). Markers of microbial translocation (sCD14 and EndoCAb) have been shown to independently predict clinical outcomes such as progression to AIDS and development of serious non-AIDS events in HIV-1 controllers (Leon et al., 2015).

A study conducted in Uganda did not find differences in LPS, LBP and EndoCAb levels between individuals with different HIV-1 phenotypes (LTNPs, standard progressors and rapid progressors) (Redd et al., 2009). This was confirmed by a study in India, which found that plasma levels of sCD14, LPS and IgM antibodies did not differ significantly when rapid progressors ($CD4^+$ T cells reaching <350 cells/µl within 3-4 years of infection) were compared to viraemic slow progressors ($CD4^+$ T cells > 500 cells/µl for >7 years with VL >10 000 RNA copies/ml) and slow progressors ($CD4^+$ T cells > 500 cells/µl for >10 years with VL <10 000 RNA copies/ml) (Negi et al., 2017). In addition, Kim et al. (2014) did not find differences in markers of microbial translocation (LPS and sCD14) between elite controllers and healthy HIV-1 uninfected individuals. The inconsistencies in the role of microbial translocation in HIV-1 controllers might be due to the differences in the definitions of the HIV-1 controller subgroups, the methods they used or population differences, however they necessitate further research into the extent of microbial translocation in HIV-1 controllers.

Th17 cells are preserved in elite controllers compared to individuals with advanced HIV-1 infection (Hartigan-O'Connor et al., 2011), which might partially account for the reduced levels of microbial translocation in this group. A higher frequency of Th17 cells has also been observed in elite controllers compared to healthy HIV-1 uninfected individuals (Kim et al., 2014). Th17 cells are also preserved in the blood and gastrointestinal tract of SIV infected sooty mangabeys (Brenchley et al., 2008). As previously mentioned, Th17 cells play a crucial role in the maintenance of the integrity of intestinal epithelial barrier (Brenchley and Douek, 2012).

1.6.5 Intestinal damage in HIV-1 infection

Since the impairment of the gut barrier is a cause of microbial translocation, studies have also looked at the association between markers of intestinal damage and microbial translocation, immune activation and HIV-1 infection. Plasma levels of intestinal fatty acid binding protein (I-FABP) are commonly used as a marker of intestinal (enterocyte) damage (Adriaanse et al., 2013). I-FABP is constitutively expressed by enterocytes, and is released into the circulatory system upon intestinal cell death (Al-Saffar et al., 2017).

In untreated people with HIV-1, I-FABP levels have been shown to increase over time (El Kamari et al., 2019). The epithelial gut damage is initiated early on during the HIV-1 infection and the I-FABP levels are not normalised by ART (Jenabian et al., 2015, Sereti et al., 2017). I-FABP levels were found to be significantly higher in HIV-1 progressors compared to elite controllers and healthy HIV-1 uninfected individuals (Cheru et al., 2018), and remained higher in ART treated individuals with undetectable viral loads compared to elite controllers and HIV-uninfected individuals (Cheru et al., 2018). In contrast, other studies found that I-FABP levels were not significantly different between ART suppressed people living with HIV-1 and HIV-1 uninfected controls (Younas et al., 2019) and between elite controllers and HIV-1 uninfected controls (Isnard et al., 2020). While some studies found that I-FABP levels correlated with markers of microbial translocation (LPS) and immune activation (sCD14) (Hunt et al., 2014, Cheru et al., 2018), other studies did not see the correlations (Romero-Sanchez et al., 2012, Isnard et al., 2020). I-FABP plasma levels have also been reported to be predictors of mortality in ART treated PLWH (Hunt et al., 2014).

Regenerating islet-derived 3 α (REG3α) is also a marker of gut damage. REG3α is selectively secreted by Paneth cells in the small intestine in the presence of bacteria, and it is translocated into the circulatory system if the integrity of the gut barrier is compromised (Marafini et al., 2014). Levels of REG3α were found to be elevated in untreated and ART treated individuals

with HIV-1, and elite controllers compared to HIV-1 uninfected individuals (Isnard et al., 2020). Unlike I-FABP levels, REG3α levels were found to correlate with markers of disease progression, viral load and CD4+ T cell count (Isnard et al., 2020). Furthermore, REG3α correlated with epithelial gut damage (I-FABP), microbial translocation (LPS) and immune activation (sCD14) in individuals with HIV-1 (Isnard et al., 2020).

1.7 Study rationale

Several genetic variants in the genes (*LBP*, *CD14*, *TLR4*, and *LY96* genes) encoding receptors and accessory molecules required for the LPS recognition and signalling through the TLR4 receptor have been shown to affect their respective protein expression levels and responses to endotoxins. In addition, these variants have been associated with susceptibility to and severity of various immune-related diseases, thereby further emphasising the key role of these four proteins in immune responses. Little however is known about genetic variation in these genes in African populations. Most of the genetic characterisation and disease association studies on variants within the *LBP*, *CD14*, *TLR4* and *LY96* genes have been conducted in European, American, or Asian populations. Given that African populations are genetically more diverse and show lower levels of linkage disequilibrium (LD) than non-African populations, different variants may be impacting gene regulation and/or protein function in populations of African descent. Regulatory or functional variants in these genes play an important role in the regulation of the TLR4 signalling pathway; they can result in the down- or up-regulation of the TLR4 mediated immune response to LPS. Therefore, understanding of the baseline expression levels of the TLR4 signalling molecules in different populations is essential, and characterisation of genetic variants in the genes encoding for these molecules in the South African black population can help identify population-specific variants, which potentially affect immune responses to LPS.

Furthermore, individuals able to control HIV-1 infection in the absence of antiretroviral treatment (HIV-1 controllers), particularly elite controllers, represent an important study model for informing HIV-1 vaccines, targets for HIV-1 therapies, and strategies to achieve an HIV-1 functional cure (Deeks et al., 2021). Other clinical phenotypes (viraemic controllers, high VL LTNPs, progressors) as well as HIV-uninfected individuals provide important comparator groups to better understand the extent of immune activation, fuelled in part by microbial translocation. The extent of microbial translocation in elite controllers is still not clearly understood, and no microbial translocation studies have been carried out in South African black elite controllers, illustrating the need for studies of this nature. Lastly, to the best of our

knowledge, association studies investigating the representation of *LBP*, *CD14*, *TLR4* and *LY96* gene variants in HIV-1 controllers have not been undertaken.

1.8 Aims and objectives of the study

1.8.1 Aims of the study

The overall aims of this study were to characterise constitutive expression levels of the LPS recognition and signalling receptors and accessory molecules in two ethnically distinct healthy South African populations (black and white population groups), to provide a detailed description of the genetic variation and LD patterns in the genes (the *LBP-CD14-TLR4-LY96* complex) encoding for these molecules in the black South African population, and to assess the representation of the genetic variants in HIV-1 controllers. We also aimed to investigate the state of microbial translocation in people living with HIV-1 with different clinical phenotypes (HIV-1 controllers and progressors).

1.9 Specific objectives

i. To measure constitutive cell surface expression levels of CD14 and TLR4, and LPS- and LTA-modulated TLR4 expression, on whole blood monocytes and neutrophils, and constitutive plasma levels of sCD14, LBP and MD-2 – to explore the influence of ethnicity, sex and age in healthy HIV-uninfected black and white South Africans.

ii. To describe genetic variation in the coding and non-coding regions of the *LBP*, *CD14*, *TLR4* and *LY96* genes using whole genome sequencing (WGS) data from healthy black South Africans, and to compare to select populations from the 1000 Genomes Project.

iii. To compare the representation of the *LBP*, *CD14*, *TLR4* and *LY96* genetic variants, determined from WGS data from black South African HIV-1 controllers (elite controllers, high viral load long-term non-progressors, and viraemic controllers), to that of the HIV-uninfected black South African controls (described above under objective ii).

iv. To measure and compare plasma levels of LBP, sCD14, MD-2, and REG3α between different groups of HIV-1 controllers (elite controllers, high viral load long-term non-progressors), progressors and healthy HIV-uninfected controls.

CHAPTER 2

Expression levels of LPS recognition and signalling molecules, LBP, CD14, TLR4, and MD-2 in two ethnically divergent South African populations

2.1 Introduction

Lipopolysaccharide (LPS), a component of gram-negative bacterial cell walls, elicits innate immune responses through the activation of the TLR4 receptor (Ulevitch and Tobias, 1995, El-Zayat et al., 2019). The activation of the TLR4 receptor is complex, and involves the use of LPS-binding protein (LBP), cluster of differentiation 14 (CD14), and Lymphocyte antigen 96 (LY96) also known as Myeloid differentiation factor 2 (MD-2) (Raetz and Whitfield, 2002, Lu et al., 2008).

TLR4 is a type 1 transmembrane protein mainly expressed by cells of myeloid origin, such as monocytes, macrophages, and a subset of dendritic cells (Muzio et al., 2000). Lipopolysaccharide recognition and signalling is initiated by the binding of LBP to the lipid A moiety of the LPS (Tobias et al., 1989, Wright et al., 1990). LPS-binding protein is a 60 kDa glycosylated soluble acute phase protein constitutively produced by the hepatic cells of the liver (Schumann et al., 1990, Grube et al., 1994). It catalytically transfers the LPS to CD14, a 55 kDa glycoprotein which exists in two forms, a glycosylphosphatidylinositol-anchored protein (mCD14), and a soluble form (sCD14) derived from cleavage of the mCD14 or direct secretion from the liver (Ulevitch and Tobias, 1995, Su et al., 1999). The membrane-bound CD14 is a surface expressed molecule on monocytes, macrophages, and neutrophils (Su et al., 1999). In non-mCD14 expressing cells, LBP transfers the LPS to sCD14 (Frey et al., 1992). CD14 subsequently transfers LPS to the hydrophobic pocket within the MD-2 protein, which is bound to the extracellular domain of TLR4 (Shimazu et al., 1999, Park et al., 2009). MD-2 is a 20–30 kDa acute phase glycoprotein that is either directly secreted as a soluble active protein by primary cells or binds to the TLR4 in the endoplasmic reticulum and facilitates the specific glycosylation and trafficking of TLR4 to the cell surface (Visintin et al., 2001, Nagai et al., 2002). The binding of LPS to MD-2 in the TLR4-MD-2 complex, results in TLR4 conformational changes that initiate downstream signal transduction resulting in the production of pro-inflammatory cytokines, which activate innate immune responses (Park et al., 2009).

A large number of studies have shown that the expression levels of CD14, LBP, TLR4, and MD-2 associate with the risk and severity of various inflammatory, autoimmune, and infectious diseases and cancers, including asthma, cardiovascular diseases, diabetes, inflammatory bowel syndrome and HIV-1 infection (Nockher et al., 1994, Opal et al., 1999, Sandanger et al., 2009, González-Reyes et al., 2010, Sandler et al., 2011, Liao et al., 2015a, El-Zayat et al., 2019). It is interesting that some of these diseases exhibit population and sex differences (Fairweather and Rose, 2004, Campbell and Tishkoff, 2008, Molodecky et al., 2012, Özdemir and Dotto,

2017, Hertz and Schneider, 2019). Additionally, plasma levels of sCD14 and LBP are often used as indirect markers of microbial translocation, which is defined as the passage of microbes or microbial products such as LPS from the gastrointestinal tract into the circulation system (Marchetti et al., 2013). Microbial translocation has been identified as one of the major causes of persistent inflammation and immune activation in people living with HIV-1 (Brenchley et al., 2006) and has also been implicated in obesity, diabetes, inflammatory bowel diseases (IBD), hepatitis B and C virus infection, fatty liver disease, pancreatitis, and cardiovascular diseases (Creely et al., 2007, Pussinen et al., 2011, Brenchley and Douek, 2012).

Given the importance of the expression levels of these LPS recognition and signalling molecules in the risk and severity of various immune related diseases, a better understanding of the baseline expression levels of these molecules taking ethnicity, sex and age into account is essential. Therefore, this study quantified constitutive expression of cell surface TLR4 and CD14 (mCD14) and plasma levels of soluble CD14, LBP, and MD-2 - in two ethnically divergent South African populations (healthy HIV-1 uninfected black and white individuals), and investigated the influence of sex and age on the expression levels. CD14 also serves as a coreceptor for lipoteichoic acid (LTA) found in gram-positive organisms, however signalling occurs through TLR2 and not TLR4 (Schroder et al., 2003). As TLR4 in the main signalling molecule within the axis under study that leads to LPS-induced responsiveness, a whole blood model was utilized to establish the effect of two stimuli (LPS and LTA) on TLR4 expression levels on monocytes and neutrophils.

2.2 Methods

2.2.1 Study participants

Age- and sex-matched HIV-1 uninfected black (n=17) and white (n=21) South African volunteers were recruited from the National Institute for Communicable Diseases, Johannesburg. The median age of the black and white individuals was 32 (range: 26 - 64 years) and 35 (range: 25 - 59 years), respectively. Black South Africans were comprised of 52.9% females (n=9) and white South Africans were comprised of 52.4% females (n=11). Two white males were excluded from the flow cytometry analysis, as a result of technical issues with sample acquisition.

Written informed consent was obtained from all the participants. Ethical clearance was obtained from the Human Research Ethics Committee at the University of the Witwatersrand, Johannesburg (Certificate number: M190995; Appendix 2).

2.2.2 Quantification of whole blood CD14 and TLR4 expression levels using flow cytometry

Within one hour of collection, sodium heparin-anticoagulated whole blood from participants was (i) stained immediately with fluorochrome-conjugated antibodies for immunophenotyping by flow cytometry - in order to determine constitutive surface expression levels of CD14 and TLR4 on monocytes and neutrophils, and (ii) TLR4 expression on monocytes and neutrophils was quantitated but following a period of LPS and LTA stimulation. For each individual, 1 ml of whole blood was incubated with 1 µg/ml of LPS from *Escherichia coli* (strain O111:B4; Sigma-Aldrich, Saint Louis, Missouri, USA) or 20 µg/ml of LTA from *Staphylococcus aureus* (Sigma-Aldrich, Saint Louis, Missouri, USA). A control tube with no stimulant (unstimulated) was also included. The blood was incubated for 4 hours at 37°C with 5% $C0_2$. After incubation, 100 µl of whole blood from each tube was immediately stained for the flow cytometric analysis of TLR4 expression.

All fluorochrome-conjugated antibodies were obtained from BD Biosciences (San Jose, California, USA). The CD14 flow cytometry panel consisted of two antibodies, CD16 BV510 (3G8) to identify neutrophils and CD14 PE (MφP6) to identify monocytes and for quantification of the number of CD14 molecules bound per cell. The TLR4 panel consisted of three antibodies, CD14 FITC (M5E2) to identify monocytes, CD16 BV510 (MφP6) to identify neutrophils, and TLR4 PE (TF901) for quantification of the number of TLR4 molecules bound per cell.

Briefly, 100 µl of whole blood from each individual was incubated with the appropriate antibodies for each panel at room temperature for 15 minutes in the dark. Red blood cells were lysed for 10 minutes in the dark at room temperature using 2 ml of FACS lysing solution (BD Biosciences, San Jose, California, USA). Samples were centrifuged at 2000 rpm for 5 minutes at room temperature and the supernatant was discarded. Samples were then washed with 3 ml of FACSFlow™ (BD Biosciences, San Jose, California, USA), and centrifuged at 2000 rpm for 5 minutes at room temperature. The supernatant was discarded, and the cells were resuspended in 250 µl of FACSFlow™ for acquisition on the flow cytometer. BD™ Comp Beads anti-Mouse Ig, κ and BD™ Comp Beads negative control (BD Biosciences, San Jose,

California, USA) were stained with each monoclonal antibody for 15 minutes in the dark at room temperature, washed twice with 2 ml FACSFlow™, and resuspended in 250 µl FACSFlow™ for acquisition on the flow cytometer.

Flow cytometry was performed using the four-laser BD LSRFortessa™ flow cytometer (BD Biosciences, San Jose, California, USA). Instrument setup was performed daily before acquisition. BD™ CS&T Beads (BD Biosciences, San Jose, California, USA) were run on the instrument for quality control, to characterize, track and report performance measurements of the instrument. Rainbow Calibration Particles (BD Biosciences, San Jose, California, USA) were then used for standardization. On each run, Rainbow Calibration Particles were used to set voltages to ensure that results were reproducible and to account for variation on the instrument. Compensation controls tubes were acquired after setting up voltages, then compensation was calculated and applied before the sample tubes were acquired.

A minimum of 9000 monocytes were acquired for each sample and data was analysed using FlowJo version 9.8.1 (Tree Star, San Carlos, California, USA). The gating strategy is shown in **Figure 2.1**. Results were reported as geometric mean (GM) for both CD14 and TLR4 expression.

Quantification of CD14 and TLR4 molecules bound per cell on monocytes and neutrophils was carried out using the commercially available Quantibrite™ system (BD Biosciences, San Jose, California, USA). The Quantibrite™ system consists of a set of four pre-calibrated beads, which are used to generate a standard curve that is used for the calculation of antibodies bound per cell (ABC). The use of the Quantibrite™ system, together with PE-conjugated monoclonal antibodies specific to CD14 or TLR4, allowed for the geometric mean (GM) to be converted into the number of CD14 or TLR4 molecules per cell (CD14 or TLR4 density).

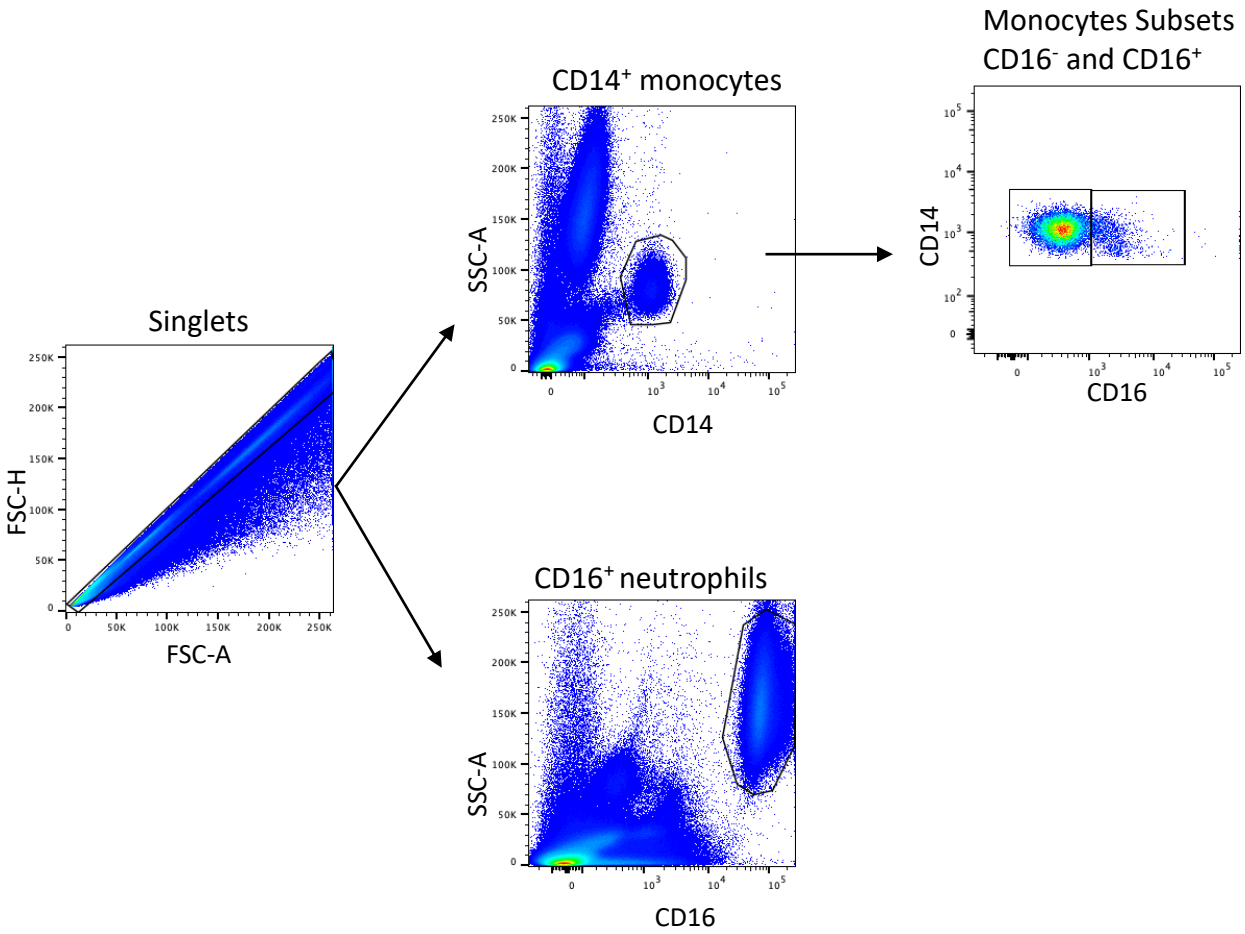


Figure 2.1: Gating strategy for identifying monocytes and neutrophils in whole blood. Doublets were excluded from the analysis by gating on singlets (FSC-H vs FSC-A). Monocytes for both panels were gated on $CD14^+$ cells and SSC-A. Monocytes subsets were gated on $CD14^+$ and $CD16^+$ and $CD16^-$ cells. Neutrophils were gated on $CD16^+$ cells and SSC-A from singlets.

2.2.3 Quantification of sCD14, LBP, and MD-2 plasma levels

Commercially available quantitative enzyme-linked immunosorbent assay (ELISA) kits were used to measure EDTA-anticoagulated plasma levels of sCD14 (R&D Systems, Minneapolis, Minnesota, USA), LBP (Hycult Biotech, Uden, the Netherlands), and MD-2 (Sigma-Aldrich, Saint Louis, Missouri, USA) as per the manufacturer's instructions. The limits of detection for the LBP, sCD14, and MD-2 ELISA kits were 4.4 ng/ml, 250 pg/ml, and 1.6 ng/ml, respectively. Samples were assayed in duplicate and results averaged. The Versamax microplate reader (Molecular Devices, San Jose, California, USA) was used to measure optical density (OD) at an absorbance of 450 nm and 540 nm, and the Softmax Pro 6.5 software (Molecular Devices, San Jose, California, USA) was used for analysis.

2.2.4 Statistical analysis

GraphPad Prism version 8.01 for Windows (GraphPad Software, La Jolla, California, USA) and Statistica™ software (TIBCO Software, Palo Alto, California, USA) were used for statistical analyses. All statistical tests were 2-tailed, with p values <0.05 considered statistically significant. The Mann-Whitney non-parametric U test was used to compare the expression levels of TLR4, CD14, LBP, and MD-2 between groups (i.e black and white South Africans and between males and females). Comparisons of expression levels between cell subsets and the TLR4 expression levels between unstimulated, LPS-stimulated, and LTA-stimulated whole blood monocytes and neutrophils were performed using the Wilcoxon signed-rank test. The spearman correlation coefficient was used to assess the correlation between age and expression levels of TLR4, CD14, LBP, and MD-2, as well as correlations between the expression levels of TLR4, CD14, LBP, and MD-2.

2.3 Results

2.3.1 CD14 and TLR4 expression on monocytes and neutrophils

Expression levels of CD14 and TLR4 on monocytes and neutrophils were measured as molecules bound per cell (density) in healthy individuals. Monocytes ($CD14^+$) were divided into two subsets namely, classical ($CD14^+CD16^-$) and intermediate ($CD14^+CD16^+$) monocytes.

CD14 expression on classical ($CD14^+CD16^-$) monocytes was significantly higher than the expression on the intermediate ($CD14^+CD16^+$) monocytes (p=0.044, **Figure 2.2A**) in black South Africans. In white South Africans, although the same relationship was present, it did not reach significance (p=0.066**, Figure 2.2A**). The CD14 expression on classical ($CD14^+CD16^-$) monocytes was also significantly higher than the expression on the intermediate ($CD14^+CD16^+$) monocytes (p=0.007, data not shown) when the black and white South African populations were combined (total group). As expected CD14 expression was significantly lower on neutrophils (approximately 100-fold lower) when compared to total monocytes, and the monocyte subsets for both populations and the total group (p<0.0001, for all comparisons). CD14 expression on total monocytes, the monocyte subsets, and neutrophils did not differ significantly between black and white South Africans (**Figure 2.2A**).

TLR4 expression on classical ($CD14^+CD16^-$) monocytes was significantly lower than the expression on the intermediate ($CD14^+CD16^+$) monocytes for both the black and white South Africans (**Figure 2.2B**) and the total group (p<0.0001, for all comparisons). Unlike the stark

cell type-specific differences in membrane CD14 expression, the range of the TLR4 expression levels (density) was similar on neutrophils and monocytes. However, TLR4 expression on total monocytes was significantly higher compared to neutrophils in the black South Africans (p=0.040), with a trend observed in the white South Africans (p=0.066); the populations combined strengthened this observation (p=0.004). When TLR4 expression levels on $CD14^{+}CD16^{-}$ monocyte subsets were compared to that of neutrophils, both the black (p=0.078) and white (p=0.096) individuals had trends of higher TLR4 expression on $CD14^{+}CD16^{-}$ monocytes compared to neutrophils; when population groups were combined, the difference was significant (p=0.013). Significantly higher TLR4 expression on both $CD14^{+}CD16^{+}$ subsets of monocytes were also observed in both the black and white South Africans (p=0.003 and p<0.0001, respectively). TLR4 expression levels on total monocytes, both monocyte subsets, and neutrophils were not significantly different between black South Africans and white South Africans (**Figure 2.2B**).

2.3.2 Plasma levels of LBP, sCD14, and MD-2

Black South Africans had significantly higher plasma levels of LBP (p<0.0001, **Figure 2.3A**) and a trend to lower levels of sCD14 (p=0.095, **Figure 2.3B**) compared to white South Africans. Just over half (n=21, 55.3 %) of healthy South Africans had detectable levels of MD-2, with the number of individuals with undetectable levels not significantly different between the two population groups. One white individual with plasma MD-2 of 1 352 ng/ml (range 0-209 ng/ml) was considered an outlier and excluded from further analysis. No significant differences were seen when plasma levels of MD-2 were compared between black and white South Africans (**Figure 2.3C**), even when individuals with undetectable levels were excluded from analysis (data not shown).

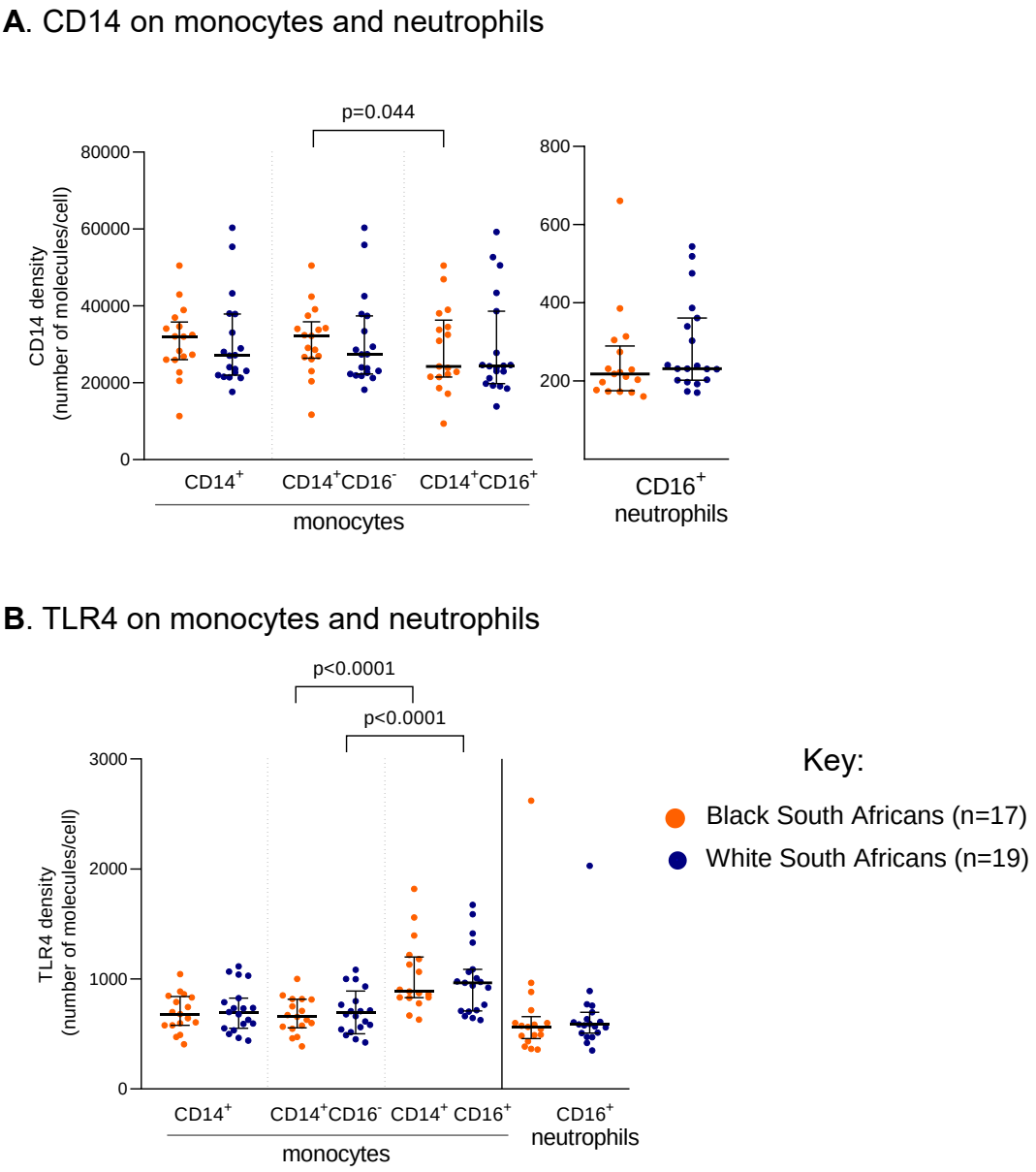


Figure 2.2: Comparisons of CD14 and TLR4 density on monocytes and neutrophils from HIV-uninfected black (n=17) and white (n=19) South Africans. (**A**) CD14 expression on monocytes and neutrophils, and (**B**) TLR4 expression on monocytes and neutrophils. Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges. The Mann-Whitney U test was used for comparison.

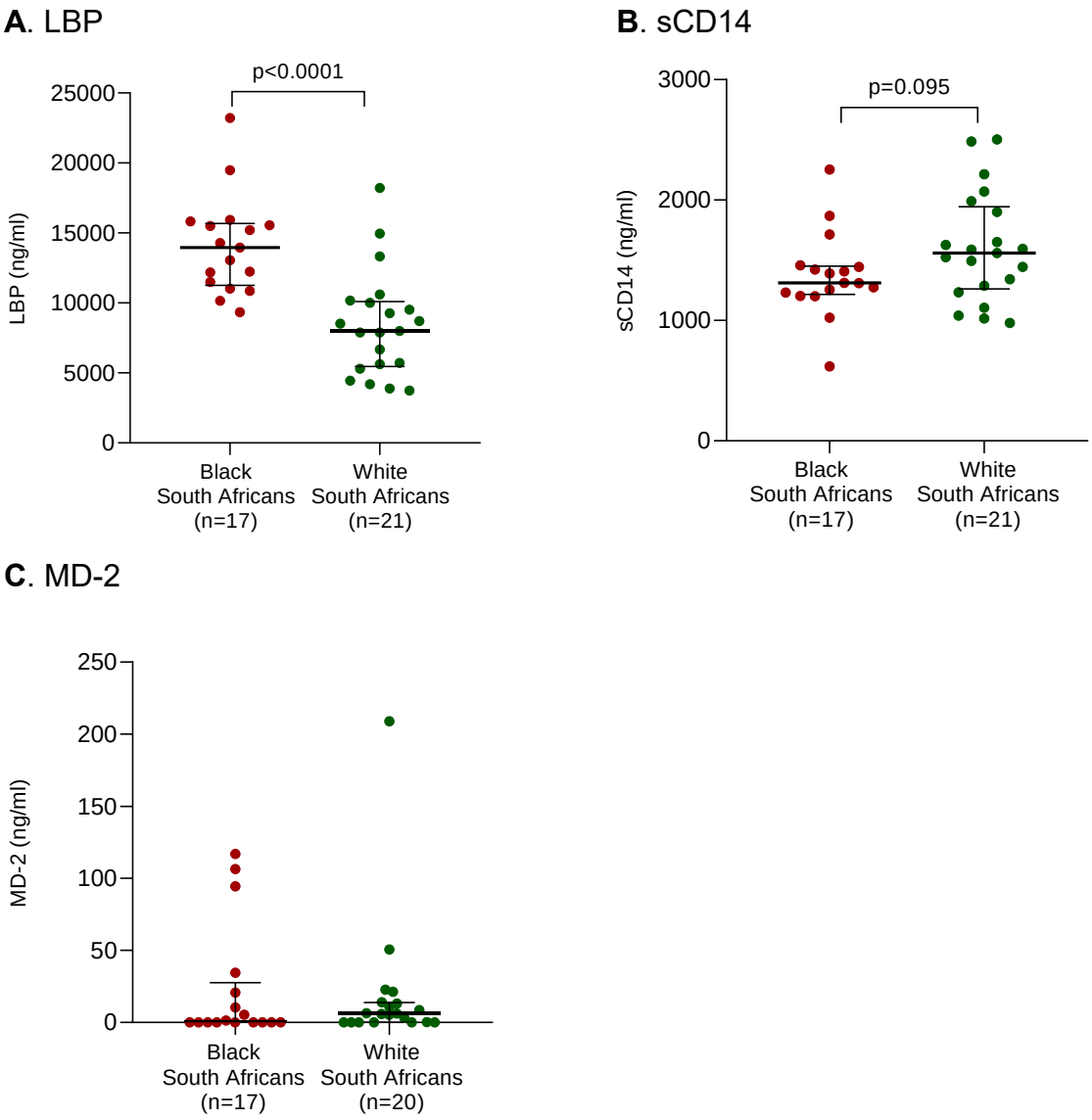


Figure 2.3: Comparison of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 plasma levels between healthy black South Africans (n=17) and white South African (n=21). Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges. The Mann-Whitney U test was used for comparison.

2.3.3 Effect of sex on the expression levels of CD14 and TLR4 and plasma levels of LBP, sCD14, and MD-2

Individuals were stratified according to sex and expression levels of the CD14 and TLR4 receptors and plasma levels of LBP, sCD14, and MD-2 compared in the total group of black and white individuals combined, and in each group separately. No significant sex differences were observed in CD14 density on monocytes and neutrophils for any of the comparisons (**Figure 2.4A i, A ii, and A iii**).

TLR4 expression on total monocytes (p=0.036) and $CD14^+CD16^+$ (p=0.033) and $CD14^+CD16^-$ (p=0.020) monocyte subsets was significantly higher in females compared to males (**Figure 2.4B i**) for the total group. Upon stratification into population groups, it was evident that this relationship was due to the white South African females, which had significantly higher TLR4 expression on their total monocytes (p=0.016) and $CD14^+CD16^+$ (p=0.016) and $CD14^+CD16^-$ (p=0.009) monocyte subsets compared to males (**Figure 2.4B ii and iii**). No sex differences in TLR4 expression were evident on neutrophils (**Figure 2.4i – 2.4B iii**).

Comparison of LBP, sCD14 and MD-2 plasma levels between the total group of males and females revealed no sex differences (data not shown). Stratification according to population group also showed no sex differences in plasma levels of LBP, sCD14 and MD-2 (**Figure 2.5**). It is noteworthy that the LBP plasma levels were significantly higher in black females compared to white females (p<0.0001), but black males compared to white males only showed a trend of higher LBP, suggesting that females are mostly responsible for the population differences in LBP plasma levels.

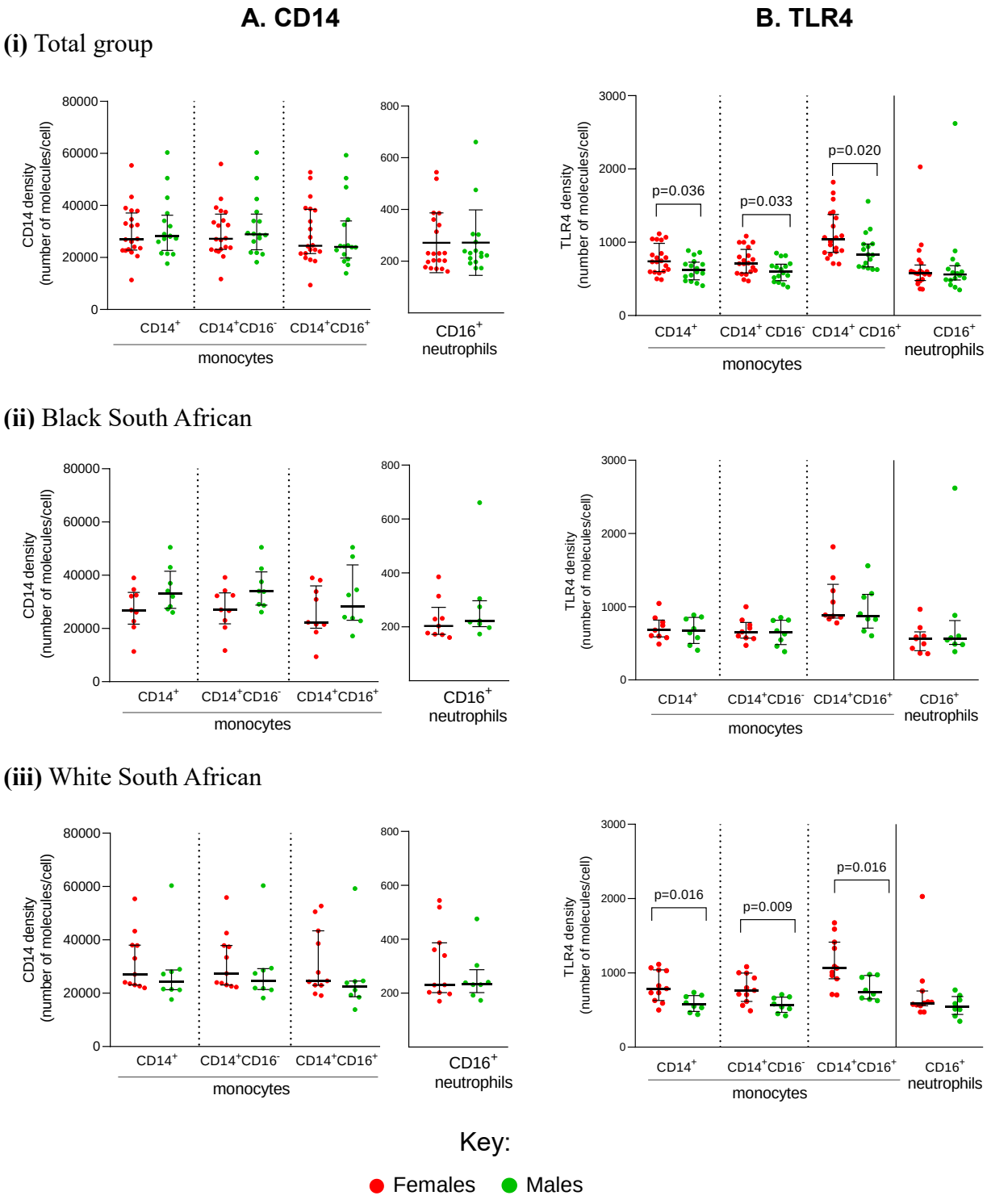


Figure 2.4: Comparisons of (**A**) CD14 and (**B**) TLR4 expression on $CD14^+$ monocytes and $CD16^+$ neutrophils between females and males. (**i**) Total South African females (n=19) and total South African males (n=16), (**ii**) black South African females (n=9) and males (n=8) (**iii**) white South African females (n=11) and males (n=8). Each dot represents a single individual and the horizontal lines represent the medians and interquartile ranges. The Mann-Whitney U test was used for comparisons.

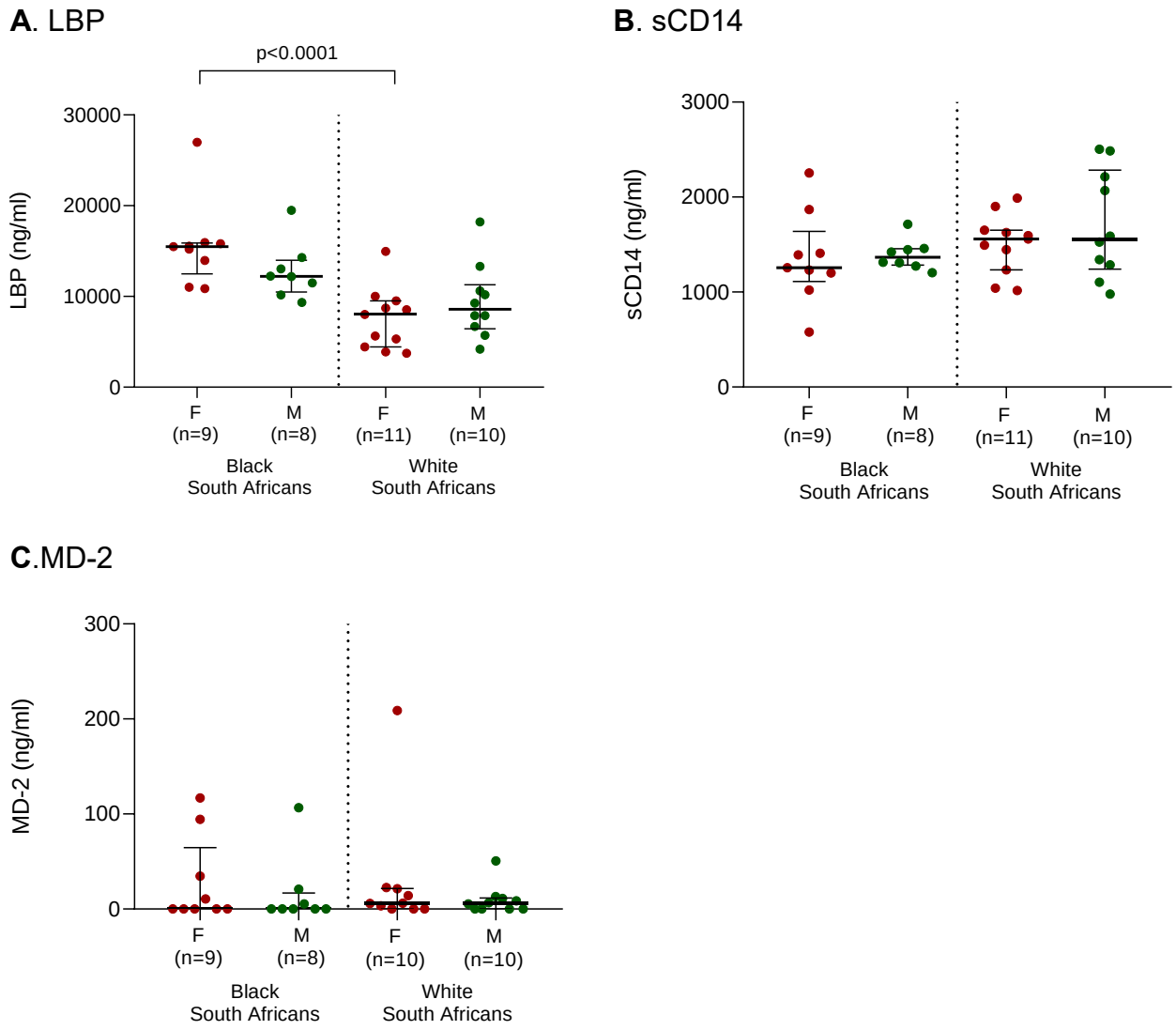


Figure 2.5: Comparisons of plasma levels of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 between females and males among black and white South Africans. Each dot represents a single individual and the horizontal lines represent the medians and interquartile ranges. The Mann-Whitney U test was used for comparisons.

2.3.4 Effect of age on the expression levels of CD14, TLR4 and plasma levels of LBP, sCD14, and MD-2

Age has been shown to affect the expression levels of various genes and proteins. Our participants ranged in age from 25 to 64 years. **Figure 2.6** shows a heatmap of the correlations between age and constitutive CD14 and TLR4 expression, and the plasma levels of LBP, sCD14, and MD-2.

Only a few significant correlations were observed, clearly driven by one or other sex in the specific population groups. In black South African males, the expression of CD14 on $CD14^{+}CD16^{+}$ subsets of monocytes (r=-0.759, p=0.029) and neutrophils (r=-0.759, p=0.029),

as well as the TLR4 expression on neutrophils negatively correlated with age (r=0.795, p=0.018). In white females, CD14 expression on total monocytes (r= -0.7, p=0.016), $CD14^{+}CD16^{-}$ (r=0.7, p=0.016) and $CD14^{+}CD16^{+}$ (r= -0.627, p=0.039) subsets of monocytes, and neutrophils (r=-0.718, p=0.013) negatively correlated with age. This was strikingly different from the positive (non-significant) correlations observed with age in white males. No significant correlations between the plasma levels of LBP, CD14, and MD-2 with age were observed for any of the comparisons.

2.3.5 Correlation between constitutive CD14 and TLR4 expression, and LBP, sCD14, and MD-2 plasma levels

The relationship between cell surface expression of CD14 and TLR4, and the plasma levels of LBP, sCD14, and MD-2 was investigated. **Figure 2.7** shows the results depicted by a heatmap.

First analysed was the expression of CD14 and TLR4 on monocytes versus neutrophils. CD14 expression on total monocytes positively correlated with CD14 expression on neutrophils in the total group of our healthy controls (r=0.421, p=0.011). This significant correlation could be attributed to the black population group since it was stronger in black South Africans (r=0.662, p=0.004) and was not significant in white South Africans. The significant positive correlation between the CD14 expression on total monocytes and neutrophils was maintained when participants were analysed according to sex in total females (r=0.531, p=0.016) and according to both sex and ethnicity in black males (r=0.762, p=0.028). Similarly, the expression of TLR4 on total monocytes also positively correlated with TLR4 expression on neutrophils in the total group (r=0.512, p=0.001), total females (r= 0.456, p=0.043) and total males (r=0.594, p=0.015).

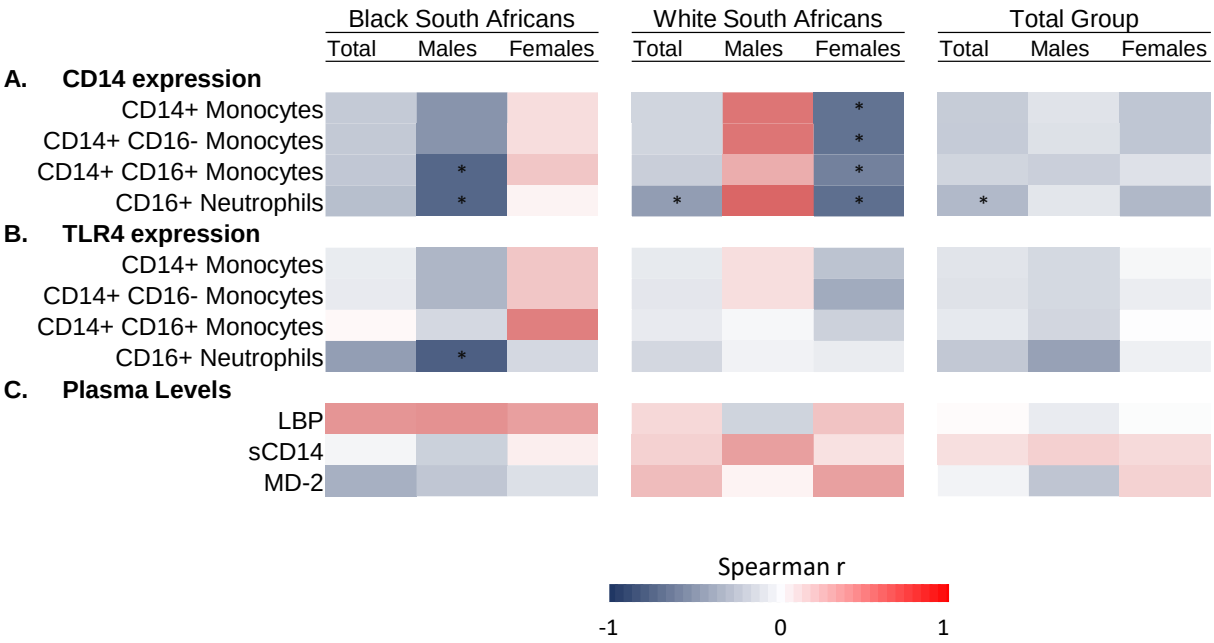


Figure 2.6: Heatmap showing correlations (Spearman's rank) between age and the expression of (**A**) CD14 and (**B**) TLR4 on $CD14^+$ monocytes, the monocyte subsets ($CD14^+CD16^-$; $CD14^+CD16^+$), and $CD16^+$ neutrophils, and (**C**) plasma levels of LBP, sCD14 and MD-2. The colours represent the direction and strength of the correlation with blue representing a negative correlation, red representing a positive correlation, and white representing a lack of correlation (*p<0.05).

Next, the relationship between CD14 and TLR4 expression was analysed. CD14 expression on total monocytes positively correlated with that of TLR4 in the total group (r=0.404, p=0.014), the total female group (r=0.523, p=0.018), and black South Africans (r=0.512, p=0.036). A strong positive correlation was also observed between the CD14 and TLR4 expression on neutrophils in the total group (r=0.624, p=0.00005), total female group (r=0.577, p=0.008), total male group (r=0.735, p=0.001), white South Africans (r=0.560, p=0.013), black South Africans (r=0.644, p=0.005), and black South African males (r=0.929, p=0.0009). Additionally, CD14 expression on neutrophils also correlated positively with TLR4 expression on monocytes ($CD14^+$) in the total group (r=0.335, p=0.046) and total males (r=0.618, p=0.011), which was driven by the strong correlation observed with the black males (r=0.810, p=0.015).

In white South Africans females, LBP correlated positively with sCD14 (r=0.636, p=0.035), but a similar correlation was absent in white males. Black South African males, however, had a significant negative correlation between LBP and sCD14 (r=-0.738, p=0.037). No significant

correlations were observed between the plasma levels of MD-2 with the plasma levels of LBP or sCD14.

Lastly, the relationships between TLR4 and CD14 expression on monocytes and neutrophils and plasma levels of sCD14, LBP, and MD-2 was investigated. There were no significant correlations between the surface expression of CD14 and TLR4 with plasma levels of sCD14, LBP, and MD-2 in either the total participant group combined or when stratified by population group. However, when stratified according to both sex and ethnicity, in black females, TLR4 expression on total monocytes positively correlated with LBP plasma levels (r=0.750, p=0.020). CD14 density on monocytes correlated positively with MD-2 plasma levels in black South African males (r=0.736, p=0.038). In white South African males, TLR4 expression on total monocytes (r=0.714, p=0.047) and CD14 expression on neutrophils (r=0.833, p=0.01) positively correlated with plasma levels of sCD14.

2.3.6 Evaluating the effect of LPS and LTA stimulation on TLR4 expression

To determine whether TLR4 expression on monocytes and neutrophils is differentially regulated in black and white South Africans post stimulation, whole blood was stimulated with LPS and LTA, and TLR4 expression compared between stimulated and unstimulated conditions.

In black South Africans, compared to unstimulated controls, LPS stimulation significantly downregulated TLR4 (**Figure 2.8 A i**) on total monocytes (p=0.031), $CD14^{+}CD16^{-}$ (p=0.035) and $CD14^{+}CD16^{+}$ (p=0.017) monocyte subsets, and on neutrophils (p=0.036). In comparison, LTA stimulation significantly downregulated TLR4 on $CD14^{+}CD16^{+}$ monocytes (p=0.039, **Figure 2.8A i**) but significantly upregulated TLR4 on neutrophils (p=0.002, **Figure 2.8A** i) when compared to unstimulated controls. Compared to LPS stimulation, LTA stimulation resulted in significantly higher TLR4 expression on neutrophils (p=0.002, **Figure 2.8A i**), but not on monocytes.

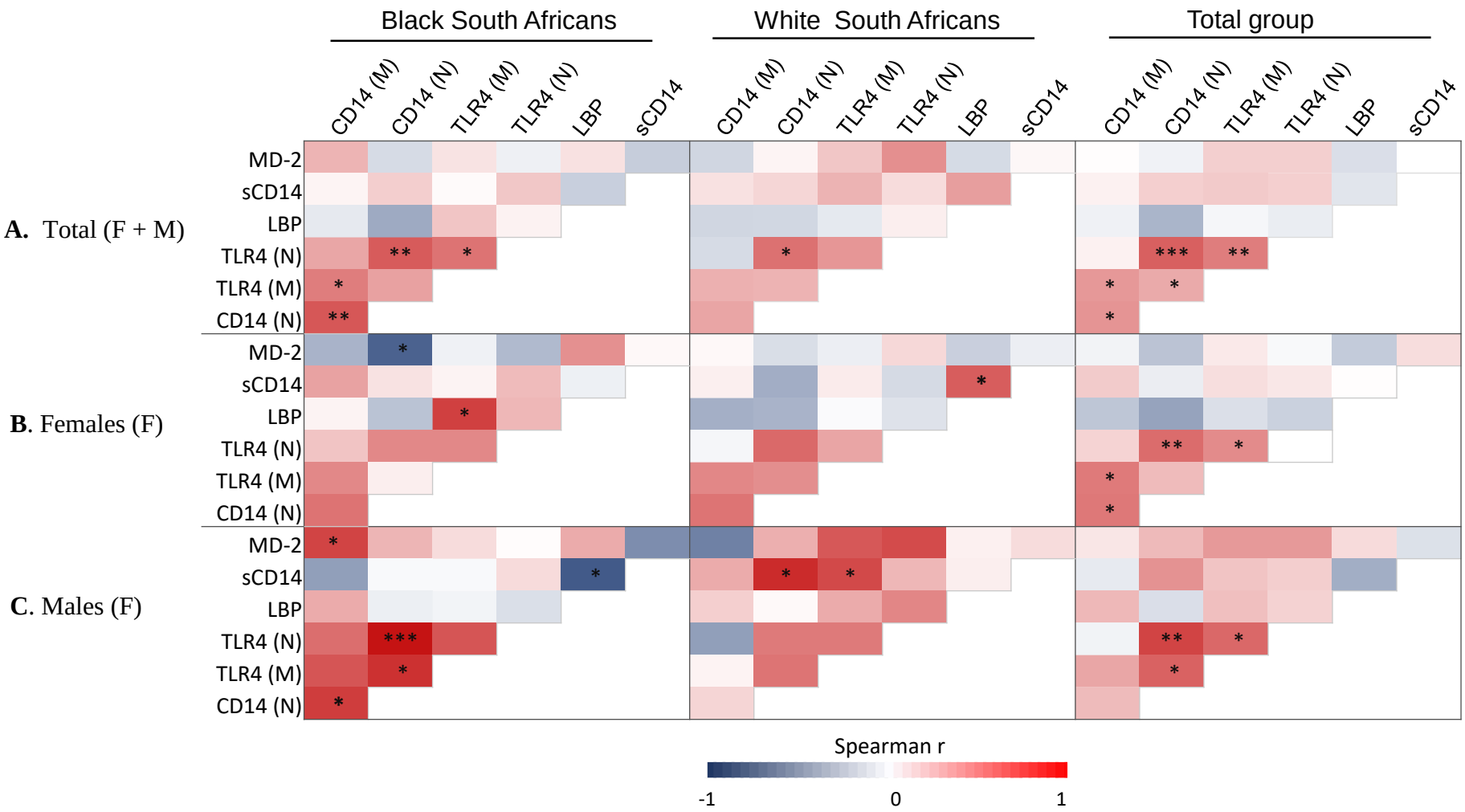


Figure 2.7: Heatmap showing correlations between expression levels of TLR4, mCD14, sCD14, LBP, and MD-2 in black South Africans, white South Africans and the combined total group of our study participants. (**A**) Total black South Africans (n=17), total white South Africans (n=21), and the total group of our study participants (n=38). (**B**) Black females (n=9), white females (n=11) and total females (n=20). (**C**) Black males (n=8), white males (n=10), and total males (n=18). CD14 (M): CD14 expression on $CD14^+$ monocytes, CD14 (N): CD14 expression on $CD16^+$ neutrophils, TLR4 (M): TLR4 expression on $CD14^+$ monocytes and TLR4 (N): TLR4 expression on $CD16^+$ neutrophils. The colours represent the direction and strength of the correlation with blue representing a negative correlation, red representing a positive correlation and white representing a lack of correlation (*p<0.05, **p<0.01, ***p<0.001).

In contrast, in white South Africans (**Figure 2.8B i**), compared with unstimulated controls, LPS stimulation significantly upregulated TLR4 on total monocytes (p=0.003), $CD14^+CD16^-$ (p=0.0.03) and $CD14^+CD16^+$ (p=0.010) monocyte subsets, and on neutrophils (p=0.005). Similarly, LTA stimulation (**Figure 2.8B i**) significantly upregulated TLR4 on total monocytes (p<0.0001), $CD14^+CD16^-$ (p<0.0001) and $CD14^+CD16^+$ (p=0.003) monocyte subsets, and on neutrophils (p=0.008). LTA stimulation compared to LPS stimulation resulted in significantly higher TLR4 expression on total monocytes (p=0.026) and the $CD14^+CD16^-$ monocyte subset (p=0.016), but not on the $CD14^+CD16^+$ monocyte subset or neutrophils (**Figure 2.8B i**).

When the black and white individuals were stratified by sex, it was found that in black South Africans the significant downregulation of TLR4 by LPS stimulation compared to unstimulated controls on total monocytes, monocytes subsets, and neutrophils was lost in both females (**Figure 2.8A ii**) and males (**Figure 2.8A iii**). The significant downregulation of TLR4 density in $CD14^+CD16^+$ monocytes in LTA stimulated compared to unstimulated controls was maintained in females (p=0.039, **Figure 2.8A ii**), as was the significant upregulation of TLR4 on neutrophils (p=0.012, **Figure 2.8A ii**) compared to LPS stimulation.

In white South Africans, we found that in females (**Figure 2.8B ii**), the significant upregulation of TLR4 by LPS stimulation was maintained in total monocytes (p=0.024), $CD14^+CD16^-$ monocyte subsets (p=0.019), and neutrophils (p=0.019). However, in males LPS stimulation only resulted in TLR4 upregulation on neutrophils (p=0.02, **Figure 2.8B iii**). Similar to the total group of white South Africans, in females LTA stimulation, resulted in the upregulation of TLR4 on total monocytes (p=0.001), $CD14^+CD16^-$ (p=0.001) and $CD14^+CD16^+$ (p=0.037) monocyte subsets, and neutrophils (p=0.003, **Figure 2.8B ii**). In males, LTA stimulation resulted in significant upregulation of TLR4 only on total monocytes (p=0.008) and $CD14^+CD16^-$ monocyte subsets (p=0.008) (**Figure 2.8B ii and 2.8B iii**).

Overall, the patterns of responsiveness (unstimulated, LPS-stimulated, LTA-stimulated) were influenced by cell type (monocytes vs neutrophils), and generally shared between females and males within each population group, and significance was attained in the total population groups because of larger sample numbers when females and males were combined. It was particularly noteworthy that TLR4 is mainly downregulated in monocytes and neutrophils of black individuals when whole blood was LPS-stimulated, whereas the opposite occurred in white individuals.

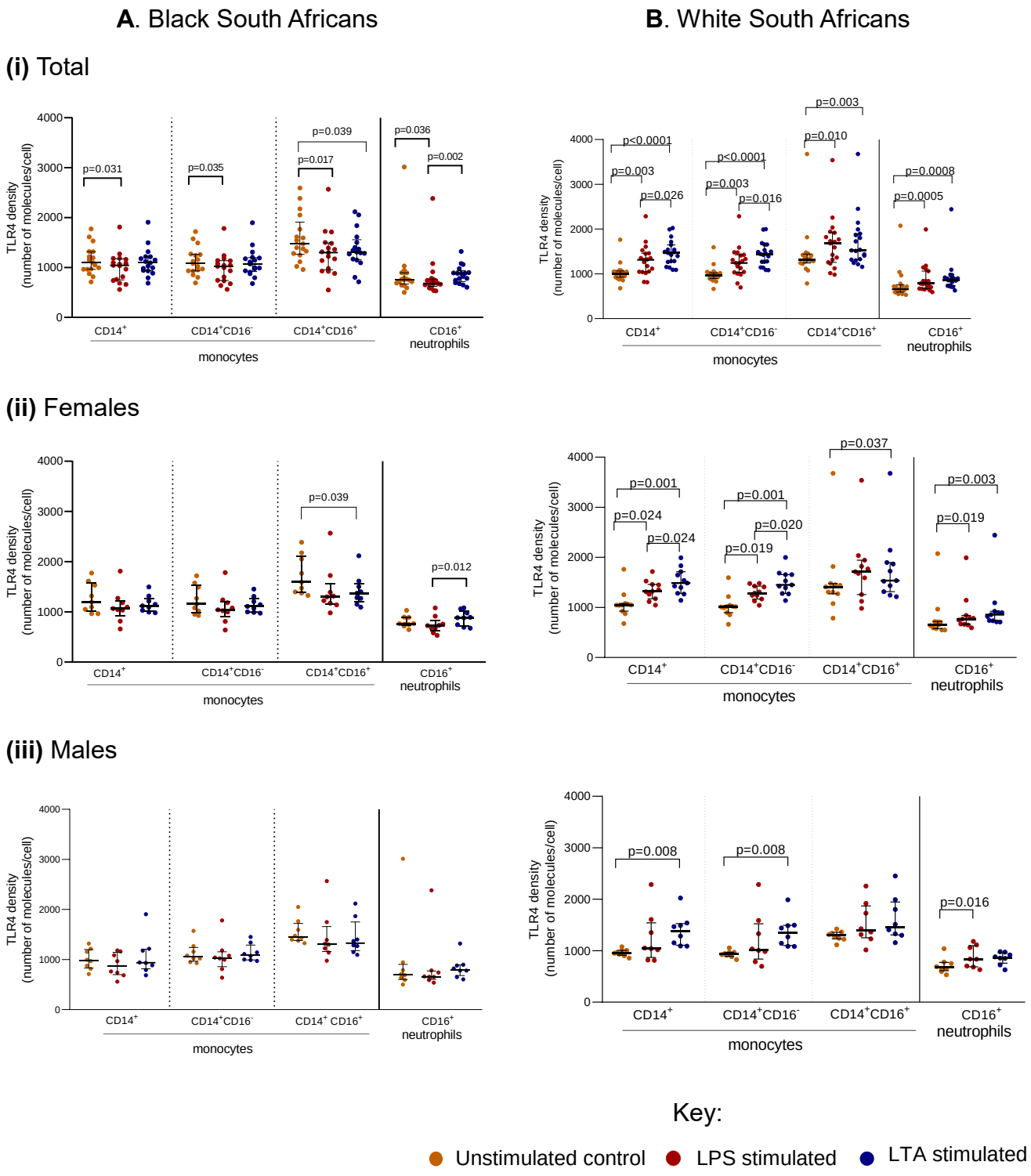


Figure 2.8: Comparisons of TLR4 expression between unstimulated, LPS- and LTA-stimulated monocytes and neutrophils in whole blood. (**A**) (i) black South Africans (n=17), (**ii**) black South African females (n=9), (**iii**) black South African males (n=8), (**B**) (**i**) white South Africans (n=19), (**ii**) white South African females (n=11), and (**iii**) white South African males (n=8). Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges. The Wilcoxon signed-rank test used for comparisons. Significant p-values ($p<0.05$) are indicated.

2.4 Discussion

Expression levels of the receptors (CD14 and TLR4) and soluble accessory molecules (sCD14, LBP, and MD-2) required for the recognition and response to LPS, are altered in various immune-related diseases (Nockher et al., 1994, Opal et al., 1999, Sandanger et al., 2009, González-Reyes et al., 2010, Sandler et al., 2011, Liao et al., 2015a). A range of these immune related diseases such as HIV-1, diabetes, and inflammatory bowel disease show sex and population differences (Fairweather and Rose, 2004, Campbell and Tishkoff, 2008, Molodecky et al., 2012, Özdemir and Dotto, 2017, Hertz and Schneider, 2019). Given that the distribution of genetic polymorphisms differ between different populations, with African populations being more genetically diverse (Tishkoff et al., 2009), different variants in the *CD14*, *LBP*, *TLR4*, and *MD-2* might be regulating the expression levels of these genes in different population groups. In addition, differences in genetic variation and expression levels of various genes have been previously described in ethnically divergent South African populations (Paximadis et al., 2021, Picton et al., 2010, Picton et al., 2017, Picton et al., 2012). Therefore, understanding of the baseline expression levels of the LPS recognition and signalling molecules in different sexes and ethnically divergent South African populations is essential.

To address the influence of population, sex and age on molecules under study in this thesis - the constitutive monocyte and neutrophil expression levels of TLR4, CD14 (mCD14) and plasma levels of sCD14, LBP, and MD-2 were determined in healthy HIV-uninfected individuals from the black and white South African populations. In addition, we investigated the effect of LPS and LTA stimulation on TLR4 expression levels on monocytes and neutrophils in these populations. The latter questioned if TLR4 expression might be differentially regulated according to population group and sex. As TLR4 is the main signalling molecule in the immune axis that engages with LPS, this component of the study was confined to TLR4 only. LTA was included to assess responsiveness compared to LPS.

A limitation, here, is that the expression of CD14 was not quantified in the same way as TLR4 expression was quantified. This could have provided some additional insights into CD14 as a functional cofactor upon LPS/LTA stimulation. CD14 expression could not be quantified using the Quantibrite^{TM} PE beads as CD14 was conjugated to FITC in the TLR4 panel. In addition, because the staining protocol used CD14 to gate monocytes without the inclusion of another cell marker, such as HLA-DR (which stains all monocytes, including CD14- monocytes, brightly), further delineation of monocytes subsets beyond only classical and intermediate subsets was not possible (Abeles et al., 2012).

Findings demonstrated the following: (i) There were no significant population differences in the constitutive expression levels of CD14 (mCD14 or sCD14), TLR4 and MD-2. However, LBP plasma levels were significantly higher in black compared to white South Africans. (ii) TLR4 showed population-specific sex differences in constitutive expression levels. (iii) There were population and sex-dependent correlations between age and the constitutive expression levels of mCD14 and TLR4. (iv) LPS and LTA stimulation differentially affected TLR4 expression in whole blood monocytes and neutrophils of black and white South Africans.

LBP is an acute-phase protein constitutively expressed in relatively low levels under normal physiological conditions (Schumann et al., 1990, Grube et al., 1994) and plasma levels increase drastically in response to microbial translocation, and in the presence of infections and inflammation (Grube et al., 1994, Schumann and Zweigner, 1999). In the presence of LPS, LBP has been shown to have concentration-dependent dual roles, which affect the activation of the TLR4 receptor. In low concentrations, LBP binds and catalytically transfers LPS to CD14 to initiate inflammatory cytokine responses (Zweigner et al., 2001). In high concentrations, however, LBP can down-regulate immune responses to LPS by transferring the LPS or LPS/CD14 complexes to lipoproteins for neutralisation, thereby inhibiting the activation of the TLR4 receptor (Vesy et al., 2000, Zweigner et al., 2001). In our study, plasma levels of LBP were found to be higher in healthy black when compared to white South Africans. To our knowledge, no previous studies have compared the baseline expression levels of LBP between black and white individuals.

Given that LBP is an indirect marker of microbial translocation, higher plasma levels of LBP in black South Africans could suggest higher levels of microbial translocation in healthy black compared to white South Africans. In the presence of microbial translocation in people living with HIV-1, several studies have found a positive correlation between sCD14 and LBP plasma levels (Brenchley et al., 2006, Abad-Fernández et al., 2013). Such a relationship was not observed in our HIV-uninfected black South Africans. However, a positive correlation was detected in white South African females, consistent with what has been described in healthy individuals from Spain irrespective of sex (Gonzalez-Quintela et al., 2013). In contrast, LBP negatively correlated with sCD14 in black South African males, which was unexpected. These findings suggest that the higher LBP plasma levels in black compared to white South Africans could be driven in part by other factors such as genetic differences, and/or environmental factors that cause inflammation unrelated to microbial translocation. Of note, black individuals

have been shown to have high levels of systemic inflammation markers (IL-6, CRP, E-selectin and Fibrinogen) compared to white individuals (Stepanikova et al., 2017).

In microbial translocation studies, sCD14 is used as a marker of LPS-induced monocyte activation (Marchetti et al., 2013). Soluble CD14 levels correlate with LPS and markers of inflammation in people living with HIV-1 (Brenchley et al., 2006, Sandler et al., 2011, Abad-Fernández et al., 2013). However, in Belgium, a study found that sCD14 correlated with LPS in white but not in African individuals living with HIV-1 and was thus not a good marker of microbial translocation in African individuals (De Voeght et al., 2016). Although not significant, sCD14 plasma levels were higher (a trend) in white compared to black South Africans. This is consistent with findings from other studies, which also found that sCD14 levels were lower in black individuals compared to white individuals irrespective of disease status (Reiner et al., 2013, De Voeght et al., 2016, Olson et al., 2020). Similar to LBP, in high concentrations, sCD14 can down-regulate immune responses to LPS, through the binding and transfer of LPS to plasma lipoproteins (Kitchens et al., 2001). Interestingly, high sCD14 levels are a risk factor for stroke and coronary heart disease in black but not white individuals (Olson et al., 2020), suggesting that these populations may have masked differences in the regulation and/or function of this molecule, which needs to be better understood.

There are marked differences in innate and adaptive immune responses in males and females, which result in differences in susceptibility to inflammatory diseases, infectious diseases, and cancers (Klein and Flanagan, 2016). Compared to males, females have stronger innate and adaptive immune responses that ultimately result in better clearance of pathogens. This enhanced immune function, however, contributes to a higher susceptibility to autoimmune and inflammatory diseases (Schröder et al., 1998, Klein and Flanagan, 2016). Compared to women, men have been shown to have a higher susceptibility to infections and sepsis and subsequent morbidity and mortality (Schröder et al., 1998, Offner et al., 1999). Additionally, sex differences have been associated with differential constitutive expression levels of select genes in various tissues (Oliva et al., 2020), which potentially affect immune responses to foreign and self-antigens (Klein and Flanagan, 2016).

Results from our study showed that white South African females have higher TLR4 expression on total monocytes and monocyte subsets compared to white males, accounting for the higher TLR4 expression in total (black and white) females compared to total males. This is in keeping with higher expression levels of TLR4 found in whole blood in females (n=67) compared to

males (n=120) reported in the GTEx (The Genotype Tissue Expression project) portal (https://www.gtexportal.org/home/). Although sex differences were not observed in TLR4 expression in monocytes, a previous study found higher TLR4 expression on neutrophils in males compared with females (Aomatsu et al., 2013). Beyond its role in recognition of pathogen-associated molecular patterns (PAMPs) such as LPS, the TLR4 receptor has been shown to play a role in tissue injury and repair (Yu et al., 2010). Activation of the TLR4 receptor by damage-associated molecular pattern molecules (DAMPs) such as DNA-binding protein high-mobility group box 1 (HMGB1) released by necrotic cells and injured tissues results in sterile inflammation (Yu et al., 2010, Molteni et al., 2016). This sex difference in the expression levels of TLR4 in white but not black South Africans in our study suggests that there could be population specific sex differences in TLR4 functionality in vivo. Our findings using the whole blood stimulation model to evaluate TLR4 expression in response to LPS/LTA supports this likelihood.

We did not find significant sex differences in sCD14 plasma levels in either population group. This is contrary to other studies, which have shown higher levels of sCD14 in females compared to males in American-based populations comprised of African Americans, Caucasians and Hispanics (Reiner et al., 2013, Olson et al., 2020, Stanislawski et al., 2021). Additionally, contrary to the findings of Gonzalez-Quintela et al. (2013), in which higher LBP plasma levels were associated (trend, not significant) with males in Spanish individuals, we found no significant sex differences in LBP plasma levels.

In addition to age being a risk factor for various diseases such as cancer, cardiovascular disease, type 2 diabetes, and infectious diseases such as COVID-19, the expression levels of various genes in different tissues vary with age (Glass et al., 2013, Peters et al., 2015, Yang et al., 2015, Henkens et al., 2022). Here, the expression of CD14 on monocytes and neutrophils showed population- and sex-specific correlations with age. In white South Africans, CD14 expression on monocytes and neutrophils decreased with age in females, and although not significant, the relationship was the opposite in males. Similar to white South African females, CD14 expression on non-classical monocytes and neutrophils decreased with age in black males. Furthermore, we found that TLR4 expression on neutrophils negatively correlated with age in black males. The negative correlation between TLR4 and CD14 density and age could be a result of age-related changes in monocyte phenotypes (De Maeyer and Chambers, 2021). Interestingly, while we saw a relationship between mCD14 and age, we did not find a significant correlation between sCD14 with age. Studies in mixed race American cohorts have,

however, found that sCD14 increased with age (Reiner et al., 2013, Olson et al., 2020, Stanislawski et al., 2021), thus highlighting the danger of extrapolating findings from studies conducted in different population groups.

No significant correlations between age and plasma levels of LBP were detected in this study irrespective of ethnicity or sex. This is in agreement with findings from a study conducted in a population of predominantly Caucasian adults older than 60 years of age, where they did not find a correlation between LBP plasma levels with age even after adjusting for race, gender and body mass index (Stehle et al., 2012). In contrast, a study conducted on healthy individuals from Spain found a positive correlation between age and LBP plasma levels (Gonzalez-Quintela et al., 2013). In animal models, intestinal permeability and microbial translocation increases with age and this is driven by age-related inflammation and microbial dysbiosis (Thevaranjan et al., 2017). A recent study in healthy humans, however, found that age did not affect the functional capacity of the intestinal barrier (Wilms et al., 2020). More studies are thus needed in different population groups to determine the relationship between gut permeability and age and consequently the effect on the downstream markers of microbial translocation.

The differential regulation of TLR4 expression in response to LPS/LTA stimulation of whole blood, according to ethnicity and sex, was striking. In response to LPS stimulation, TLR4 was downregulated on monocytes and neutrophils in black South Africans, but the opposite occurred in white South Africans – TLR4 expression was upregulated. Given that TLR4 is activated by LPS, an increase in TLR4 expression would be expected. Our findings in white individuals are consistent with those of a study conducted in Norway, which found that stimulation with different concentrations of LPS increased TLR4 mRNA expression in monocytes and polymorphonuclear cells (Muzio et al., 2000). Although ethnicity was not specified, over 90% of the Norwegian population are white. Interestingly, in our study, when white South Africans were stratified by sex, the effect of LPS stimulation on TLR4 expression was lost in males but remained in females. This suggests that the sex differences in baseline TLR4 expression levels in white South Africans may influence the immune response to LPS. White females had a more substantial upregulation of TLR4 with both stimuli. Sex differences in response to LPS stimulation have been previously reported. Compared to females, males have been shown to produce higher levels of pro-inflammatory cytokines in response to LPS stimulation; however, other studies did not find sex differences in cytokine production after LPS stimulation (Bruunsgaard et al., 1999, Heesen et al., 2002, Aulock et al., 2006, Lefèvre et

al., 2012, Beenakker et al., 2020). Furthermore, a higher cytokine response in response to LPS stimulation was present in younger men compared to younger women, however this relationship was not seen in older individuals (Bruunsgaard et al., 1999).

In mice, downregulation of TLR4 post LPS stimulation has been reported (Nomura et al., 2000). This is consistent with the downregulation seen in black South African population, which could be attributed to endocytosis of the TLR4 receptor and/or endotoxin tolerance. LPS-induced downregulation of TLR4 expression on monocytes through endocytosis, thereby attenuating inflammatory immune responses, has been reported (Husebye et al., 2006). Endotoxin/LPS tolerance is a diminished response to LPS post an initial exposure to LPS (Biswas and Lopez-Collazo, 2009). Given that LBP plasma levels are an indirect marker of microbial translocation, higher plasma levels of LBP seen in black South Africans could be indicative of more LPS exposure, thereby resulting in endotoxin tolerance after LPS stimulation and subsequent downregulation. The complex interaction between LPS, LBP, and TLR4 and its consequential effect on cytokine production and the immune response in different population groups needs to be studied further.

Purified LTA does not stimulate TLR4 in NF-κβ reporter cell lines as measured by the expression of CD25 (Han et al., 2003). Although LTA is not a TLR4 ligand, LTA significantly upregulated TLR4 expression in whole blood on both monocytes and neutrophils in white South Africans, and on neutrophils in black South Africans. Given that CD14 is also a receptor for LTA (Schroder et al., 2003), LTA stimulation could have indirectly upregulated the TLR4 expression through the activation of monocytes and neutrophils via CD14. Differential cytokine induction through either stimulant, affected by ethnicity, sex or age, could also indirectly affect surface expression of TLR4 or of CD14.

The main strength of this study is that we collectively determined the expression levels of the LPS signalling receptors and accessory molecules in the same individuals, which allowed us to also investigate the relationship between these molecules. Although the small sample sizes were a limitation of this study, especially when investigating sex differences, some of the associations were strong regardless of the small numbers and form a strong basis for future investigations. Additionally, due to the exploratory nature of this study, we did not correct for multiple comparisons, therefore some of the findings could be false positives. Verification of these findings in larger replication cohorts is required. Another weakness of this study is that

we did not obtain demographic information from our study participants. For example, weight, BMI and fasting status at the time of blood collection can influence the plasma levels of LBP.

Taken together, findings show that there are baseline population differences in LBP plasma levels between black and white South Africans and population-specific sex differences (TLR4), as well as, population, and sex-specific influence of age on the expression levels of CD14 and TLR4. These findings suggest that there are distinct differences in factors, which regulate the expression levels of these molecules between black and white South Africans, which could translate to differences in innate immune responses between these two populations. Importantly, our findings underscore the importance of taking ethnicity, sex, and age into consideration when quantitating biological markers in relation to any clinical outcomes of infection/disease.

CHAPTER 3

Variation in the genes for LPS recognition receptors, CD14 and TLR4, in the black South African population: comparison with other populations and prevalence in HIV-1 controllers

3.1 Introduction

Cluster of differentiation 14 (CD14) and Toll-like receptor 4 (TLR4) are key molecules in the activation of the innate immune response by lipopolysaccharide (Lu et al., 2008). CD14 is a glycoprotein found either as a glycosylphosphatidylinositol-anchored protein (mCD14) on the surface of monocytes, macrophages, and neutrophils, or as a soluble form (sCD14) derived from direct secretion from the liver or from cleavage of the mCD14 (Su et al., 1999). Although CD14 has a binding domain for LPS, it lacks a transmembrane domain (Dauphinee and Karsan, 2006) and thus TLR4 is required for intracellular signalling. TLR4 is a type 1 transmembrane protein consisting of an extracellular domain, a transmembrane signalling domain, and a cytoplasmic domain (Kang and Lee, 2011). Like CD14, TLR4 is mostly expressed by cells of myeloid origin (Muzio et al., 2000).

CD14 is encoded by the *CD14* gene located on chromosome 5 q23−q31 (Goyert et al., 1988). The *CD14* gene is approximately 1.9 kb in length and has two exons. A single nucleotide polymorphism (SNP) located in the *CD14* gene promotor region (rs2569190, also known as -159C>T or -260C>T) has been extensively studied. This variant has been associated with the transcriptional activity of the *CD14* gene promotor (LeVan et al., 2001). Homozygosity for the minor allele of the rs2569190 SNP (T/T) has been associated with higher plasma levels of sCD14 and higher mCD14 density on monocytes (Baldini et al., 1999, Hubacek et al., 1999, Eng et al., 2004, Lin et al., 2007, Levan et al., 2008). Minor allele homozygosity for rs2569190 has also been associated with higher levels of TNF-α and IL-6 and lower levels of IL-10 upon stimulation with LPS, and higher levels of TNF-α after stimulation with *Chlamydiae* (Eng et al., 2004, Lin et al., 2007). Several studies have shown that this rs2569190 SNP also associates with the risk and severity of several diseases including myocardial infarction, tuberculosis, periodontitis, asthma, and sepsis (Hubacek et al., 1999, Unkelbach et al., 1999, Shimada et al., 2000, Laine et al., 2005, Rosas-Taraco et al., 2007, Wu et al., 2017).

TLR4 is encoded by the *TLR4* gene located on chromosome 9 q32-33 (Rock et al., 1998). The *TLR4* gene is composed of three exons and is approximately 19 kb in length. Two nonsynonymous *TLR4* SNPs, namely rs4986790 (Asp299Gly) and rs4986791 (Thr399Ile), have been extensively studied. These amino acid changes are within the TLR4 extracellular domain (Rallabhandi et al., 2006) and have been shown to affect TLR4 responsiveness to inhaled LPS in humans (Arbour et al., 2000, Newport et al., 2004, Rallabhandi et al., 2006). The rs4986790 and rs4986791 *TLR4* SNPs are in strong linkage disequilibrium (LD) in Caucasians, but not in West African and South African populations (Arbour et al., 2000,

Newport et al., 2004, Kresfelder et al., 2011), illustrating the importance of population-specific studies. These *TLR4* SNPs have been associated with the risk and severity of a range of infectious and non-infectious diseases including cancers, periodontal disease, sepsis as well as HIV-1 infection (Noreen et al., 2012, Vidyant et al., 2019). In people living with HIV-1, rs4986790 and rs4986791 have been associated with high viral loads (Pine et al., 2009) and late $CD4^+$ T-cell recovery after antiretroviral therapy (Yong et al., 2016). Another TLR4 variant of significance is the rs11536889 SNP which is located in the 3' untranslated region of the *TLR4* gene. This variant significantly upregulates TLR4 expression in monocytes (Sato et al., 2012) and has also been associated with several diseases (Zheng et al., 2004, Hishida et al., 2009, Zhou et al., 2011, Liu et al., 2018).

It is well established that African populations are more genetically diverse compared to non-African populations with different LD patterns. The *CD14* and *TLR4* genes, however, have not been extensively studied in African populations. In this study, using whole genome sequencing (WGS) data, we describe genetic variation and linkage disequilibrium patterns in the *CD14* and *TLR4* genes in black South African individuals and compare to data available from the 1000 Genomes Project (Auton et al., 2015). Furthermore, WGS data available for a cohort of black South African HIV-1 controllers, allowed us the opportunity to determine putative associations of interest in the context of HV-1 control based on overrepresentation/underrepresentation of various genetic variants relative to the uninfected control individuals.

3.2 Methods

3.2.1 Study participants

Whole genome sequencing (WGS) data for 87 HIV-1 uninfected black South African individuals were used for the characterisation of the *CD14* and *TLR4* genetic variation and linkage disequilibrium analysis as well as for comparing to the South African HIV-1 controllers (described below). These black South African individuals are a subset of the previously described AWI-Gen H3Africa Consortium study (Ramsay et al., 2016, Ali et al., 2018). For purposes of describing the genetic variation, these individuals will be referred to as the South African (SA) population, and for purposes of comparison to the HIV-1 controllers, these individuals will be referred to as South African (SA) controls since they are serving as a healthy (absence of HIV-1 infection) reference group.

In addition, WGS data for 39 black South African HIV-1 controllers were available from a study conducted in our research unit, and first used towards a study of genes involved in absorption, distribution, metabolism, and excretion (ADME) of drugs in African populations (da Rocha et al., 2021). HIV-1 controllers are broadly defined as individuals who can control the HIV-1 infection in the absence of antiretroviral therapy (ART). They were recruited from four different sites in Johannesburg, South Africa, namely Chris Hani Baragwanath Hospital, Helen Joseph Hospital, Life Brenthurst Hospital, and Toga Laboratories. The HIV-1 controllers were comprised of elite controllers (ECs, n=21), viraemic controllers (VCs, n=6), and high viral load long-term non-progressors (HVL LTNPs, n=12). Elite controllers were defined as people living with HIV-1 (PLWH), who in the absence of ART can maintain a viral load (VL) of less than 50 RNA copies/ml. Viraemic controllers were comprised of individuals who can maintain VL > 50 RNA copies/ml but < 2000 RNA copies/ml. Lastly, HVL LTNPs were defined as PLWH that can maintain CD4+ T cell counts > 500 cells/µl in the presence of VL > 10,000 copies/ml for > 7 years without the use of ART.

Written informed consent was obtained from all study participants and this study was approved by the University of the Witwatersrand Human Research Ethics Committee (HREC).

3.2.2 Whole genome sequencing

Whole genome sequencing of the 87 black South African controls was conducted by the Broad Institute Genomic Services (Massachusetts Institute of Technology, Cambridge, Massachusetts, USA). The TruSeq Nano DNA Library Prep Kits were used for sample preparation and the WGS was performed on an Illumina TenX (150 bp) to a minimum depth of coverage of 30x. WGS data for the HIV-1 controllers were also generated commercially by Edinburgh Genomics (Edinburgh, Scotland). The Illumina SeqLab workflow system and high coverage sequencing (~40X) on the Illumina HiSeqX platform were used to generate the WGS data.

Analysis of the WGS data was performed using the Broad Institute best practice pipeline. De-multiplexing and adapter trimming was done using Bcl2fastq (v2.17.1.14) and BCBio-Nextgen (v0.9.7) was used to analyse the data. Sequencing reads were aligned to the hg38 reference using Bwa-mem (v0.7.13), and Samblaster (v0.1.22) was used to mark duplicated fragments. Variant calling was performed using HaplotypeCaller implemented in GATK (v4.0.8.0). Variant Quality Score Recalibration implemented in GATK (v4.0.8.0) was used for filtering of variants (SNPs and indels). Appropriate quality control metrics were applied to the data and

bad calls were filtered out. Variants that passed the quality assessment were used for subsequent analysis. However, variants that were only detected in single individuals could still be due to a sequencing error, and would therefore need to be validated using other methodologies. It should be noted that in this study, wildtype vs. mutant allele refer to major and minor alleles in our South African population which do not always correspond with the 1000 Genomes Project populations (i.e., a particular variant may have an allele that is reported as the minor allele in the SA population, when in fact it is the predominant allele in the 1000 Genomes Project populations.

3.2.3 Linkage disequilibrium analysis and Hardy-Weinberg equilibrium

The Haploview software version 4.2 was used to calculate LD between pairs of SNPs. Both D' and R^2 values were calculated and used as measures of LD to infer putative haplotypes in the black South African population (n=87). Variants with a minor allele frequency of less than 1% were excluded when inferring haplotypes. The frequencies of the putative haplotypes were estimated using the Haploview software and manually confirmed by counting the number of variant alleles making up the putative haplotype and dividing by the total number of alleles. Deviation from Hardy-Weinberg equilibrium for all SNPs detected in the *CD14* or *TLR4* genes was calculated using the exact tests of Wigginton, Cutler, and Abecasis (Wigginton et al., 2005) implemented through the Haploview software version 4.2. The South African controls and HIV-1 controllers were tested independently.

3.2.4 Data analysis

To compare allelic and genotypic representation of *CD14* and *TLR4* variants, Fisher's exact tests were performed using an in-house developed and validated data analysis pipeline (Reveley et al., unpublished). Variants across the entire gene as well as 2 kb upstream and downstream of the genes were used for analysis. All statistical tests were two-sided and statistical significance was set at $p < 0.05$ and 95% confidence intervals (CI) of odds ratios (OR) were also calculated. The pipeline was also used to obtain allelic data for *CD14* and *TLR4* variants available from the 1000 Genomes Project [Yoruba in Ibadan, Nigeria (YRI, n=108) and Luhya from Webuye, Kenya (LWK, n=99) African subpopulations, European population (EUR, n= 503), East Asian population (EAS, n=504), and South Asian population (SAS, n=489)], for comparisons with the SA population. Genotypic data for variants showing LD in our population group were obtained from 1000 Genomes Project reference populations to carry out linkage disequilibrium analysis for comparative purposes.

To predict the consequences of nonsynonymous SNPs, we used the Variant Effect Predictor (VEP) tool (https://www.ensembl.org/Tools/VEP) which generates SIFT (Sorting Intolerant from Tolerant) and PolyPhen-2 (Polymorphism Phenotyping v2) prediction scores. We used RegulomeDB v2.1 (https://regulomedb.org/) database to predict the function effect of non-coding variants (intronic, and 5' and 3' flanking region variants). RegulomeDB is an integrated database that provides functional annotation of non-coding human variants based on high throughput data sets from ENCODE and other sources such as the GTEx project (Boyle et al., 2012, Dong et al., 2022). The RegulomeDB annotations are based on data such as transcription factor binding sites, protein motifs, chromatin state, expression quantitative trait loci (eQTL), and computational predictions (DNase footprinting, position weight matrices). RegulomeDB uses a scoring system that classifies variants into categories 1 to 7, with lower categories representing variants which are more likely to be located in a functional region. This study focused on identifying category 1 variants, which are likely to be in a functional region and are thus more likely to affect gene expression. Category 1 variants are further divided into subcategories (1a – 1f), with a variant scored as 1a having the highest confidence on functionality (Boyle et al., 2012).

3.3 Results

3.3.1 Genetic characterisation of the *CD14* and *TLR4* genes

3.3.1.1 Single nucleotide polymorphisms in the *CD14* gene

Forty-six variants were identified in the *CD14* gene and flanking regions. The distribution and frequency of these variants with comparisons to select populations (YRI, LWK, EUR, EAS, and SAS) from the 1000 Genomes Project are shown in **Table 3.1**. Forty-one of the identified *CD14* variants were SNPs and six of these SNPs were present in single individuals with a minor allele frequency (MAF) of 0.6%.

We identified four SNPs in the coding region (exon 2) of the *CD14* gene - one resulted in a synonymous mutation (rs4914), and three resulted in nonsynonymous mutations (rs754747693, rs377235098, and rs11556179). Representation of the synonymous SNP (rs4914; Leu367=) was significantly higher in the SA population compared to the European population, but not different when compared to the African subpopulations (YRI and LWK) and Asian populations (EAS and SAS). The rs11556179 SNP, which results in a substitution of glutamic acid with lysine at position 341 (Glu341Lys) is predicted to be 'possibly damaging'

and ‘deleterious’ by PolyPhen-2 and SIFT, respectively, and appears to be exclusive to African populations. Data are not available in 1000 Genomes Project for the rs754747693 (threonine to isoleucine, Thr129Ile) and rs377235098 (valine to isoleucine, Val333Ile) SNPs, which both occur at low frequency in our SA population. These SNPs are both predicted to be benign and tolerated by SIFT and PolyPhen-2.

Only one SNP (rs1581449468) was identified in the intronic region of the *CD14* gene. This SNP was only found in a single individual in the SA population, and no data for this SNP were available in the 1000 Genomes Project. Most SNPs detected in *CD14* were found in the 5’ and 3’ flanking regions of the gene (25 and 11 SNPs, respectively). A SNP located in the 3’UTR was a newly identified variant (designated *CD14* N1-1) – this polymorphism was however only found in one individual and thus requires verification. Comparisons with the 1000 Genomes Project data showed extensive differences in representation of the *CD14* 5’ and 3’ flanking region variants when the SA population was compared to the European and Asian populations. As expected, the SA population showed a strong similarity to the African subpopulations, however more similarity was seen with the east African LWK population, with only one significantly different SNP representation (rs2569191) between these two populations.

Table 3.1: Minor allele frequencies of *CD14* gene variants in the black South African population and comparison to select populations from the 1000 Genomes Project.

Location on gene	Position (Chr:5)	Accession number	Base change (wt/mut)	n(minor allele frequency, %) SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
	140635623	rs2569192	G/C	39 (22.4)	43 (19.9)	42 (21.2)	131 (13.0)	261 (25.9)	215 (22.0)
	140635596	rs5744446	T/C	17 (9.8)	18 (8.3)	22 (11.1)	0	1 (0.1)	0
	140635541	rs5744447	G/A	3 (1.7)	11 (5.1)	6 (3.0)	0	2 (0.2)	0
	140635490	rs73271519	C/T	17 (9.8)	18 (8.3)	22 (11.1)	0	1 (0.1)	0
	140635324	rs5744448	G/T	14 (8.0)	33 (15.3)	20 (10.1)	0	5 (0.5)	0
	140635201	rs866722455	C/G	2 (1.1)	-	-	-	-	-
	140635128	rs568528336	T/G	2 (1.1)	1 (0.5)	0	0	0	0
	140635028	rs3138076	T/C	17 (9.8)	50 (23.1)	25 (12.6)	298 (29.6)	255 (25.3)	247 (25.3)
	140634792	rs2915863	T/C	38 (21.8)	39 (18.1)	42 (21.2)	538 (53.6)	426 (42.1)	472 (48.3)
	140634698	rs186291587	C/T	1 (0.6)	1 (0.5)	0	0	0	1 (0.1)
	140634532	rs3138078	C/A	17 (9.8)	50 (23.1)	25 (12.6)	298 (29.6)	250 (24.9)	247 (25.3)
	140634506	rs6875483	G/T	16 (9.2)	16 (7.4)	10 (5.1)	0	0	0
5' Flanking region	140634475	rs189034252	T/C	1 (0.6)	5 (2.3)	0	0	0	0
	140634418	rs116340037	G/A	4 (2.3)	4 (1.9)	3 (1.5)	0	0	0
	140634318	rs2569191	T/C	56 (32.2)	73 (33.8)	86 (43.4)	577 (57.4)	491 (48.6)	516 (52.8)
	140634298	rs5744451	A/C	16 (9.2)	16 (7.4)	10 (5.1)	0	0	0
	140633982	rs5744454	T/G	17 (9.8)	50 (23.1)	25 (12.6)	298 (29.6)	255 (25.3)	247 (25.3)
	140633846	rs75652866	G/C	16 (9.2)	17 (7.9)	10 (5.1)	0	0	0
	140633807	rs1477426797	**TG/T**	1 (0.6)	-	-	-	-	-
	140633794	rs1247251479	T/C	14 (8.0)	-	-	-	-	-
	140633784	rs112865855	T/C	14 (8.0)	33 (15.3)	20 (10.1)	0	5 (0.5)	0
	140633722	rs5744455	G/A	3 (1.7)	17 (7.9)	5 (2.5)	298 (29.6)	244 (24.3)	248 (25.4)
	140633688	rs142256657	**A/AT**	14 (8.0)	33 (15.3)	20 (10.1)	0	5 (0.5)	0
	140633514	rs1314258554	C/T	2 (1.1)	0	0	0	0	0
	140633331	rs2569190	G/A	58 (33.3)	68 (31.5)	85 (42.9)	575 (57.2)	491 (48.6)	516 (52.8)
	140633206	rs34424920	T/C	2 (1.1)	4 (1.9)	4 (2.0)	0	0	0
	140633177	rs142540319	A/C	4 (2.3)	1 (0.5)	9 (4.5)	0	0	0
Intron 1	140632997	rs1581449468	G/A	1 (0.6)	-	-	-	-	-
	140632598	rs754747693	G/T	1 (0.6)	-	-	-	-	-
	140632229	rs770147646	**TG/T**	2 (1.1)	-	-	-	-	-
Exon 2	140631987	rs377235098	C/T	2 (1.1)	-	-	-	-	-
	140631963	rs11556179	C/T	5 (2.9)	1 (0.5)	4 (2)	0	0	0
	140631883	rs4914	C/G	12 (6.9)	14 (6.5)	12 (6.1)	24 (2.4)	110 (10.9)	49 (5.0)
	140631787	rs1369331514	C/T	2 (1.1)	-	-	-	-	-
	140631764	rs5744456	A/T	3 (1.7)	16 (7.4)	2 (1)	0	1 (0.1)	0
	140631730	rs2563298	C/A	37 (21.3)	36 (16.7)	37 (18.7)	133 (13.2)	261 (25.9)	215 (22.0)
	140631336	rs10691483	**A/AGTGT**	3 (1.7)	-	-	-	-	-
	140631233	*CD14* NI-1	G/A	1 (0.6)	-	-	-	-	-
3' Flanking region	140631141	rs12108825	A/G	22 (12.6)	8 (3.7)	13 (6.6)	0	0	0
	140630615	rs79468340	G/A	8 (4.6)	8 (3.7)	2 (1.0)	0	0	0
	140630586	rs1332242490	**TTGATGAAGTG/T**	1 (0.6)	-	-	-	-	-
	140630479	rs137971674	C/T	4 (2.3)	4 (1.9)	3 (1.5)	0	0	0
	140630368	rs867339288	C/T	2 (1.1)	-	-	-	-	-
	140630155	rs141887474	G/A	4 (2.3)	3 (1.4)	4 (2.0)	0	0	0
	140630056	rs113191362	C/G	8 (4.6)	1 (0.5)	5 (2.5)	0	0	0
	140629752	rs148099171	G/A	1 (0.6)	0	2 (1.0)	0	12 (1.2)	1 (0.1)

wt/mut: wildtype/mutant referring to major and minor alleles, respectively; **SA**: South Africa; **YRI**: Yoruba in Ibadan, Nigeria; **LWK**: Luhya from Webuye, Kenya; **EUR**: European population; **EAS**: East Asian population; **SAS**: South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); -: No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic); *CD14*. NI-1: *CD14* newly identified variant. Indels are indicated with bold base changes.

Using the RegulomeDb database to predict the effects of the *CD14* 5' and 3' flanking region SNPs on *CD14* gene regulation, within the 5' flanking region, a common SNP across populations (rs2569190) had the lowest RegulomeDb rank of 1a, and three SNPs (rs112865855, rs5744455, and rs2915863) had a RegulomeDb rank of 1b. These two latter

SNPs were also common across all populations studied. Furthermore, we found six SNPs with a RegulomeDb rank of 1f in the 5' flanking region, five of those SNPs (rs5744454, rs2569191, rs3138078, rs3138076, and rs2569192) were common SNPs across populations, while one SNP (rs5744448) was largely restricted to African populations (had < 1% MAF in EAS population). In the 3' flanking region, three SNPs (rs148099171, rs12108825, and rs2563298) had a RegulomeDb rank of 1f – while the rs2563298 SNP was common across populations, the rs12108825 was only found in African populations, and the rs148099171 SNP was a rare variant across all populations and absent in the YRI and EUR populations.

3.3.1.2 Single nucleotide polymorphisms in the *TLR4* gene

One hundred and seven variants were identified in the *TLR4* gene and flanking regions. The distribution and frequency of the *TLR4* variants compared to the 1000 Genomes Project populations, are shown in **Table 3.2**. Thirty of these variants were found in single individuals with MAFs of 0.6% - eight of these variants were newly identified variants in the SA population (*TLR4* NI-1 to *TLR4* NI-8), and as mentioned for *CD14*, will need to be verified. These newly identified variants were distributed throughout the gene with four in the 3'UTR, one in the 5'UTR, one in each of the two introns and one in exon 3.

Eleven *TLR4* SNPs were found in the coding region of the gene (exon 3), of which five (*TLR4* NI-4, rs201050092, rs11536884, rs5030719 and rs55786277) were amongst the variants with 0.6% MAF. The newly identified *TLR4* NI-4 results in a synonymous mutation (Ser142=). The rs201050092 (Ile301Thr), rs11536884 (Leu385Phe), rs5030719 (Gln510His) and rs55786277 (Arg804Trp) SNPs all result in nonsynonymous mutations. The rs55786277 and rs11536884 SNPs are both predicted to be deleterious by SIFT, and 'probably damaging' and 'possibly damaging', by PolyPhen-2, respectively, whereas rs5030719 and rs201050092 are predicted to be 'tolerated' and 'benign' by SIFT and PolyPhen-2, respectively. No data are available on the 1000 Genomes Project for the rs201050092 SNP. The rs11536884 and rs55786277 SNPs are reported as absent in the YRI, LWK, EUR, EAS, and SAS populations. The rs5030719 SNP is exclusive to African populations and found in significantly higher proportions in the LWK population relative to the SA population. Of the six exonic SNPs found with MAF > 0.6%, two result in synonymous mutations (rs5030710, Ser105= and rs112999435, Asn517=). The rs5030710 SNP was found exclusively in African populations whereas no data are available for the rs112999435 in the 1000 Genomes Project. The four coding region SNPs causing nonsynonymous mutations (MAF > 0.6%) were as follows, rs5030713 (Gln188Arg), rs4986790 (Asp299Gly), rs5030718 (Glu474Lys) and rs140816474 (Met659Thr). Using SIFT

and PolyPhen-2, three of the nonsynonymous SNPs (rs5030713, rs4986790 and rs5030718) were predicted to be 'tolerated' and 'benign', respectively, while the rs140816474 SNP was predicted to be 'deleterious' and 'possibly damaging', respectively. Interestingly, the rs140816474 SNP was found exclusively in our SA black population with a MAF of 4%, while the rs5030713 SNP was also present in YRI at a low frequency (MAF: 0.5%). The rs4986790 SNP was found in all population groups included in this study except for the EAS population – this was surprising given its relatively high prevalence in both EUR (MAF: 5.7%) and SAS (MAF: 12.6%) population groups. The rs5030718 SNP, however, was found in African and EAS populations but was absent in EUR and SAS populations.

The majority of TLR4 SNPs present at MAF > 0.6% were located within the introns of the gene (n=39, 54.9%). Twenty SNPs with MAF > 0.6% were identified in the 3' UTR and six in the 5'UTR. Three of the 5'UTR SNPs, namely rs11536865, rs114102395, and rs11536866 were located in the promoter region of the TLR4 gene according to Ensembl (https://www.ensembl.org/index.html). Comparisons of the MAFs (MAF 0.6 1%) of the 3' and 5' UTR and intronic region *TLR4* SNPs between the SA population and the 1000 Genomes Project reference populations showed that as expected, representation of the SNPs differed extensively between the SA population and the non-African populations (EUR, EAS, and SAS), and showed similarities to the African YRI and LWK populations. What was interesting is that unlike *CD14*, the SA population showed greater similarity to YRI than LWK in *TLR4* variant representation.

Analysis of the regulatory region SNPs using RegulomeDb did not reveal any category 1 variants in the 5' flanking region. However, in the 3' flanking region, we identified one SNP (rs11536898) with a RegulomeDb rank of 1b and eight SNPs (rs7873784, rs11536891, rs11536896, rs11536897, rs1927906, rs1554973, rs17420985, and rs60103585) with a rank of 1f. While these SNPs were common across populations, the rs11536897 SNP was rare in the SAS population.

Table 3.2: Minor allele frequencies of *TLR4* gene variants in the black South African population and comparison to select populations from the 1000 Genomes Project.

				n (minor allele frequency, %)					
Location on gene	Position (Chr:9)	Accession number	Base change (wt/mut)	SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
5' flanking region	117702392	rs10983755	G/A	1 (0.6)	1 (0.5)	0	41 (4.1)	233 (23.1)	75 (7.7)
	117702447	rs1927914	G/A	33 (19)	29 (13.4)	43 (21.7)	675 (67.1)	634 (62.9)	598 (61.1)
	117702717	rs111949799	C/T	1 (0.6)	0	1 (0.5)	0	0	0
	117702866	rs10759932	T/C	26 (14.9)	61 (28.2)	32 (16.2)	154 (15.3)	242 (24)	106 (10.8)
	117703014	rs11536863	C/T	13 (7.5)	10 (4.6)	9 (4.5)	0	0	0
	117703320	rs545731765	**TAG/T**	4 (2.3)	0	0	0	0	0
	117703395	rs374195327	**GTGA/G**	1 (0.6)	-	-	-	-	-
	117703745	rs11536865	G/C	14 (8)	40 (18.5)	22 (11.1)	2 (0.2)	0	0
	117703808	rs114102395	G/C	5 (2.9)	4 (1.9)	6 (3)	0	0	0
	117703940	rs11536866	A/G	5 (2.9)	3 (1.4)	12 (6.1)	0	0	0
	117704399	*TLR4* NI-1	G/C	1 (0.6)	-	-	-	-	-
	117704458	rs112099833	G/A	1 (0.6)	0	1 (0.5)	0	0	0
Intron 1	117705259	rs143235839	C /T	4 (2.3)	0	0	0	0	0
	117705718	rs2737195	C/T	3 (1.7)	4 (1.9)	5 (2.5)	0	0	0
	117705794	rs115809387	T /C	5 (2.9)	3 (1.4)	12 (6.1)	0	0	0
	117705818	rs73655841	A /G	30 (17.2)	22 (10.2)	16 (8.1)	0	0	0
	117706032	TLR4 N1-2	T /G	1 (0.6)	-	-	-	-	-
	117706041	rs146948012	**A /AAT**	14 (8)	6 (2.8)	3 (1.5)	0	0	0
	117706110	rs1927913	A/G	3 (1.7)	5 (2.3)	5 (2.5)	0	0	0
	117706371	rs12344353	T /C	34 (19.5)	34 (15.7)	32 (16.2)	57 (5.7)	0	123 (12.6)
	117706417	rs1159304191	C/ T	1 (0.6)	-	-	-	-	-
	117706831	rs560779117	C/T	1 (0.6)	1 (0.5)	0	0	0	0
	117707089	rs78736019	A/G	4 (2.3)	6 (2.8)	3 (1.5)	0	0	0
	117707148	rs1927912	C/T	33 (19)	27 (12.5)	22 (11.1)	0	0	0
	117707272	rs7864330	T/G	34 (19.5)	34 (15.7)	32 (16.2)	57 (5.7)	0	123 (12.6)
	117707386	rs2737196	A/G	34 (19.5)	27 (12.5)	22 (11.1)	0	0	0
	117707776	rs1927911	A/G	67 (38.5)	65 (30.1)	78 (39.4)	748 (74.4)	637 (63.2)	737 (75.4)
	117707806	rs570234836	A/C	1 (0.6)	1 (0.5)	0	0	0	0
	117707870	rs1927910	A/G	39 (22.4)	33 (15.3)	24 (12.1)	0	0	1 (0.1)
	117708080	rs1413088	G/A	37 (21.3)	33 (15.3)	25 (12.6)	0	0	0
	117708086	rs1057270421	G/T	1 (0.6)	-	-	-	-	-
	117708094	rs10759933	A/C	34 (19.5)	34 (15.7)	32 (16.2)	57 (5.7)	0	123 (12.6)
	117708271	rs548662649	A/G	4 (2.3)	0	0	0	0	0
	117708542	rs146471625	T/G	3 (1.7)	2 (0.9)	0	0	0	0
Intron 2	117708780	rs2770148	G/A	37 (21.3)	33 (15.3)	25 (12.6)	0	0	0
	117709009	rs11536874	C/T	15 (8.6)	11 (5.1)	8 (4)	0	0	0
	117709013	rs11536875	C/T	34 (19.5)	28 (13)	19 (9.6)	0	0	0
	117709042	rs11536876	A/G	6 (3.4)	13 (6)	8 (4)	0	0	0
	117709080	rs11536877	A/G	34 (19.5)	28 (13)	19 (9.6)	0	0	0
	117709275	rs11536878	C/A	3 (1.7)	10 (4.6)	13 (6.6)	101 (10)	129 (12.8)	136 (13.9)
	117709382	rs111650156	C/T	1 (0.6)	0	1 (0.5)	0	0	0
	117709505	rs144008955	C/T	3 (1.7)	1 (0.5)	3 (1.5)	0	0	0
	117709613	rs188809386	A/G	1 (0.6)	1 (0.5)	0	0	0	0
	117709658	rs553313109	**A/AT**	1 (0.6)	0	1 (0.5)	0	0	0
	117709743	rs1927909	G/A	3 (1.7)	5 (2.3)	5 (2.5)	0	0	0
	117709748	rs2737197	C/T	37 (21.3)	33 (15.3)	25 (12.6)	0	0	0
	117709877	rs1927908	C/A	3 (1.7)	4 (1.9)	3 (1.5)	0	0	0
	117709979	rs144882633	G/A	1 (0.6)	1 (0.5)	2 (1)	0	0	0
	117710053	rs11536880	A/G	34 (19.5)	28 (13)	20 (10.1)	0	0	0
	117710079	rs11536881	A/C	1 (0.6)	1 (0.5)	8 (4)	22 (2.2)	0	23 (2.3)
	117710334	rs11536882	**C/CTTTG**	6 (3.4)	7 (3.2)	15 (7.6)	418 (41.6)	637 (63.2)	377 (38.5)
	117710452	rs12377632	T/C	6 (3.4)	11 (5.1)	21 (10.6)	441 (43.8)	636 (63.1)	395 (40.4)
	117710486	rs1927907	C/T	25 (14.4)	55 (25.5)	35 (17.7)	154 (15.3)	222 (22.0)	105 (10.7)
	117710487	rs11536883	C/T	13 (7.5)	7 (3.2)	3 (1.5)	0	0	0
	117710896	rs539376532	C/A	1 (0.6)	1 (0.5)	0	0	0	0
	117710902	rs10983756	C/T	16 (9.2)	21 (9.7)	22 (11.1)	57 (5.7)	0	122 (12.5)
	117711060	rs2770146	T/C	25 (14.4)	20 (9.3)	27 (13.6)	273 (27.1)	2 (0.2)	228 (23.3)
	117711098	*TLR4* NI-3	G/T	1 (0.6)	-	-	-	-	-
	117711188	rs117343502	G/A	1 (0.6)	10 (4.6)	4 (2)	57 (5.7)	0	122 (12.5)
	117711556	rs5030717	A/G	14 (8)	50 (23.1)	28 (14.1)	113 (11.2)	7 (0.7)	33 (3.4)
	117711589	rs5030725	T/G	43 (24.7)	45 (20.8)	46 (23.2)	4 (0.4)	0	0
	117711896	rs374101533	A/T	5 (2.9)	0	1 (0.5)	0	0	0
	117711921	rs2149356	T/G	33 (19)	31 (14.4)	46 (23.2)	691 (68.7)	638 (63.3)	615 (62.9)
	117711979	rs5030727	C/T	12 (6.9)	16 (7.4)	10 (5.1)	1 (0.1)	0	0
	117712004	rs5030728	G/A	25 (14.4)	20 (9.3)	27 (13.6)	273 (27.1)	2 (0.2)	228 (23.3)
	117712061	rs5030729	A/G	44 (25.3)	33 (15.3)	27 (13.6)	0	0	0
	117712250	rs5030644	**CT/C**	45 (25.9)	33 (15.3)	28 (14.1)	0	0	0
	117712373	rs5030643	**CT/C**	44 (25.3)	33 (15.3)	27 (13.6)	0	0	0

	117712443	rs5030710	T/C	45 (25.9)	35 (16.2)	27 (13.6)	0	0	0
	117712550	TLR4 NI-4	T/C	1 (0.6)	-	-	-	-	-
	117712691	rs5030713	A/G	2 (1.1)	1 (0.5)	0	0	0	0
	117713024	rs4986790	A/G	5 (2.9)	8 (3.7)	19 (9.6)	57 (5.7)	0	123 (12.6)
	117713030	rs201050092	T/C	1 (0.6)	-	-	-	-	-
Exon 3	117713283	rs11536884	G/T	1 (0.6)	0	0	0	0	0
	117713548	rs5030718	G/A	9 (5.2)	13 (6)	11 (5.6)	0	4 (0.4)	0
	117713658	rs5030719	G/T	1 (0.6)	7 (3.2)	14 (7.1)	0	0	0
	117713679	rs112999435	C/T	8 (4.6)	-	-	-	-	-
	117714104	rs140816474	T/C	4 (2.3)	0	0	0	0	0
	117714538	rs55786277	C/T	1 (0.6)	0	0	0	0	0
	117714804	rs11536885	A/C	2 (1.1)	0	0	0	0	0
	117714824	TLR4 NI-5	**TAA/T**	1 (0.6)	-	-	-	-	-
	117714872	rs113464348	A/G	10 (5.7)	-	-	-	-	-
	117715169	rs1052811222	C/T	4 (2.3)	-	-	-	-	-
	117715396	rs11536887	A/G	14 (8)	22 (10.2)	8 (4)	0	0	0
	117715511	rs11536888	C/T	11 (6.3)	9 (4.2)	9 (4.5)	0	0	0
	117715671	TLR4 NI-6	A /T	1 (0.6)	-	-	-	-	-
	117715752	rs113017335	C /T	1 (0.6)	-	-	-	-	-
	117715754	rs7869402	C /T	29 (16.7)	62 (28.7)	44 (22.2)	41 (4.1)	38 (3.8)	71 (7.3)
	117715793	rs200721952	G /A	10 (5.7)	-	-	-	-	-
	117716061	TLR4 NI-7	G A	1 (0.6)	-	-	-	-	-
	117716658	rs7873784	G/C	47 (27)	56 (25.9)	54 (27.3)	134 (13.3)	104 (10.3)	111 (11.3)
	117716691	TLR4 NI-3	A/G	1 (0.6)	-	-	-	-	-
3' flanking region	117716762	rs144852071	G/T	5 (2.9)	1 (0.5)	6 (3)	0	0	0
	117717059	rs11536891	T/C	52 (29.9)	56 (25.9)	62 (31.3)	134 (13.3)	106 (10.5)	110 (11.2)
	117717109	rs183679927	G/A	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	117717204	rs11536892	G/A	5 (2.9)	12 (5.6)	11 (5.6)	0	0	0
	117717274	rs11536893	G/A	2 (1.1)	8 (3.7)	9 (4.5)	0	0	1
	117717456	rs11536896	T/C	52 (29.9)	56 (25.9)	62 (31.3)	134 (13.3)	106 (10.5)	110 (11.2)
	117717496	rs145801336	**TACTG/T**	3 (1.7)	9 (4.2)	6 (3)	61 (6.1)	99 (9.8)	55 (5.6)
	117717732	rs11536897	G/A	14 (8)	17 (7.9)	21 (10.6)	52 (5.2)	4 (0.4)	51 (5.2)
	117717818	rs1829376677	A/G	1 (0.6)	-	-	-	-	-
	117717837	rs1927906	T/C	74 (42.5)	116 (53.7)	77 (38.9)	102 (10.1)	38 (3.8)	185 (18.9)
	117717932	rs11536898	C/A	47 (27)	56 (25.9)	53 (26.8)	112 (11.1)	103 (10.2)	106 (10.8)
	117718208	rs141831136	C/T	2 (1.1)	3 (1.4)	5 (2.5)	0	0	0
	117718534	rs1554973	C/T	27 (15.5)	35 (16.6)	148 (74.7)	236 (23.5)	141 (14.0)	295 (30.2)
	117718615	rs17420985	T/C	52 (29.9)	56 (25.9)	62 (31.3)	134 (13.3)	103 (10.2)	110 (11.2)
	117718616	rs60103585	G/A	52 (29.9)	56 (25.9)	61 (30.8)	113 (11.2)	103 (10.2)	106 (10.8)

wt/mut: wildtype/mutant. **SA**: South Africa; **YRI**: Yoruba in Ibadan, Nigeria; **LWK**: Luhya from Webuye, Kenya; **EUR**: European population; **EAS**: East Asian population and **SAS**: South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); -: No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic); chr: 9: Chromosome 9; *TLR4* NI: *TLR4* newly identified variant. Indels are indicated with bold base changes.

3.3.1.3 Indels in the *CD14* and *TLR4* genes

Five indels were detected across the *CD14* gene and flanking regions (**Table 3.1**). Three of the indels were deletions located in the 5'UTR (rs1477426797), the 3'UTR (rs1332242490), and in exon 2 (rs770147646). The exon 2 deletion (rs770147646) causes a frameshift mutation, which introduces a premature stop codon in the gene. Data for these three deletions were not available in the 1000 Genomes Project. Both the 5'UTR and the 3'UTR deletions were only found in a single individual (MAF: 0.6%), however the exonic deletion was detected in two individuals (MAF: 1.1%). The two insertions detected were in the 3'UTR (rs10691483) and 5'UTR (rs142256657). The 5'UTR insertion (rs142256657) is common among the African

populations, present in the EAS population (MAF < 1%) but absent in the EUR and SAS populations, and was predicted to have a RegulomeDb rank of 1b. No 1000 Genomes Project data are available for the 3' UTR (rs10691483) insertion (however this insertion was detected in three SA individuals (MAF: 1.7%).

Six indels with a MAF > 0.6% were detected across the *TLR4* gene and flanking regions (**Table 3.2**). Two of the indels were insertions located within the first and second introns (rs146948012 and rs11536882, respectively). The rs146948012 insertion was found exclusively in African populations, however had a significantly higher presentation in the SA population compared to the YRI and LWK populations. In contrast, the rs11536882 insertion was found across all the populations, with similar representations in the African populations, but significantly higher representation in the non-African populations (EUR, EAS, and SAS) compared to the SA population. A third insertion in intron 2 (rs553313109) was found in a single individual (MAF=0.6%; **Table 3.2**) and was also present in the LWK population, but absent in the remaining populations. Of the four deletions with MAF > 0.6% detected, one was located in the 2 kb upstream region (rs545731765), and interestingly was unique to the SA population - found in four individuals with a MAF of 2.3%. Two deletions (rs5030644 and rs5030643) found in the second intron of the *TLR4* gene were exclusive to African populations, and had significantly higher representation in the SA population compared to the YRI and LWK populations. The fourth deletion (rs145801336) located in the 3'UTR, was found across all populations and representation of this deletion was significantly higher in non-African populations compared to the SA population. In addition, two rare deletions (MAFs=0.6%) were identified in the 5' and 3' flanking regions (**Table 3.2**). No 1000 Genomes Project data were available for these two deletions.

3.3.2 Linkage disequilibrium and identification of putative haplotypes in the *CD14* and *TLR4* genes

3.3.2.1 CD14

Nine putative haplotypes across the *CD14* gene and flanking regions were identified. The largest of these putative haplotypes, designated *CD14* Hap-1 (**Figure 3.1**), is comprised of seven SNPs (rs142256657, rs112865855, rs1247251479, rs5744454, rs3138078, rs3138076, and rs5744448) in strong LD (r^2>0.78; haplotype frequency, HF: 8.0%). *CD14* Hap-1 appears to be comprised of two smaller putative haplotypes, one comprised of four SNPs (rs142256657, rs112865855, rs1247251479, and rs5744448) in complete LD (Hap-1a; r^2=1, HF: 8%), and the second comprised of three SNPs (rs5744454, rs3138078, rs3138076) in complete LD (Hap-1b;

r^2=1, HF=9.8%). Interestingly, visual inspection of the data revealed that while Hap-1a was always present in individuals with Hap1b, Hap-1b was found in the absence of Hap-1a. Two putative haplotypes, both comprised of three SNPs each, designated *CD14* Hap-2 (rs867339288, rs377235098, and rs866722455; r^2=1; HF: 1.1%) and *CD14* Hap-3 (rs75652866, rs5744451, and rs6875483; r^2=1; HF: 9.2%) were also identified (**Figure 3.1)**. In addition, six *CD14* putative haplotypes comprised of two SNPs each were identified (**Table 3.3**). We termed these putative haplotypes *CD14*-2-SNP Hap-1 (rs2569190, and rs2569191; r^2=0.89, HF: 31.6%), -2-SNP Hap-2 (rs2563298, and rs2569192; r^2=0.93, HF: 21.3%) -2-SNP Hap-3 (rs73271519 and rs5744446; r^2=1, HF: 9.8%), -2-SNP Hap-4 (rs137971674 and rs116340037; r^2=1, HF: 2.3%), -2-SNP Hap-5 (rs1369331514 and rs1314258554; r^2=1, HF: 1.1%) and -2-SNP Hap-6 (rs5744456 and rs5744455; r^2=1, HF: 1.7%).

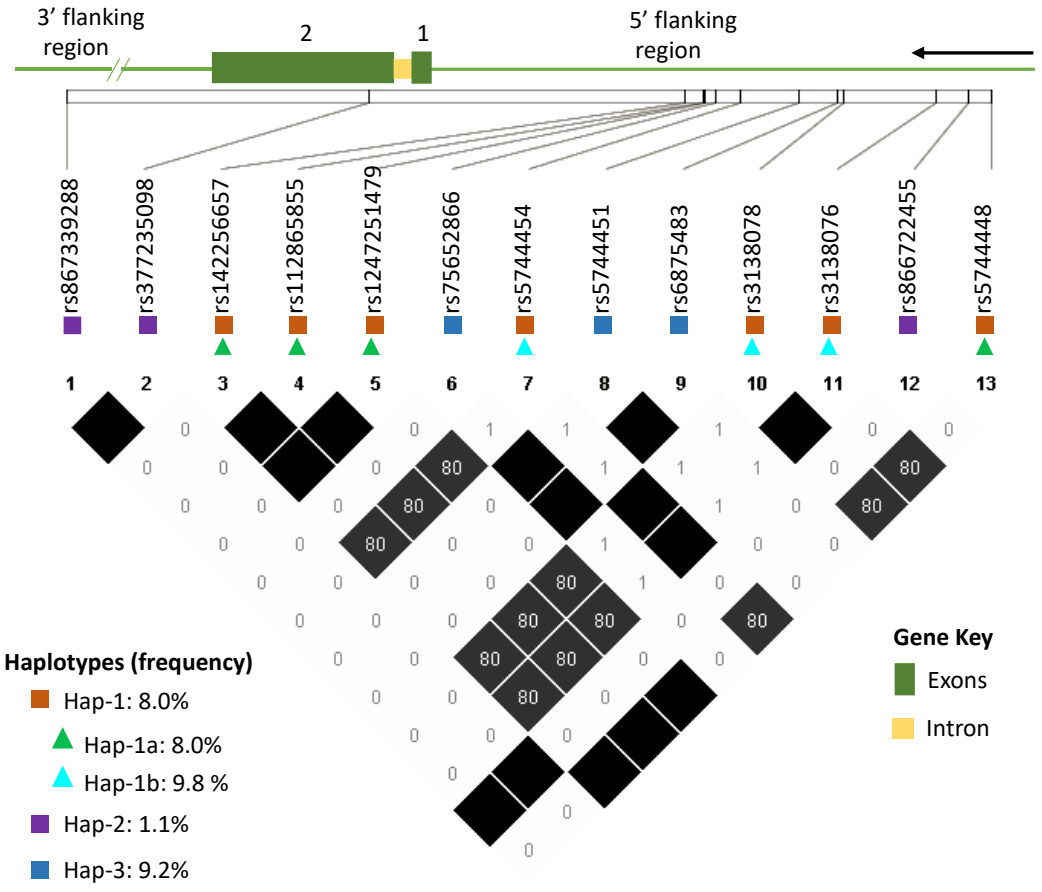


Figure 3.1: Linkage disequilibrium (LD) plot showing linkage between variants across the *CD14* gene in the black South African population. Variants present at frequencies <1% were excluded from analysis. Variants not in LD with any other variants are not shown. Values inside each diamond represent the r^2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

Table 3.3: *CD14* 2-SNP putative haplotypes pairwise LD in the black South African control population and 1000 Genomes Project reference populations.

		Variant Pair		Pairwise linkage disequilibrium					
Gene	Haplotype (2-SNP)	Variant 1	Variant 2	SA	YRI	LWK	EUR	EAS	SAS
CD14	Hap-1	rs2569190	rs2569191	89	90	94	100	99	100
	Hap-2	rs2563298	rs2569192	93	80	85	100	98	100
	Hap-3	rs73271519	rs5744446	100	100	100	100	-	-
	Hap-4	rs137971674	rs116340037	100	100	100	-	-	-
	Hap-5	rs1369331514	rs1314258554	100	-	-	-	-	-
	Hap-6	rs5744456	rs5744455	100	94	39	0	-	-

-: No variant data available on the 1000 Genomes Project., r^2 given in %. **SA**: South Africa; **YRI**: Yoruba in Ibadan, Nigeria; **LWK**: Luhya from Webuye, Kenya, **EUR**: Europeans; **EAS**: East Asians and **SAS**: South Asians.

Linkage disequilibrium patterns of the SNPs making up the *CD14* putative haplotypes in the SA population were compared to those of the 1000 Genomes Project - data however were not available for the Hap-1a-associated rs1247251479 SNP. Comparison of LD for Hap-1a (in the absence of rs1247251479) and Hap-1b SNPs in YRI and LWK, showed similar LD to the SA population (**Figure 3.2A and 3.2B)**. In Asian populations (EAS and SAS), the SNPs comprising Hap-1a were absent, however all Hap-1b SNPs were present in complete LD. In the EUR population, while Hap-1b SNPs were present in strong LD ($r^2 > 0.96$), only two of the Hap-1a SNPs (rs142256657 and rs112865855) were present and in complete LD – interestingly, these SNPs did not show any LD with Hap-1a SNPs as is seen in African populations (**Figure 3.2C)**. Comparison of Hap-2 was not possible since data for Hap-2 SNPs were not available in the 1000 Genomes Project. With regards to Hap-3, SNPs were found in complete LD in LWK, and strong LD ($r^2 > 0.93$) was seen in the YRI population (**Figure 3.2A and 3.2B)** – Hap-3-associated SNPs were however absent in the EUR, EAS, and SAS populations. Comparison of the 2-SNP putative haplotypes (**Table 3.3**) revealed that the LD patterns for 2-SNP Hap-1 and 2-SNP Hap-2 were similar across populations. Complete LD was present in YRI, LWK and EUR for 2-SNP Hap-3 and complete LD of the SNPs comprising 2-SNP Hap-4 was present in YRI, and LWK. Data for the SNPs comprising 2-SNP Hap-5 are not present in the 1000 Genomes project and only YRI showed a strong LD for 2-SNP Hap-6 SNPs.

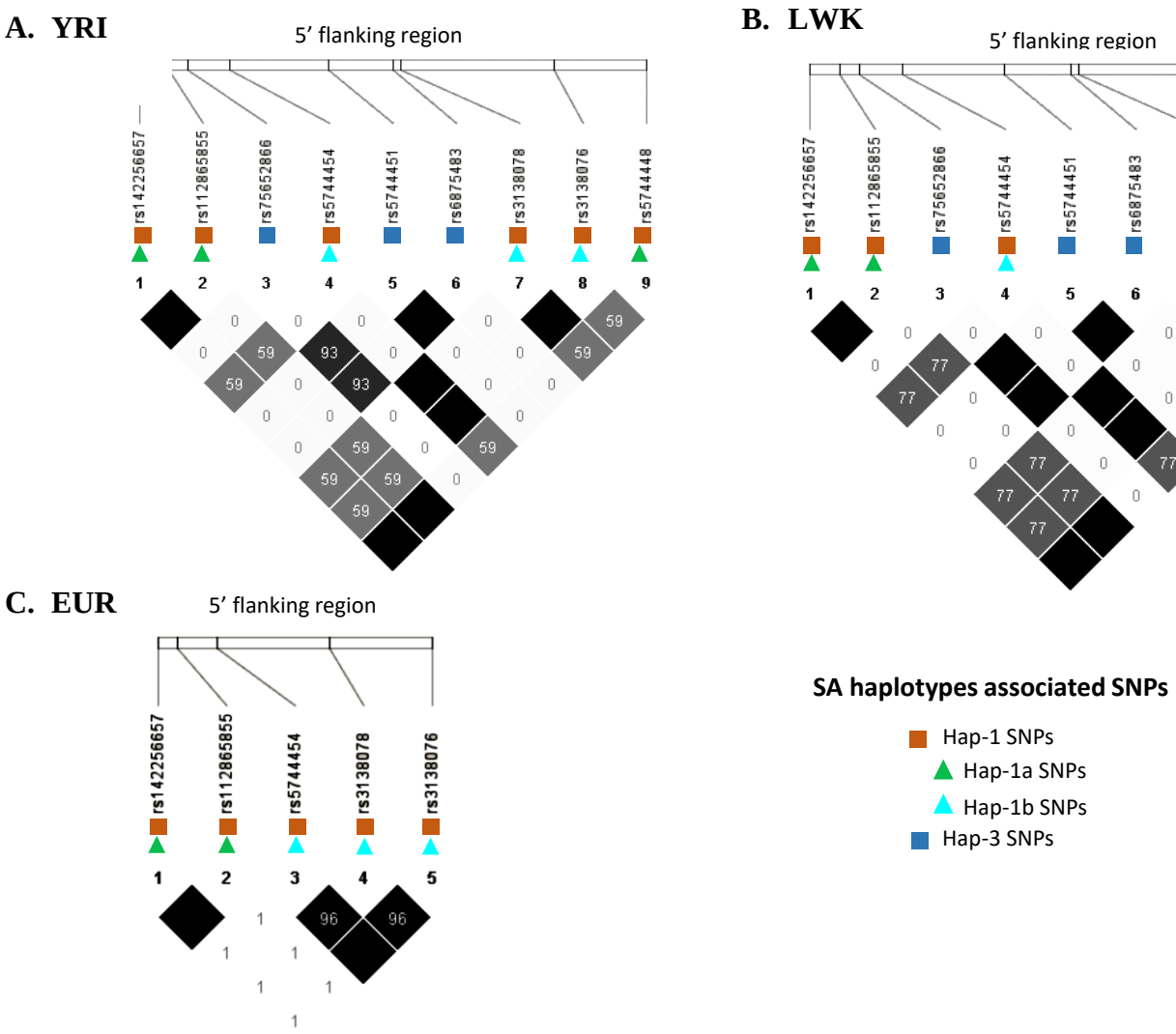


Figure 3.2: Linkage disequilibrium plot showing linkage between *CD14* variants in the (**A**) Yoruba in Ibadan, Nigeria (YRI, n=108), (**B**) Luhya from Webuye, Kenya (LWK, n=99), (**C**) Europeans (EUR, n=503), from the 1000 Genomes Project found to be in linkage disequilibrium in black South Africans. Values inside each diamond represent the r^2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

3.3.2.2 TLR4

A total of 16 putative haplotypes were detected across the *TLR4* gene and flanking regions. Putative haplotypes comprised of three or more SNPs were named in order of complexity. **Figure 3.3** shows the LD plots for putative haplotypes Hap-1 to Hap-8. The largest *TLR4* putative haplotype was comprised of 10 SNPs in strong LD (r^2 > 0.7, HF: 17.2%) and designated as *TLR4* Hap-1 (rs73655841, rs1927912, rs2737196, rs1927910, rs1413088, rs2770148, rs11536875, rs11536877, rs2737197, and rs11536880). Two *TLR4* putative haplotypes were comprised of six SNPs each, namely Hap-2 (rs11536866, rs143235839, rs115809387, rs548662649, rs140816474 and rs1052811222; r^2 > 0.8, HF: 2.7%) and Hap-3 (rs7873784, rs11536891, rs11536896, rs11536898, rs17420985 and rs60103585; r^2 > 0.7, HF: 2.3%). In addition, three *TLR4* putative haplotypes, each comprised of four SNPs each, were detected, namely, Hap-4 (rs2737195, rs1927913, rs1927909, and rs1927908; r^2=1, HF: 1.7%), Hap-5 (rs5030729, rs5030644, rs5030643, and rs5030710; r^2 > 0.9, HF: 25.3%) and Hap-6 (rs1927914, rs2770146, rs2149356, and rs5030728; r^2 > 0.7, HF: 14.4%). Lastly, two three-SNP putative haplotypes were detected in *TLR4*, namely Hap-7 (rs10759933, rs12344353 and rs7864330; r^2=1, HF: 19.5%) and Hap-8 (rs112999435, rs113464348 and rs200721952; r^2 > 0.7, HF: 4.6%).

Seven of these putative haplotypes (shown in **Table 3.4**) were comprised of 2 variants each, and similar to *CD14*, were named numerically as *TLR4*-2-SNP Hap-1 (rs11536882 and rs12377632; r^2=1, HF: 3.4%), -2-SNP Hap-2 (rs146948012 and rs11536883; r^2=0.92, HF: 7.5%), -2-SNP Hap-3 (rs11536893 and rs141831136; r^2=1, HF: 1.1%), -2-SNP Hap-4 (rs10759932 and rs1927907, r^2=0.95, HF: 14.4%), -2-SNP Hap-5 (rs11536876 and rs11536892; r^2=0.82, HF: 2.9%), -2-SNP Hap-6 (rs11536865 and rs5030717; r^2=1, HF: 8.0%) and -2-SNP Hap-7 (rs114102395 and rs144852071; r^2=1, HF: 2.9%).

Comparison of the LD patterns to the 1000 Genomes Project populations showed similar LD pattern for SNPs making up *TLR4* Hap-1 in African LWK and YRI populations (**Figure 3.4 and 3.5**), but not in non-African populations (**Figure 3.6**). Data were not available on the 1000 Genomes Project for four of the SNPs making up Hap-2, however data for the two SNPs available, namely rs11536866 and rs115809387, showed they were in complete LD in YR1 and LWK, but absent in the non-African populations. Hap-3 was conserved across populations, with strong LD for all SNPs across populations (**Figure 3.4 and 3.5**). *TLR4* Hap-4, Hap-5, and Hap-6 were all detected in the LWK and YRI populations. In non-African populations, SNPs

making Hap-4 and Hap-5 were absent and although SNPs making up Hap-6 were present, the putative haplotype was divided into two blocks, with the rs2770146 and rs5030728 SNPs in complete LD, and the rs1927914 and rs2149356 SNPs in strong LD across the non-African populations, with no LD between the two blocks. Hap-7 was detected in all populations except for the EAS population, while data for SNPs making up Hap-8 were absent in the 1000 Genomes Project populations studied. Comparisons of the TLR4 2-SNP putative haplotypes (**Table 3.4**), showed that the *TLR4*-2-SNP Hap-1 and -2-SNP Hap 4 were found across all populations with relatively strong LD (depending on the population). On the other hand, SNPs making up the remaining 2-SNP putative haplotypes were only present in the African YRI and LWK populations with varying levels of LD - strong LD for *TLR4*-2-SNP Hap-2 and -2-SNP Hap-6 SNPs, but weak LD for *TLR4*-2-SNP Hap-3 and -2-SNP Hap-5. It should be mentioned, however, that for *TLR4*-2-SNP Hap-3, only two SA individuals had these two SNPs, and thus the haplotype inference may not be accurate, whereas in the YRI and LWK populations, more individuals were assessed - this result would therefore need to be verified using a larger SA population group. Interestingly, for *TLR4*-2-SNP hap-7, both the SA and LWK populations showed complete LD for the two SNPs whereas the YRI population had very weak LD (r^2=0.25).

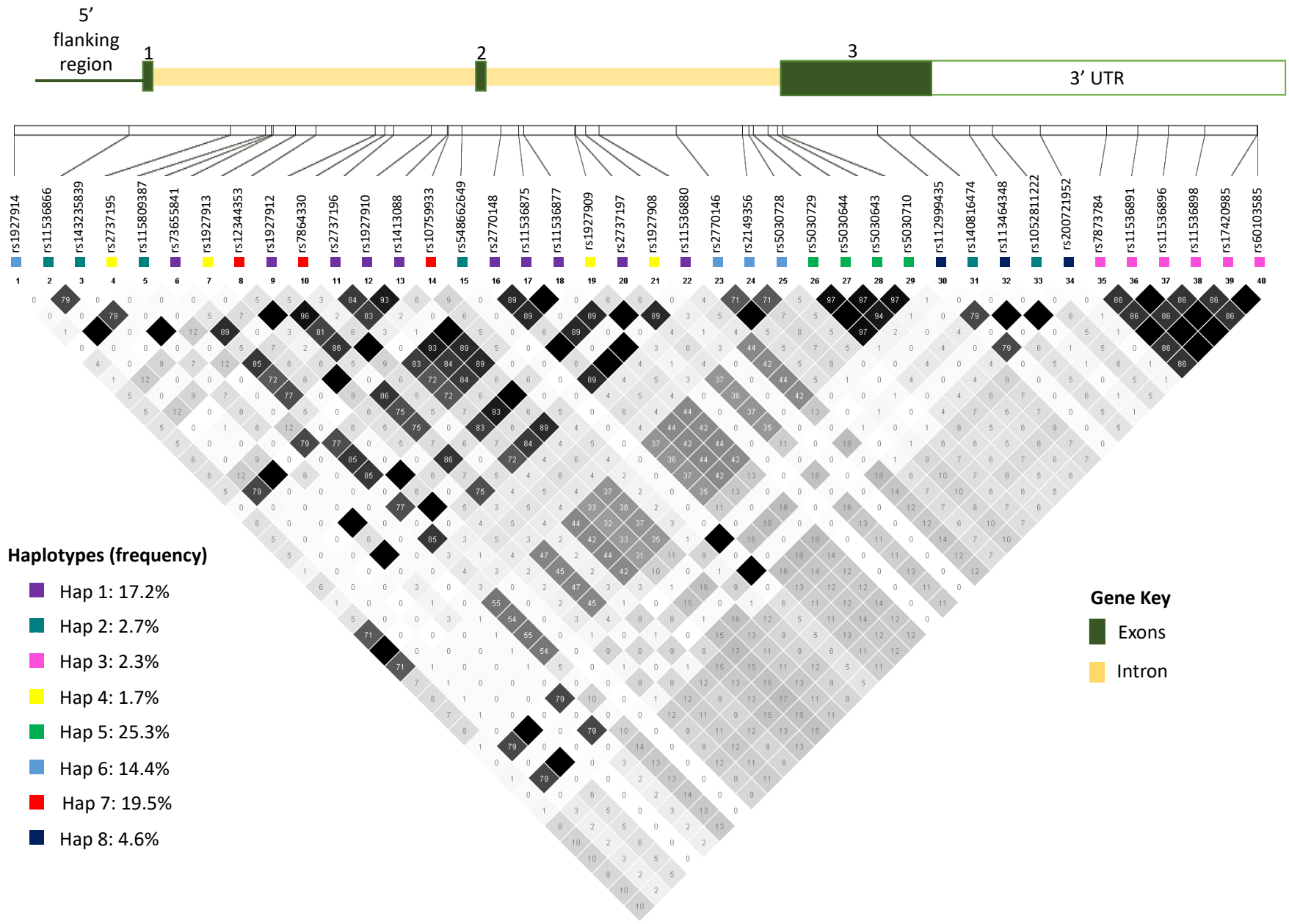


Figure 3.3: Linkage disequilibrium plot showing linkage between variants across the *TLR4* gene in the black South African population. Variants present at frequencies <1% were excluded from analysis. Variants not in LD with any other variants not shown. Values inside each diamond represent the r^2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

Table 3.4: *TLR4* 2-SNP putative haplotypes pairwise LD in the black South African control population and 1000 Genomes Project reference populations.

		Variant Pair		Pairwise linkage disequilibrium					
Gene	Haplotype (2-SNP)	Variant 1	Variant 2	SA	YRI	LWK	EUR	EAS	SAS
TLR4	Hap-1	rs11536882	rs12377632	100	68	69	87	82	91
	Hap-2	rs146948012	rs11536883	92	85	100	-	-	-
	Hap-3	rs11536893	rs141831136	100	37	54	-	-	-
	Hap-4	rs10759932	rs1927907	95	62	83	88	100	99
	Hap-5	rs11536876	rs11536892	82	28	39	-	-	-
	Hap-6	rs11536865	rs5030717	100	75	75	-	-	-
	Hap-7	rs114102395	rs144852071	100	25	100	-	-	-

-: No variant data available on the 1000 Genomes Project., r^2 given in %. **SA**: South Africa; **YRI**: Yoruba in Ibadan, Nigeria; **LWK**: Luhya from Webuye, Kenya, **EUR**: Europeans; **EAS**: East Asians and **SAS**: South Asians.

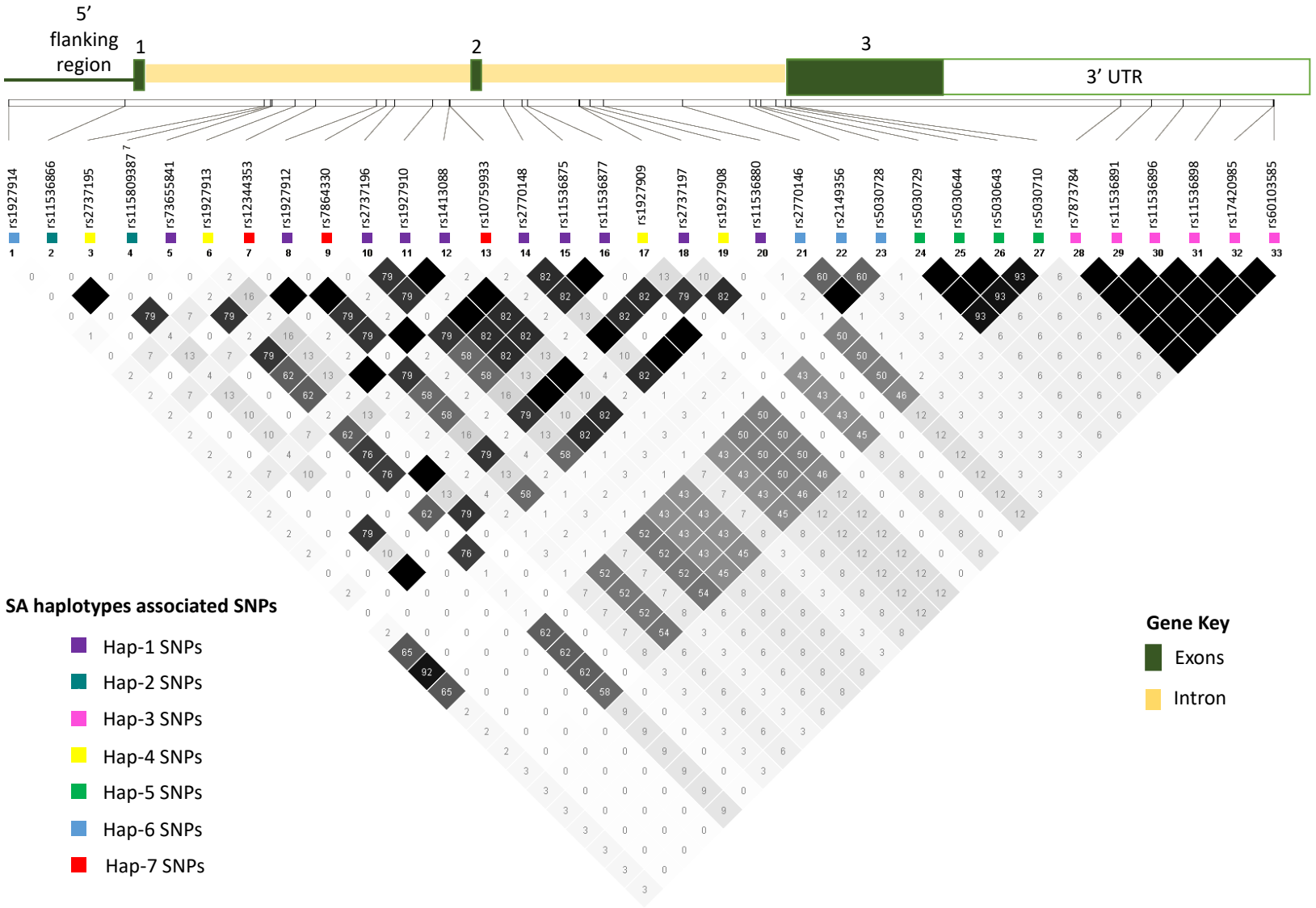


Figure 3.4: Linkage disequilibrium plot showing linkage between *TLR4* variants in the Yoruba in Ibadan, Nigeria (YRI, n=108), from the 1000 Genomes Project found to be in linkage disequilibrium in black South Africans. Values inside each diamond represent the r^2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

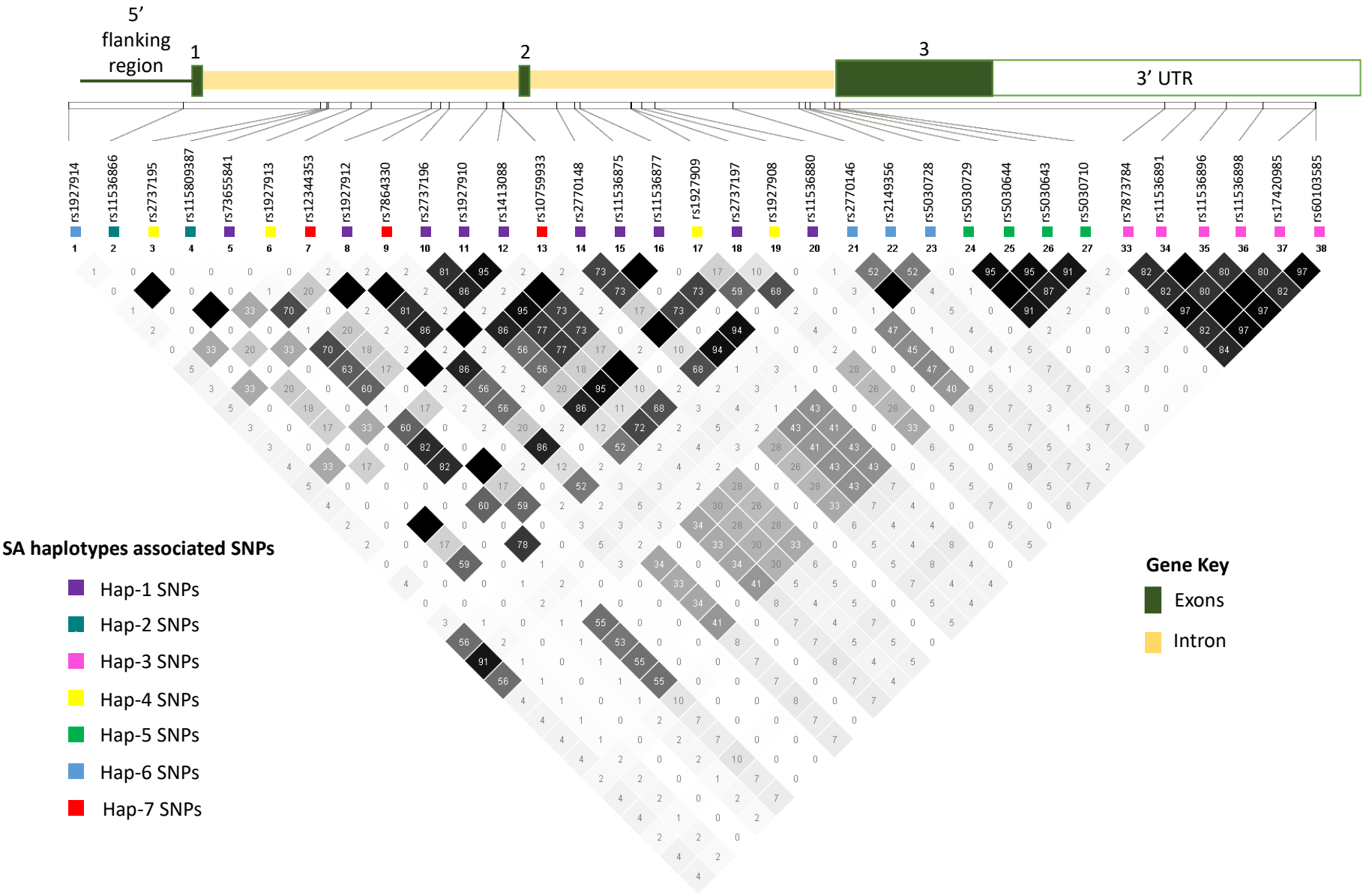


Figure 3.5: Linkage disequilibrium plot showing linkage between *TLR4* variants in the Luhya from Webuye, Kenya (LWK, n=99), from the 1000 Genomes Project found to be in linkage disequilibrium in black South Africans. Values inside each diamond represent the r2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

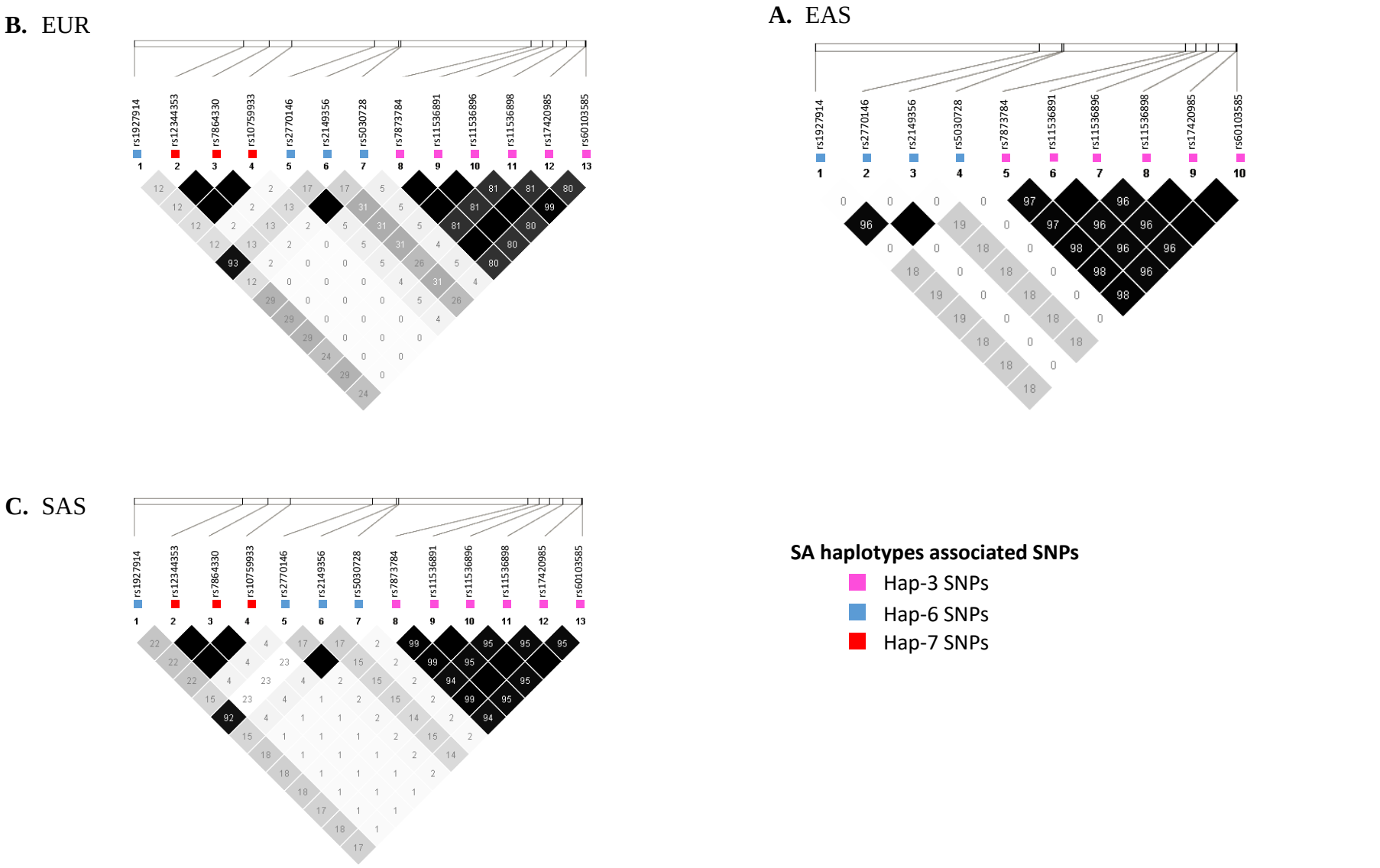


Figure 3.6: Linkage disequilibrium plot showing linkage between *TLR4* variants in (**A**) Europeans (EUR, n=503), (**B**) East Asians (EAS, n=504), and (**C**) South Asians (SAS, n=489) from the 1000 Genomes Project found to be in linkage disequilibrium in black South Africans. Values inside each diamond represent the r2 value in %. Shading in each diamond represent the strength of linkage with black shading representing complete LD and shades of grey representing lower LD and white repressing no LD.

3.3.3 Comparisons of the representation of the *CD14* and *TLR4* variants and putative haplotypes between South African controls and HIV-1 controllers

3.3.3.1 *CD14*

Comparing the representation of all identified *CD14* variants between the SA controls and HIV-1 controllers revealed the MAF of only one *CD14* SNP (rs186291587) to be significantly higher in ECs compared to the SA controls (p=0.024; OR=13.3, CI: 1.3 – 131.4). According to Ensembl (https://www.ensembl.org/index.html), the rs186291587 SNP is a regulatory region variant (CTCF binding site) located in the 2 kb upstream region of the *CD14* gene. Heterozygosity for the rs186291587 SNP was also significantly higher in ECs compared to SA controls (p=0.023; OR=14.3, CI: 1.4 – 108.1). No individuals were homozygous for the rs186291587 SNP minor allele. This SNP was absent in the in VCs and HVL LTNPs, and both allelic and genotypic representations of this SNP did not differ when VCs and HVL LTNPs were compared to the SA control group or to ECs. Representation of the *CD14* putative haplotypes did not differ significantly between SA controls and HIV-1 controllers or HIV-1 controller subgroups.

3.3.3.2 *TLR4*

Comparing the representation of all identified *TLR4* variants between the SA controls and HIV-1 controllers showed several variants to be significantly different between the groups (**Table 3.5**). The MAF of a newly identified intronic SNP (*TLR4* NI-2; **Table 3.2**) was significantly higher in HVL LTNPs compared to the SA controls (p=0.006; OR=24.71, CI: 2.46-248.51). Although this NI-2 SNP was only found in a single SA control individual, its presence in three HVL LTNP individuals verifies the authenticity of the newly identified polymorphism. A second intronic SNP, rs115919003 (C/T), not detected in the SA control population, was detected in all three HIV-1 controller groups, and MAF (3.8%) as well as heterozygosity (7,7%) for this variant was significantly higher in the total HIV-1 controller group (n=39) compared to the SA controls (MAF: p=0.029; heterozygosity: p=0.028). Although not detected in the SA controls, the rs115919003 SNP is found in the 1000 Genomes Project YRI and LWK population groups with MAFs of 1.4% and 1.5%, respectively. Minor allele frequency of a third intronic SNP, rs78736019, was also overrepresented in the HIV-1 controllers combined (p=0.038), and in HVL LTNPs (p=0.040) when compared to SA controls. A fourth intronic SNP, rs5030727, was also significantly overrepresented in HVL LTNPs (p=0.039; OR=3.55, CI: 1.13-11.18) compared to SA controls, and interestingly, this relatively prevalent SNP was absent in the VCs. Differences in the representation of a common (MAF: 25.9%) African

deletion were also observed - MAF of the rs5030644 deletion was found to be overrepresented in ECs compared to the SA controls (p=0.037; OR=2.15, CI: 1.07-4.33).

With respect to the coding regions SNPs, comparisons revealed that both MAF (p=0.039; OR=15.73, CI: 1.36-180.65) and heterozygosity (p=0.038; OR=17.2, CI: 1.43-207.08) of the rs11536884 SNP, were significantly overrepresented in HVL LTNPs compared to SA controls. The MAFs of two 3'UTR SNPs, namely, rs113017335 (p=0.006; OR=24.71, CI: 2.46-248.51) and rs1554973 (p=0.044; OR=0.367, CI: 0.413-0.943) were significantly higher in HVL LTNPs compared to SA controls. Furthermore, rs1554973 SNP minor allele homozygosity was overrepresented in the total HIV-1 controller group (p=0.028; OR=0.094, CI: 0.010-0.888) and HVL LTNPs (p=0.033; OR=0.049, CI: 0.004-0.625) compared to SA controls. Lastly, heterozygosity for two additional 3'UTR SNPs, rs11536898 (p=0.040; OR=0.39, CI: 0.163-0.934) and rs60103585 (p=0.039; OR=0.055, CI: 0.189-1.617) showed an underrepresentation in HIV-1 controllers compared to the SA control population.

Although the representation of the *TLR4* putative haplotypes did not differ significantly between the SA controls and HIV-1 controllers, or between the HIV-1 controller subgroups, when looking at the HIV-1 controller data, the *TLR4* NI-2 intronic and 3' UTR rs113017335 SNPs, both showing significant overrepresentation in HVL LTNPs are in complete LD (r^2=1), and thus represent an additional 2-SNP putative haplotype in *TLR4*, which was not detected in the SA control population due to only a single individual having these variants (i.e. analysis using the combined SA control and HIV-1 controller groups revealed the LD between these variants).

Table 3.5: Comparison of allelic and genotypic frequencies of *TLR4* variants between South African controls and HIV-1 controllers.

Allelic comparisons

		Minor allele frequency					HICs vs SA controls			ECs vs SA controls			VC vs SA controls			HV LTNPs vs SA controls		
Variant	**Minor allele**	**SAs**	**HICs**	**ECs**	**VCs**	**HVL LTNPs**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**
NI-2 (9:117706032)	G	1(0.6)	3(3.8)	0	0	3(12.5)	6.92	0.708-67.61	0.089	0	0-NaN	1	0	0-NaN	1	24.71	2.46-248.51	0.006
rs115919003	T	0	3(3.8)	1(2.4)	1(8.3)	1(4.2)	∞	NaN-∞	0.029	∞	NaN-∞	0.194	∞	NaN-∞	0.065	∞	NaN-∞	0.121
rs78736019	G	4(2.3)	7(9.0)	3(7.1)	1(8.3)	3(12.5)	4.19	1.19-14.76	0.038	3.27	0.703-15.20	0.135	3.86	0.40-37.57	0.286	6.07	1.27-29.01	0.04
rs5030727	T	12(6.9)	8(10.3)	3(7.1)	0	5(20.8)	1.54	0.604-3.940	0.45	1.04	0.279-3.859	1	0	0-NaN	1	3.55	1.13-11.18	0.039
rs5030644	C	45(25.9)	26(33.3)	18(42.9)	2(16.7)	6(25.0)	1.43	0.802-2.561	0.203	2.15	1.07-4.33	0.037	0.573	0.121-2.717	0.733	0.956	0.357-2.57	1
rs11536884	T	1(0.6)	3(3.8)	1(2.4)	0	2(8.8)	6.92	0.708-67.61	0.089	4.22	0.258-68.88	0.352	0	0-NaN	1	15.73	1.36-180.65	0.039
rs113017335	T	1(0.6)	3(3.8)	0	0	3(12.5)	6.92	0.708-67.61	0.089	0	0-NaN	1	0	0-NaN	1	24.71	2.46-248.51	0.006
rs1554973	T	27(15.5)	20 (25.6)	10(23.8)	2(16.7)	8(33.3)	0.533	0.277-1.024	0.079	0.588	0.259-1.335	0.252	0.918	0.191-4.426	1	0.367	0.413-0.943	0.044

Heterozygous genotype comparisons

		Genotype frequency					HICs vs SA controls			ECs vs SA controls			VC vs SA controls			HV LTNPs vs HCs		
Variant	**Genotype**	**SAs**	**HICs**	**ECs**	**VCs**	**HVL LTNPs**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**
rs115919003	CT	0	3(7.7)	1(4.8)	1(16.7)	1(8.3)	∞	NaN-∞	0.028	∞	NaN-∞	0.194	∞	NaN-∞	0.065	∞	NaN-∞	0.121
rs11536884	GT	1(1.1)	3(7.7)	1(4.8)	0	2(16.7)	7.17	0.721-71.23	0.089	4.3	0.258-71.726	0.353	0	0-NaN	1	17.2	1.429-207.08	0.038
rs11536898	CA	39(44.8)	9(23.1)	6(28.6)	1(16.7)	2(16.7)	0.39	0.163-0.934	0.04	0.564	0.193-1.646	0.433	0.226	0.025-2.016	0.217	0.251	0.051-1.232	0.106
rs60103585	GA	38(43.7)	9(23.1)	6(28.6)	1(16.7)	2(16.7)	0.38	0.159-0.919	0.039	0.055	0.189-1.617	0.307	0.221	0.025-1.978	0.215	0.246	0.050-1.209	0.104

Homozygous mutant genotype comparisons

		Genotype frequency					HICs vs SA controls			ECs vs SA controls			VC vs SA controls			HV LTNPs vs HCs		
Variant	genotype	SAs	HICs	ECs	VCs	HVL LTNPs	OR	CI	P	OR	CI	P	OR	CI	P	OR	CI	P
rs1554973	TT	1(1.1)	4(10.3)	2(9.5)	0	2(16.7)	0.094	0.010-0.888	0.028	0.107	0.009-1.265	0.095	∞	NaN-∞	1	0.049	0.004-0.625	0.033

SAs: South African (SA) controls (n=87), **ECs**: elite controllers (n=21), **VCs**: Viraemic controllers (n=6), **HVL LTNPs**: high viral load long-term non-progressors (n=12). NI-newly identified. Variant IDs highlighted in the same colour represent variants showing linkage disequilibrium. Significantly different p values between respective groups being compared are indicated by red font.

3.4 Discussion

The essential role of CD14 and TLR4 in the activation of innate immune responses by LPS is indisputable (Lu et al., 2008). Regulatory and functional polymorphisms in the genes encoding these molecules have been reported to associate with the risk and severity of various diseases (Hubacek et al., 1999, Laine et al., 2005, Rosas-Taraco et al., 2007, Noreen et al., 2012, Wu et al., 2017, Vidyant et al., 2019). In African populations, studies documenting the variation in the *CD14* and *TLR4* genes are scarce.

In this study, we have characterised the variation (SNPs and indels) and putative haplotypes of *CD14* and *TLR4* in the black South African population. Findings showed that although both allelic representation and putative haplotype structure in the South African population closely resembles that of the African-based populations (using the Yoruba and Luhya populations as representative west and east African populations, respectively), there are several distinct differences. A study that conducted full-exon sequencing of four TLR pathway genes (TLR2, TLR4, TLR6, and Toll-interleukin-1 Receptor (TIR) domain-containing adaptor protein - TIRAP) on individuals from Uganda and South Africa (mixed race, n=33; black, n=8, and white, n=7) found the Ugandan and the mixed race South African populations to be similar to the Africans (Maasai in Kinyawa, Kenya, African ancestry in Southwest USA, LWK and YRI) in the 1000 Genomes Project (Baker et al., 2012). Results are also consistent with findings from previous studies in our laboratory, which found that representation of select genetic variants in *RICH-2*, *BST-2*, and *IL-8*, in black South African populations, closely mirror those of the African Yoruba and Luhya populations (Picton et al., 2017, Dias et al., 2020, Paximadis et al., 2021). As expected, the South African black population showed significant differences in the *CD14* and *TLR4* genes when compared to the European and Asian populations, again reinforcing that data from studies conducted in these populations should not be extrapolated to South African black populations.

This study has also resulted in the description on some putative novel variants in these two genes - a newly identified *CD14* variant and eight newly identified *TLR4* variants. It is to be expected that newly identified variants will be present in low frequency, and these were all detected in single individuals (MAF: 0.6%). For newly identified variants that were not detected in the HIV-1 controller cohort, verification of these variants will be needed, either by genotyping larger cohorts, or by using alternative methods (for example Sanger sequencing) for genotyping the same individuals. In addition, using RegulomeDb, we have identified 5' and 3' flanking region variants which are ranked in category 1, suggesting that there are located in

regions of functional significance and are likely to affect expression of these molecules (Dong et al., 2022). With respect to *CD14* variants, four of the RegulomeDb rank 1 variants identified, comprised of three SNPs (rs112865855, rs5744448 and rs12108825) and one indel (rs142256657) were restricted to the African populations. These variants might therefore be involved in gene regulation in African populations and may in part be responsible for the lower expression levels of sCD14 in black compared to white individuals (Reiner et al., 2013, De Voeght et al., 2016, Olson et al., 2020). In the *TLR4* gene, all identified RegulomeDb category 1 variants found in the 5' and 3' flanking regions were common variants across populations studied, except for the rs11536897 SNP, which was rare in the SAS population. However, the MAFs of the RegulomeDb category 1 variants in the SA population were either significantly higher or lower compared to non-Africans, suggesting that these variants could still have differential effects on gene regulation in the SA population.

For both *CD14* and *TLR4* putative haplotypes, not unexpectedly, the LD patterns in the SA population more closely resembles those of the YRI and LWK populations compared to the non-African populations (EUR, EAS, and SAS). However, a *TLR4* 3' flanking region putative haplotype (*TLR4* Hap-3), comprised of six RegulomeDb category 1 variants (1b - rs11536898; 1f - rs7873784, rs11536891, rs11536896, rs17420985 and rs60103585) was found to be conserved across all populations studied, possibly indicating a functionally important haplotype. Data on *CD14* haplotype structure are scarce in the literature. A *TLR4* haplotype involving the rs4986790 (Asp299Gly) and rs4986791 (Thr399Ile) SNPs has however been extensively studied - these two *TLR4* SNPs are in strong LD in Caucasian but not in African populations (Arbour et al., 2000, Newport et al., 2004). In this study, only the rs4986790 SNP is present in the black SA population, with a MAF of 2.9%, which is not significantly different from the EUR population, but significantly lower when compared to the LWK and SAS populations. This is in agreement with findings from a study conducted on black South African children (n=405), which found the rs4986790 SNP in low frequency (4.4%) with the rs4986791 SNP detected in at a MAF of <0.5% (Kresfelder et al., 2011). In contrast, Baker et al. (2012) failed to detect both the rs4986790 and rs4986791 SNPs in a cohort of South Africans with black (n=8), white (n=7), and mixed race (n=33) individuals, but this may be due to the small sample numbers in this study.

HIV-1 controllers are a subset of PLWH that can naturally control the HIV-1 infection in the absence of ART. They are comprised of elite controllers (ECs) who naturally suppress the HIV-1 viral load (VL) to < 50 RNA copies/ml, viraemic controllers (VCs), who suppress the VL to

< 2000 RNA copies/ml, and high viral load long term non-progressors (HVL LTNPs) who can maintain CD4+ T cell counts > 500 cells/µl despite high VL (VL > 10 000 RNA copies/ml) (Gurdasani et al., 2014). The HVL LTNPs are a rare phenotype, similar to simian immunodeficiency virus (SIV)-infected sooty mangabeys who do not progress to AIDS despite high levels of viral replication (Silvestri et al., 2003, Rotger et al., 2011). HIV-1 controllers have been shown to have lower levels of immune activation and inflammation compared to progressors (Choudhary et al., 2007). Given that CD14 and TLR4 play major roles in immune activation and inflammation, we have also investigated the potential association of *CD14* and *TLR4* polymorphisms with HIV-1 control in a cohort of HIV-1 controllers from the black South African population.

A common *CD14* polymorphism (rs2569190), located in the promotor region of the gene, has been shown to increase CD14 expression. This polymorphism enhances the transcriptional activity of the *CD14* gene by decreasing the binding affinity of the GC box for the Sp3 transcription factor (LeVan et al., 2001), a transcription factor known for inhibiting promotor activity in various genes. The rs2569190 SNP has been associated with the risk and severity of various diseases (Hubacek et al., 1999, Unkelbach et al., 1999, Shimada et al., 2000, Laine et al., 2005, Rosas-Taraco et al., 2007, Wu et al., 2017). In the current study, the rs2569190 SNP was also a prevalent variant in the SA population (MAF: 33.3%) and in LD (r^2=0.9) with another SNP located in the 5' flanking region (rs2569191). Interestingly, this 2-SNP putative haplotype was found in all populations we assessed from the 1000 Genomes Project. The present study, however, found no differences in representation of the rs2569190 SNP between the SA controls and HIV-1 controllers, or the HIV-1 controller subgroups. Therefore, this variant either does not associate with HIV-1 control in this population group, or the study is underpowered to detect the association.

An overrepresentation of both MAF (7.1%) and heterozygosity (14.1%) of the *CD14* rs186291587 SNP was found in elite controllers compared to SA controls (0.6% and 1.1%, respectively). The rs186291587 SNP is a rare variant, absent in the LWK, EUR and EAS populations, and found in relatively low frequencies in the YRI (0.5%) and SAS (0.1%) populations. Rare variants are defined as variants with a MAF less than 1%, and these variants are predicted to have larger phenotypic effects than those of common variants (MAF>5%) (Chattopadhyay and Lu, 2020). To the best of our knowledge, no previous studies have found disease-related associations for the rs186291587 SNP or associations with CD14 expression levels (cell surface or soluble). The rs186291587 SNP is located in a CTCF binding site. CTCF

is a versatile zinc-finger transcription factor that can act as either a transcription activator or repressor when bound close to gene promotors (Lobanenkov et al., 1990, Burton et al., 2002), and functions as an enhancer blocker (Bell et al., 1999). Given the important functions of the CTCF binding site, variants affecting this site could potentially affect gene expression. Overall, results suggest differential regulation of the *CD14* gene in elite controllers compared to controls. Further studies would need to be conducted to elucidate the role of this SNP, its effect on CD14, and its potential role in HIV-1 control.

The association of two nonsynonymous *TLR4* SNPs (Asp299Gly, rs4986790; and Thr399Ile, rs4986791) with a range of infectious and non-infectious diseases, including HIV-1, has been well documented (Noreen et al., 2012, Vidyant et al., 2019). Several studies have shown that the rs4986790 SNP, rather than the rs4986791 SNP, is responsible for the structural changes in the TLR4 extracellular domain, resulting in reduced LPS signalling (Arbour et al., 2000, Ohto et al., 2012, Long et al., 2014). These TLR4 polymorphisms have been associated with susceptibility to acquiring HIV-1 infection in Indians (Vidyant et al., 2019), high viral loads (Pine et al., 2009), late $CD4^+$ T-cell recovery after ART (Yong et al., 2016), and an increased risk of serious illnesses in PLWH (Papadopoulos et al., 2010). No studies have investigated the representation of TLR4 polymorphisms in HIV-1 controllers. In this study, we did not find differences in the representation of the rs4986790 SNP between HIV-1 controllers and SA controls. Again, as mentioned above for CD14, the failure to see differences in the representation of the rs4986790 SNP in this study could be related to differences in ethnicities between the studies or an underpowered study. The *TLR4* 3'UTR SNP, rs11536889, which has been shown to affect TLR4 expression levels on monocytes, and to associate with a range of diseases (Sato et al., 2012), is absent in our black South African population – both SA controls and HIV-1 controllers.

While no differences in representation of the extensively studied TLR4 variants were found among the HIV-1 controllers relative to controls, several other *TLR4* variants (SNPs and indels) were significantly overrepresented in HIV-1 controllers, and interestingly, these were predominantly located in intronic sequences. The bulk of the DNA sequences in human genes are made up of introns (Rigau et al., 2019). Introns play an important role in the regulation of gene expression, alternative splicing, and mRNA export (Sorek and Ast, 2003, Valencia et al., 2008, Beaulieu et al., 2011, Rigau et al., 2019). Variants in the intronic regions can affect gene expression and studies have reported associations between intronic variants and diseases (Rigau et al., 2019, Nair et al., 2021). Although some intronic SNPs may not have functional

consequences, they may be in LD with variants of functional significance. Several *TLR4* SNPs, including two intronic SNPs (a newly identified SNP and rs5030727) were exclusively overrepresented in HVL LTNPs when compared to SA controls. The rs5030727 SNP is a common intronic variant across the African populations (LWK and YRI), rare in the EUR population and absent in both Asian populations. The newly identified SNP (*TLR4* NI-2) was also significantly higher in HVL LTNPs compared to elite controllers, and in complete LD with the rs113017335 SNP located in the 3' UTR (discussed below). Another intronic SNP, rs78736019, also showed MAF overrepresentation in both the HVL LTNPs and in the total controller group compared to the SA controls, and yet another intronic SNP, rs115919003, was not detected in the SA controls, and showed MAF overrepresentation in the total controller group compared to the SA controls. In elite controllers, there was an overrepresentation of a common intronic deletion, rs5030644, compared to the SA controls. To the best of our knowledge, no other studies have reported disease associations with these intronic *TLR4* variants and thus these results should be regarded as tentative associations that should be explored in larger cohorts.

There was only one TLR4 coding region SNP (rs11536884 SNP), which was significantly overrepresented in HVL LTNPs compared to the controls. This SNP is a rare nonsynonymous variant that results in a leucine to phenylalanine substitution at position 385, a substitution of a non-polar amino acid to an aromatic amino acid. Given that this variant is predicted to be deleterious by predictive software, and thus could potentially alter the function of TLR4, further exploration of this variant is warranted.

The 3' UTR is a non-coding region that contains *cis*-elements and plays a key role in gene transcription through its role in localization, stability, and translation of messenger RNAs (Mayr, 2017, Mayr, 2019), and thus consequently, variants in the 3'UTR have potential effects on the regulation of gene expression. A rare TLR4 3' UTR SNP (rs113017335) was exclusively overrepresented in HVL LTNPs compared to controls. The rs113017335 SNP was absent in elite controllers, and the representation of this SNP was also significantly higher in HVL LTNPs compared to elite controllers. As mentioned above, this 3' UTR SNP is in complete LD with the intronic *TLR4* NI-2 and exhibits identical associations, and thus it remains to be determined which of the two variants in this 2-SNP putative haplotype are responsible for the associations seen. The miRNASNP-v3 database (http://bioinfo.life.hust.edu.cn/miRNASNP/#!/) (Liu et al., 2021) predicts that the

rs113017335 SNP results in a gain of binding sites of two microRNAs (miRNA, hsa-miR-1252-5p and hsa-miR-541-5p) and loss of binding sites for 4 miRNAs (hsa-miR-6755-5p, hsa-miR-7515, hsa-miR-3179 and hsa-miR-6841-5p). The effect of the loss or gain of these miRNA binding sites on the TLR4 expression have not been experimentally determined.

Another prevalent TLR4 3' UTR SNP, rs1554973, also showed higher MAF in HVL LTNPs compared to controls, and homozygosity the SNP was overrepresented in both the HVL LTNPs and the total controller group. This variant has RegulomeDb rank of 1f, suggesting that it likely to be located in a functional region and associates with lower TLR4 expression in the muscularis layer of the esophagus according to the GTEx project (https://gtexportal.org). In addition, two other 3'UTR SNPs (rs11536898 and rs60103585 heterozygosity) were underrepresented in the total group of our HIV-1 controllers compared to SA controls. The rs11536898 and rs60103585 SNPs are in LD in both SA controls (r^2=0.86) and HIV-1 controllers (r^2=1) and form part of the *TLR4* Hap-3 (together with four other variants). There are slight differences in the LD pattern for the *TLR4* Hap-3 in the SA controls and HIV-1 controllers, which results in only the two SNPs mentioned above resulting in a significant representation difference between the groups. It is interesting to note, however, that the representation of the *TLR4* hap-3 and/or any other *CD14* or *TLR4* putative haplotypes, did not differ significantly between the SA controls and the HIV-1 controllers or between the HIV-1 controller subgroups. As previously mentioned, five of the SNPs making up *TLR4* Hap-3 had a RegulomeDb rank of 1f, and the rs11536898 SNP has a rank of 1b, suggesting that these variants are in a functional region and these variants associate with higher TLR4 expression levels in various tissues according to the GTEx project. The underrepresentation of these variants in HVL LTNPs could result in lower levels of TLR4 in this group. Higher TLR4 expression levels have been associated with AIDS (Lester et al., 2008).

Apparent from the results is that most of the associations seen with respect to *TLR4* variants in the HIV-1 controllers were seen in the HVL LTNP group. Although ideally findings need to be verified in bigger cohorts, it would be nearly impossible to do so, as our cohort of HVL LTNPs is, to the best of our knowledge, among the larger cohorts (n=12) of these rare individuals, and certainly the only cohort of black HVL LTNPs that we are aware of. The current HIV-1 test and treat guidelines make it highly unlikely that individuals with this phenotype would be detected; the same applies to any prospective recruitment of elite and viraemic controllers. A potential role of TLR4 in non-progression to AIDS in sooty mangabeys has been reported (Palesch et al., 2018). A study investigating candidate genes that affect

susceptibility to AIDS in sooty mangabey and species that progress to AIDS (i.e., macaques and humans), found a TLR4 frameshift mutation in sooty mangabeys (Palesch et al., 2018). The authors suggested that the frameshift mutation, which results in reduced TLR4 responses to LPS *in vitro*, could be a factor contributing to the lower immune activation seen in SIV-infected sooty mangabeys (Palesch et al., 2018). Given that HVL LTNPs are a phenotype similar to the SIV-infected sooty mangabey model, it can be speculated that regulatory and/or functional TLR4 variants can potentially affect TLR4 activation following microbial translocation, which consequentially results in reduced immune activation, which has been reported in HVL LTNPs (Rotger et al., 2011).

It should be noted that no differences in either *CD14* or *TLR4* variants were found in viraemic controllers, this may be due to the small sample size (n=6), however, it may also be related to different mechanisms of control that are at play. Limitations to this study include low sample numbers in the HIV-1 controller subgroups, these small sample sizes are not a true representation of the bigger population, therefore the frequencies of SNPs observed, particularly rare SNPs could be inaccurate. Nonetheless, these results form a strong basis for future studies. There was no WGS conducted for individuals who were non-controllers (progressive HIV-1 infection), therefore any shifts in representation of variants in this group relative to controls or HIV-1 controllers are not accounted for. Therefore, some important variants associated with HIV-1 control may be missed. Additionally, we did not correct for multiple comparisons due to the exploratory nature of this study. Lastly, it should be mentioned that reliable detection of indels using next-generation sequencing can be problematic and given that the SA controls and HIV-1 controller groups were sequenced at two different institutes, the frequencies of the indels observed in both groups would need to be verified.

In summary, the characterisation of genetic variation in the *CD14* and *TLR4* genes was undertaken in the black South African population, contributing to the understanding of genetic diversity in African populations. The enrichment of a *CD14* SNP in elite controllers and the higher representation of several *TLR4* variants in HVL LTNPs and elite controllers was noteworthy. Our results suggest a possible role of *CD14* and *TLR4* SNPs in HIV-1 control. However, these findings need to be verified in bigger cohorts that also include individuals with progressive HIV-1 infection.

CHAPTER 4

Variation in the genes for LPS binding proteins, LBP and MD-2, in the black South African population: comparison with other populations and prevalence in HIV-1 controllers

4.1 Introduction

Lipopolysaccharide (LPS) binding protein (LBP) and myeloid differentiation factor 2 (MD-2) are accessory molecules required for LPS recognition and signalling (Lu et al., 2008). LBP is a soluble protein synthesised by hepatocytes and binds to the lipid A moiety of LPS to initiate the innate immune responses (Tobias et al., 1989, Wright et al., 1990). LBP has been shown to have both pro and anti-inflammatory effects on immune responses to LPS, depending on the circulating levels of the molecule (Gegner et al., 1995). At normal or low concentrations, LBP catalytically transfers LPS to CD14 for TLR4 signalling, but at high concentrations, LBP can inhibit LPS immune responses by clearing the LPS or LPS/CD14 complexes without the activation of the TLR4 receptor (Zweigner et al., 2001). MD-2 is a secreted glycoprotein which is found in association with TLR4 on cell surfaces, forming a TLR4-MD-2 complex (Visintin et al., 2001, Nagai et al., 2002). The binding of LPS to the MD-2 on the TLR4-MD-2 complex initiates the downstream signal transduction which results in the production of cytokines (Park et al., 2009).

LBP is encoded by the *LBP* gene located on chromosome 20 (20q11.2) (Gray et al., 1993). The *LBP* gene is approximately 28.5 kb in length and has 14 exons (Hubacek et al., 1997). Various single nucleotide polymorphisms (SNPs) in the *LBP* gene have been associated with LBP expression levels, functionality, as well as a range of diseases. An *LBP* coding region SNP, rs2232613 (+988C>T), which introduces a proteinase cleavage site resulting in a mutant protein that has reduced LPS and lipopeptide binding capabilities, has been reported to associate with reduced cytokine production after LPS stimulation and higher pneumonia and sepsis mortality rate (Eckert et al., 2013). Another nonsynonymous *LBP* SNP (rs2232618) which results in a phenylalanine to leucine substitution, has been suggested to cause spatial conformational changes to the active centre of the C–terminal domain of LBP (Zeng et al., 2012a). This variant has been associated with the risk of sepsis and multiple organ dysfunction after trauma (Zeng et al., 2012a, Lu et al., 2018). An *LBP* promotor region SNP (rs2232571) located in a transcription factor binding site (CAAT box) has been found to be associated with higher LBP plasma levels (Chien et al., 2008). This promoter SNP has been shown to be in linkage disequilibrium (LD) with other promotor SNPs (rs1780616, rs5741812, rs1780617, rs2232575, and rs2232578) as well as a synonymous SNP (rs2232582) located in exon 2 (Chien et al., 2008, Flores et al., 2009) - these SNPs have been associated with gram-negative bacteraemia and the risk of severe sepsis (Chien et al., 2008, Flores et al., 2009) .

MD-2 is encoded by the *LY96* gene located on chromosome 8 (8q21.11) (Gray et al., 2010). The *LY96* gene is approximately 38 kb in size and is comprised of five exons (Gray et al., 2010). A promotor *LY96* SNP (rs11465996, −1625C>G) has been shown to result in increased expression of *LY96* mRNA and TNF-α in whole blood upon LPS stimulation (Gu et al., 2007), and heterozygosity for this variant has been reported to associate with the reduced risk of ulcerative colitis (Bank et al., 2014). Two *LY96* gene promotor SNPs which are in LD (rs1809441 and rs1809442) have been associated with higher MD-2 mRNA expression levels (Liao et al., 2015b). In addition, a rare *LY96* nonsynonymous SNP (+103 Thr35Ala) has been reported to result in reduced NF-κβ-activation and TNF-α production upon LPS stimulation (Hamann et al., 2004). Another MD-2 nonsynonymous SNP, rs6472812 (glycine to arginine at position 56) has been shown to result in reduced LPS responsiveness (Vasl et al., 2008).

Although variants in both the *LBP* and *LY96* genes have been shown to affect LPS responsiveness, little is known about these genes in African populations. We thus undertook this study to describe the genetic variation and LD patterns in the *LBP* and *LY96* genes in black South African control individuals, using data available from select 1000 Genomes Project populations for comparative purposes (Auton et al., 2015). We also compared the representation of the *LBP* and *LY96* variants in black South African individuals who can control their HIV-1 infection without antiretroviral treatment (HIV-1 controllers), to the South African control individuals.

4.2 Methods and materials

4.2.1 Study participants

Refer to Chapter 3, section 3.2.1 for details. Briefly, whole genome sequencing (WGS) data for 87 black South African (SA) individuals (controls; absence of HIV-1 infection) was used for the characterisation of the *LBP* and *LY96* genetic variation and linkage disequilibrium analysis, and WGS data for 39 black South African HIV-1 controllers was used for comparison to the black SA controls.

4.2.2 Whole genome sequencing

Refer to Chapter 3, section 3.2.2 for details. Briefly, WGS of the 87 black South African controls was conducted by the Broad Institute Genomic Services (Massachusetts Institute of Technology, Cambridge, Massachusetts, USA) and WGS data for the 39 HIV-1 controllers was also generated commercially by Edinburgh Genomics (Edinburgh, Scotland).

4.2.3 Linkage disequilibrium analysis and Hardy-Weinberg equilibrium

Refer to Chapter 3, section 3.2.3 for details. Briefly, the Haploview software version 4.2 was used to calculate linkage disequilibrium (LD) between pairs of SNPs and deviation from Hardy-Weinberg equilibrium for the *LBP* and *LY96* genes. For this chapter variants present at minor allele frequency < 5% were excluded.

4.2.4 Data analysis

Refer to Chapter 3, section 3.2.4 for details. Briefly, Fisher's exact test was performed using an in-house developed and validated data analysis pipeline (Reveley et al, unpublished) to compare the allelic and genotypic representation of *LBP* and *LY96* variants between the SA population and the 1000 Genomes Project reference populations and the HIV-1 controller cohort. The Variant Effect Predictor (VEP) tool was used predict the consequences of nonsynonymous SNPs and RegulomeDB was used to predict the function effects of non-coding variants.

4.3 Results

4.3.1 Genetic characterisation of the *LBP* and *LY96* genes

4.3.1.1 Single nucleotide polymorphisms and indels in the *LBP* gene

Four hundred and five variants (SNPs and indels) were identified across the *LBP* gene and the flanking regions. Due to the large size of this gene, we have listed the variants in two tables; the distribution and comparison of the minor allele frequency (MAF) of the 5' and 3' flanking regions and coding regions to select populations (YRI, LWK, EUR, EAS, and SAS) from the 1000 Genomes Project are shown in **Table 4.1**, and similarly intronic variants are listed and compared in **Supplementary Table 1**. One hundred and nine of the identified variants were present in single individuals with MAF of 0.6%.

We identified 20 SNPs in the coding regions (exons) of the *LBP* gene (**Table 4.1)**. Seven of the SNPs were synonymous variants, 12 were nonsynonymous variants and one was a newly identified nonsynonymous SNP (*LBP* NI-3). Data for two synonymous variants, namely rs111616591 (Leu39=) and rs112681398 (Leu47=), were not present in the 1000 Genomes Project. Three synonymous SNPs (rs1739654, rs2232582, rs2232596) were common across the 1000 Genomes Project populations. Representation (MAF) of the rs1739654 (Leu48=) and rs2232582 (Pro97=) synonymous SNPs was significantly higher in SA population compared to LWK, EUR, EAS, and SAS populations, while the representation of the rs2232596

(Ser204=) synonymous SNP was only higher compared to the EUR and EAS populations. The rs34144735 (Thr153=) and rs2232620 (Ala445=) synonymous SNPs were found in low frequencies and exclusively in African populations. The newly identified SNP located in exon 11 is a nonsynonymous SNP resulting in a substitution of phenylalanine with valine at position 393 (Phe393Val), predicted to be 'tolerated' and 'benign' using SIFT and PolyPhen-2, respectively - this polymorphism was however only found in a single individual and thus requires verification. Four nonsynonymous SNPs, rs372496254 (Arg119His), rs146507643 (Glu167Asp), rs143256886 (Val325Ile), and rs142356536 (Lys454Asn), were also rare and found in single individuals in the SA population and absent or rare in other 1000 Genomes Project populations. Using SIFT and Polyphen-2, only the rs372496254 (Arg119His) SNP was predicted to be 'deleterious' and 'possibly damaging', respectively, while the other three SNPs, (rs146507643, rs143256886 and rs142356536) were predicted to be 'tolerated' and 'benign', respectively. Another three nonsynonymous SNPs, rs2232615 (Ile364Thr), rs2232586 (Ser157Cys) and rs142724815 (Lys439Thr), were found exclusively in African populations. The rs2232615 SNP was predicted to be 'deleterious' and 'benign', and the rs2232586 SNP was predicted to be 'deleterious' and 'probably damaging', by SIFT and Polyphen-2, respectively. The rs142724815 SNP however was predicted to be 'tolerated' and 'benign' by SIFT and Polyphen-2, respectively. Lastly, five nonsynonymous SNPs identified, namely rs2232580 (Pro9Leu), rs5744204 (Val166Met), rs2232613 (Pro333Leu), rs2232618 (Phe436Leu) and rs2232619 (Ala445Thr), were found across the 1000 Genomes Project populations. The rs5744204 and rs2232613 SNPs were predicted to be 'deleterious' and 'probably damaging', while the rs2232580, rs2232618, and rs2232619 were predicted to be 'tolerated' and 'benign' by SIFT and Polyphen-2, respectively.

Most of the variants detected in the *LBP* gene (335 of 405, **Supplementary Table 1**) were located in the intronic regions of the gene, with four newly identified SNPs present in single individuals (MAF of 0.6%) - these newly identified SNPs would thus require verification using alternative methods. Twenty-six indels (10 insertions and 16 deletions), were identified in the introns of the *LBP* gene. Comparisons of the MAFs of the intronic variants to select populations from the 1000 Genomes Project, showed high similarity between the SA population and the African YRI and LWK populations, and pronounced differences compared to the non-African EUR, EAS and SAS populations.

We identified 23 SNPs and one indel in the 5'flanking region of the *LBP* gene (**Table 4.1)**. The identified indel (rs994320874) was an insertion present in a single individual and data for this

indel were not present in the 1000 Genomes Project populations. Six of the 5' flanking region SNPs identified were present in single individuals and data for one on these SNPs (rs6025023) were not present in the 1000 Genomes Project populations. Comparisons of the MAFs of the 5' flanking region SNPs between the SA population and 1000 Genomes Project populations revealed similarities to the African YRI and LWK populations and extensive differences to the non-African populations, as was seen with the intronic variants. In the 3' flanking region of the *LBP* gene, 21 SNPs and five indels were detected. The indels were comprised of three insertions (rs1357622257, rs35214332 and rs2083977646) and two deletions (rs1489855389, and rs2083978958). No data were available on the 1000 Genomes Project for all five indels, and only two of these indels were rare (present in single individuals) – this may point to difficulty with indel identification using next generation sequencing (NGS), and possibly to using different NGS platforms. Five of the 21 SNPs in 3' flanking region SNPs were found in single individuals, and data for four of these SNPs were not available for the 1000 Genomes Project populations. What is interesting is that comparison of MAFs of the SA population 3' flanking regions SNPs showed that, for most of the SNPs where data are available, the SA population is similar to the African LWK population and significantly different to African YRI population as well as to the non-African populations.

Using the RegulomeDB database to predict the effect of the *LBP* 5' and 3' flanking region SNPs and indels on *LBP* gene regulation, we found two 5' flanking region SNPs, namely rs2232575 and rs1780617, to have RegulomeDB ranks of 1a and 1b, respectively, and ten 5' flanking region SNPs (rs5741811, rs6014785, rs5741812, rs2232571, rs2232572, rs2232573, rs2232577, rs2232578, and rs2232579) to have a rank of 1f. While most of these SNPs were present across the 1000 Genomes Project populations, the rs6014785, rs2232577 and rs2232577 SNPs were largely restricted to African populations. In the 3' flanking region, we found two insertions with RegulomeDb ranks of 1b (rs1357622257) and 1f (rs35214332). Data for these two indels were not available in the 1000 Genomes Project.

Table 4.1: Minor allele frequencies of *LBP* genetic variants (regulatory and coding regions) in the South African population compared to select populations from the 1000 Genomes Project.

Location on gene	Position (Chr:20)	Accession number	Base change (wt/mut)	n (allele frequency, %)					
				SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
5' flanking region	38344442	rs5744197	G/T	10 (5.7)	0	0	0	0	0
	38344492	rs111768929	C/T	8 (4.6)	-	-	-	-	-
	38344539	rs1780616	C/T	72 (41.4)	83 (38.4)	68 (34.3)	361 (35.9)	449 (44.5)	328 (33.5)
	38344739	rs111403915	C/G	8 (4.6)	-	-	-	-	-
	38344907	rs6025023	G/A	1 (0.6)	-	-	-	-	-
	38344969	rs189030573	C/T	1 (0.6)	0	1 (0.5)	0	1 (0.1)	0
	38345068	rs5744198	C/T	5 (2.9)	10 (4.6)	3 (1.5)	1 (0.1)	0	0
	38345281	rs5741811	T/A	4 (2.3)	7 (3.2)	8 (4.0)	79 (7.9)	113 (11.2)	125 (12.8)
	38345297	rs369592810	G/A	1 (0.6)	1 (0.5)	2 (1.0)	0	0	0
	38345426	rs994320874	**G/GA**	1 (0.6)	-	-	-	-	-
	38345528	rs6014785	G/T	53 (30.5)	70 (32.4)	55 (27.8)	2 (0.2)	0	0
	38345596	rs5741812	A/T	79 (45.4)	64 (29.6)	54 (27.3)	125 (12.4)	471 (46.7)	272 (27.8)
	38345598	rs11536934	T/C	9 (5.2)	18 (8.3)	4 (2.0)	1 (0.1)	0	0
	38345681	rs2232571	T/C	66 (37.9)	69 (31.9)	68(34.3)	855 (85.0)	706 (70.0)	668 (68.3)
	38345688	rs551546824	G/A	1 (0.6)	0	0	0	0	1 (0.1)
	38345731	rs2232572	T/A	29 (16.7)	24 (11.1)	22 (11.1)	15 (1.5)	0	32 (3.3)
	38345754	rs1780617	G/A	53 (30.5)	104 (48.1)	109 (55.1)	881 (87.6)	712 (70.6)	718 (73.4)
	38345891	rs2232573	T/C	59 (33.9)	54 (25.0)	41 (20.7)	16 (1.6)	0	41 (4.2)
	38346109	rs2232575	C/G	20 (11.5)	32 (14.8)	25 (14.8)	151 (15.0)	238 (23.8)	311 (31.8)
	38346155	rs2232576	C/A	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38346194	rs2232577	G/T	56 (32.2)	49 (22.7)	49 (24.7)	0	0	0
	38346312	rs2232578	A/G	80 (46.0)	88 (40.7)	67 (33.8)	167 (16.6)	240 (23.8)	342 (35.0)
	38346380	rs6025037	C/T	1 (0.6)	2 (0.9)	8 (4.0)	0	0	0
	38346480	rs2232579	C/T	42 (24.1)	37 (17.1)	34 (17.2)	0	0	0
Exon 1	38346542	rs2232580	C/T	2 (1.1)	5 (2.3)	8 (4.0)	79 (7.9)	68 (6.7)	115 (11.8)
	38346631	rs111616591	C/T	8 (4.6)	-	-	-	-	-
Exon 2	38349564	rs112681398	A/G	9 (5.2)	-	-	-	-	-
	38349567	rs1739654	A/G	53 (30.5)	48 (22.2)	41 (20.7)	43 (4.3)	166 (16.5)	141 (14.4)
Exon 3	38350862	rs2232582	T/C	83 (47.7)	95 (44.0)	65 (32.8)	133 (13.2)	74 (7.3)	232 (23.7)
	38350927	rs372496254	G/A	1 (0.6)	0	0	0	0	0
Exon 4	38354374	rs34144735	A/G	5 (2.9)	6 (2.8)	0	0	0	0
	38354385	rs2232586	C/G	2 (1.1)	6 (2.8)	10 (5.1)	0	0	0
	38354411	rs5744204	A/G	74 (42.5)	38 (59.3)	120 (60.6)	993 (98.7)	1008 (99.9)	956 (97.8)
	38354416	rs146507643	G/T	1 (0.6)	0	0	0	0	0
Exon 6	38360727	rs2232596	A/G	69 (39.7)	82 (38.0)	71 (35.9)	513 (51)	119 (11.8)	340 (34.8)
Exon 9	38366820	rs143256886	G/A	1 (0.6)	2 (0.9)	0	0	0	0
Exon 10	38369011	rs2232613	C/T	13 (7.5)	14 (6.5)	27 (13.6)	73 (7.3)	1 (0.1)	43 (4.4)
	38369104	rs2232615	T/C	10 (5.7)	5 (2.3)	10 (5.1)	0	0	0
Exon 11	38370765	*LBP* NI-3	T/G	1 (0.6)	-	-	-	-	-
Exon 13	38373117	rs2232618	T/C	15 (8.6)	41 (19.0)	24 (12.1)	73 (7.3)	57 (5.7)	167 (17.1)
	38373127	rs142724815	A/C	3 (1.7)	2 (0.9)	2 (1.0)	0	0	0
Exon 14	38373945	rs2232619	G/A	26 (14.9)	24 (11.1)	34 (17.2)	11 (1.1)	1 (0.1)	24 (2.5)
	38373947	rs2232620	C/T	3 (1.7)	7 (3.2)	1 (0.5)	0	0	0
	38373974	rs142356536	G/T	1 (0.6)	0	1 (0.5)	0	0	0
3' flanking region	38376859	rs61269129	G/A	2 (1.1)	15 (6.9)	1 (0.5)	0	1 (0.1)	0
	38376894	rs745144	C/T	27 (15.5)	50 (23.1)	28 (14.1)	445 (44.2)	11 (1.1)	187 (19.1)
	38377131	rs146547376	C/T	1 (0.6)	0	3 (1.5)	0	0	0
	38377248	rs60146510	C/T	2 (1.1)	15 (6.9)	1 (0.5)	0	1 (0.1)	0
	38377270	rs73909023	T/A	8 (4.6)	27 (12.5)	18 (9.1)	75 (7.5)	60 (6.0)	171 (17.5)
	38377303	rs1256925873	T/A	1 (0.6)	-	-	-	-	-
	38377381	rs111621628	A/C	8 (4.6)	0	0	0	0	0
	38377544	rs1333186270	G/A	1 (0.6)	-	-	-	-	-
	38377677	rs151021835	G/C	2 (1.1)	3 (1.4)	2 (1.0)	0	0	0
	38377723	rs73287301	C/T	2 (1.1)	15 (6.9)	1 (0.5)	0	1 (0.1)	0
	38377778	rs73289105	G/C	2 (1.1)	15 (6.9)	1 (0.5)	0	1 (0.1)	0
	38377815	rs73289108	A/G	2 (1.1)	15 (6.9)	1 (0.5)	0	1 (0.1)	0
	38378041	rs531229970	C/A	2 (1.1)	0	0	0	0	0
	38378078	rs1357622257	**A/ACTTCCT**	7 (4.0)	-	-	-	-	-
	38378304	rs182169009	A/G	39 (22.4)	-	-	-	-	-
	38378343	rs35214332	A/AT	1(0.6)	-	-	-	-	-
	38378369	rs368426777	G/A	17 (9.8)	27 (12.5)	19 (9.6)	79 (7.9)	212 (21.0)	175 (17.9)
	38378405	rs774980081	C/T	36 (20.7)	-	-	-	-	-
	38378526	rs1426749530	A/C	2 (1.1)	-	-	-	-	-
	38378538	rs566520973	C/T	1 (0.6)	0	0	0	0	0
	38378750	rs997158205	G/A	1 (0.6)	-	-	-	-	-
	38378927	rs1391141106	C/T	3 (1.7)	-	-	-	-	-
	38378931	rs1284294001	A/C	2 (1.1)	-	-	-	-	-
	38378939	rs2083977646	**A/INDEL-1**	1(0.6)	-	-	-	-	-
	38378960	rs1489855389	**INDEL-2/G**	5 (2.9)	-	-	-	-	-
	38379009	rs2083978958	**INDEL-3/G**	63 (36.2)	-	-	-	-	-

Wt/mut: wildtype/mutant referring to major and minor alleles, respectively; **SA:** South Africa; **YRI:** Yoruba in Ibadan, Nigeria; **LWK:** Luhya from Webuye, Kenya; **EUR:** European population; **EAS:** East Asians population; **SAS:** South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); **-:** No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic). NI-3: *LBP* newly identified SNP 3. Indels are indicated with bold base changes.

INDEL-1 ACCCCGTCCGGGAGGGAGGTGGGGGGGGTCAGCCCCCCGCCCGGCCAGCCG
INDEL-2 GGGGGGTCAGCCCCCCGCCCGGCCAGCCGCCCCGTCCGGGAGGGAGGT
INDEL-3 GGGGGGTCAGCCCCCCGCCCGGCCAGCCGCCCCGTCCGGGAGGGAGGT.

4.3.1.2 Single nucleotide polymorphisms and indels in the *LY96* gene

Three hundred and forty-four variants were identified in the *LY96* gene and flanking regions. Sixty-six of these variants were found in single individuals with MAFs of 0.6%. The distribution and comparison of the MAFs to 1000 Genomes Project populations are shown in **Table 4.2** (5' and 3' flanking regions and coding regions) and **Supplementary Table 2** (introns). One synonymous SNP, rs1195824354 (His62=), present in a single individual, was identified in exon 2 of the *LY96* gene - no data were available for this variant in the 1000 Genomes Project.

As seen with the *LBP* gene, the majority of the *LY96* variants (318 of 344, **Supplementary Table 2**) were found in the intronic regions of the gene. We identified 10 novel variants in the intronic regions (*LY96* NI-1 to NI-10). Nine of the newly identified variants, comprised of three insertions, one deletion and five SNPs, were present in single individuals (MAF=0.6%), and thus will need to be verified. The newly identified SNP (NI-9) at chromosome position 8:74017606, causing a C to T change, was however present at a MAF of 2.3%. In total, 40 indels (19 deletions and 21 insertions) were detected in the *LY96* intronic regions. As seen with the *LBP* gene, comparison of the MAFs of all intronic region variants showed similarities between the SA population and the African YRI and LWK populations, and extensive differences to the non-African populations.

In the 5' flanking region (**Table 4.2**), seven SNPs and one deletion (rs200524844) were identified. Representation of the rs200524844 deletion did not differ significantly between the SA population and the African YRI and LWK populations, but MAF was significantly higher in the SA population (3.4%) compared to non-African populations (MAFs < 0.3%). The rs190775697 SNP located in the promotor region of the *LY96* gene was found exclusively in African populations. Data from the 1000 Genomes Project were not available for two 5' flanking region SNPs, both low prevalence SNPs, one present in a single individual (rs1317371076) and another (rs571561117) present in two individuals (MAF: 1.1%). Overall, no significant differences were observed in the MAFs of the 5'flanking region variants when the SA population was compared to the African YRI and LWK populations, however significant differences were detected for most of the non-African population comparisons. Using the RegulomeDB database to predict the functional effect of the *LY96* 5' flanking region SNPs and indels, we found three SNPs (rs1809442, rs1809441, and rs1809440) which are common across the 1000 Genomes Project populations, to have a Regulome DB rank of 1f.

Sixteen SNPs and one deletion (rs1277364319) were identified in the 3' flanking region of the *LY96* gene **Table 4.2**. The deletion (rs1277364319) was found in a single SA individual, and no 1000 Genomes Project data were available for this deletion. Three 3' flanking region SNPs (rs553408466, rs111949004, and rs935408891) were found in single individuals, and the rs553408466, and rs111949004 SNPs were either absent or found in relatively low frequencies in the other populations, with no data for the rs935408891 SNP in the 1000 Genomes Project. Comparisons of the MAFs of the 3' flanking region variants showed that although the SA population showed the most similarity to the African YRI and LWK populations, there were still variants that showed significant differences to these populations – as example the rs7013457 SNP has significantly lower MAF in the SA population (MAF: 20.7%) compared to all other populations studied. Using RegulomeDB to predict the functional effect of the *LY96* 3' flanking region SNPs and indels, revealed the rs11466007 SNP, almost uniquely African (detected in one EUR population individual only), with a rank of 1b, and four SNPs (rs4737403, rs6994774, rs7013457, rs4738415), common across the 1000 Genomes Project populations, with ranks of 1f.

Table 4.2: Minor allele frequencies of *LY96* genetic variants (regulatory and coding regions) in the South African population compared to select populations from the 1000 Genomes Project.

Location on gene	Position (Chr:8)	Accession number	Base change (wt/mut)	n(minor allele frequency, %)					
				SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
5' flanking region	73989727	rs11465996	C/G	24 (13.8)	28 (13.0)	19 (9.6)	289 (28.7)	193 (19.1)	286 (29.2)
	73990084	rs1317371076	A/G	1 (0.6)	-	-	-	-	-
	73990151	rs1809442	G/C	69 (39.7)	85 (39.4)	72 (36.4)	547 (54.4)	667 (66.2)	482(49.3)
	73990178	rs1809441	T/G	76 (43.7)	86 (38.8)	81 (40.9)	547 (54.4)	667 (66.2)	482(49.3)
	73990288	rs1809440	G/A	79 (45.4)	106 (49.1)	88 (44.4)	547 (54.4)	669 (66.4)	482(49.3)
	73990590	rs200524844	**GT/G**	6 (3.4)	4 (1.9)	12 (6.1)	1 (0.1)	2 (0.2)	0
	73990670	rs571561117	C/T	2 (1.1)	-	-	-	-	-
	73991299	rs190775697	A/C	4 (2.3)	1 (0.5)	1 (0.5)	0	0	0
Exon 2	74004869	rs1195824354	C/T	1 (0.6)	-	-	-	-	-
3' flanking region	74029071	rs10282832	A/C	31 (17.8)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
	74029115	rs73337540	G/C	2 (1.1)	2 (0.9)	0	0	0	0
	74029299	rs4737403	A/T	27 (15.5)	16 (7.4)	21 (10.6)	36 (3.6)	22 (2.2)	37 (3.8)
	74029439	rs11466007	C/A	29 (16.7)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
	74029579	rs6994774	T/C	59 (33.9)	64 (29.6)	74 (37.4)	147 (14.6)	109 (10.8)	97 (9.9)
	74029656	rs111467417	C/T	11 (6.3)	12 (5.6)	9 (4.5)	0	0	0
	74029728	rs147893820	C/T	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74029809	rs11993601	G/A	23 (13.2)	23 (10.6)	11 (5.6)	1 (0.1)	0	0
	74029843	rs7013457	A/G	36 (20.7)	69 (31.8)	64 (32.2)	822 (81.7)	867 (86.0)	826 (84.5)
	74029915	rs113598965	G/C	2 (1.1)	2 (0.9)	0	0	0	0
	74030001	rs7013662	T/G	70 (40.2)	207 (55.1)	96 (48.5)	823 (81.8)	867 (86.0)	826 (84.5)
	74030696	rs553408466	C/T	1 (0.6)	0	0	0	0	1 (0.1)
	74030789	rs4738415	T/A	36 (20.7)	69 (31.9)	64 (32.3)	822 (81.7)	867 (86.0)	826 (84.5)
	74030875	rs111949004	G/A	1 (0.6)	0	0	4 (0.4)	0	4 (0.4)
	74030906	rs78519436	C/T	4 (2.3)	7 (3.2)	5 (2.5)	0	0	0
	74031049	rs1277364319	**TG/T**	1 (0.6)	-	-	-	-	-
	74031061	rs935408891	C/A	1 (0.6)	-	-	-	-	-

Wt/mut: wildtype/mutant referring to major and minor alleles, respectively; **SA:** South Africa; **YRI:** Yoruba in Ibadan, Nigeria; **LWK:** Luhya from Webuye, Kenya; **EUR:** European population; **EAS:** East Asians population; **SAS:** South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); **-:** No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic). Indels are indicated with bold base changes.

4.3.2 Linkage disequilibrium and identification of putative haplotypes in the *LBP* and *LY96* genes

The large sizes of the *LBP* and *LY96* genes and the high number of variants present in these genes resulted in highly complex LD patterns and the identification of a very large number of putative haplotypes. Thus, for the purpose of this study, we decided to apply more selective (stringent) criteria in the linkage disequilibrium analysis of these two genes – only putative haplotypes comprised of variants with MAFs > 5% showing complete LD (r^2=1) will be described in this chapter. For purposes of comparison to the 1000 Genomes Project populations however, variants making up the respective SA population putative haplotypes were accessed and analysed using r^2 values > 0.65 (i.e., although only complete LD putative haplotypes (r^2=1) in the SA population were studied, when evaluating the absence or presence of the haplotype in the 1000 Genomes Project populations, less stringent criteria were applied).

A total of 26 putative haplotypes were identified across the *LBP* gene and flanking regions. Twelve of these putative haplotypes were comprised of three or more variants (range: 3 to 24; named *LBP* Hap-1 to *LBP* Hap-12 in order of decreasing complexity/variant number), and the remaining 14 putative haplotypes were each comprised of two variants (named *LBP* 2-SNP Hap-1 to *LBP* 2-SNP Hap-14). **Table 4.3** shows the composition of each putative haplotype, i.e., the variants in complete LD for each haplotype. The most complex (largest) of the putative haplotypes identified in the SA population is comprised of 21 SNPs and 3 indels (*LBP* Hap-1), all located in intronic regions. Comparison to 1000 Genomes Project populations revealed that data for nine of the 24 *LBP* Hap-1 variants were not available for all populations studied. However, analysis of the data available for the remaining 15 variants revealed that these variants were virtually absent in these populations and thus haplotype prediction could not be performed. *LBP* Hap-1 on the other hand is a relatively prevalent putative haplotype in the SA population (HF: 6.9%), and for now appears to be unique to this population. While most of the putative haplotypes identified were comprised of intronic variants, *LBP* Hap-3, and *LBP* 2-SNP-Hap-13 were comprised of a combination of intronic and exonic variants. *LBP* Hap-3 was present (r^2>0.67) across all populations studied (YRI, LWK, EUR, EAS, and SAS populations), while the *LBP* 2-SNP-Hap-13 was present in all populations (r2>0.68) except for the EAS population.

Variants making up *LBP* Hap-2, *LBP* Hap-8, *LBP* Hap-10, *LBP* 2-SNP Hap-9 and *LBP* 2-SNP Hap-10 were absent in non-African populations. *LBP* Hap-2 was only detected in LWK (r^2>0.66), and *LBP* Hap-8 was only detected in YRI (r^2=1), while *LBP* Hap-10, *LBP* 2-SNP Hap9 and *LBP* 2-SNP Hap-10 were present in both African populations (r^2 > 0.70). *LBP* 2-SNP Hap-4 was only present in YRI and LWK populations (r^2=0.96 for both), while *LBP* 2-SNP Hap-11 was only present in the YRI population (r^2=1), even though the variants comprising both these two putative haplotypes were present across all 1000 Genomes Project populations studied. *LBP* Hap-12 was present in YRI and LWK populations (r^2=1 for both) and the EUR population (r^2=1). *LBP* Hap-4, *LBP* 2-SNP Hap-6, and *LBP* 2-SNP Hap-7 were present (r^2>0.68) in all populations except the EAS population. Data were not available in the 1000 Genomes Project for the *LBP* 2-SNP Hap-14 rs2076840912 SNP. The remaining 11 putative haplotypes (i.e., *LBP* Hap-5, *LBP* Hap-6, *LBP* Hap-7, *LBP* Hap-9, *LBP* Hap-11, *LBP* 2-SNP Hap-1, *LBP* 2-SNP Hap-2, *LBP* 2-SNP Hap-3, *LBP* 2-SNP Hap-5, *LBP* 2-SNP Hap-8, and *LBP* 2-SNP Hap-12) were present across all 1000 Genomes Project populations (r^2 > 0.65).

Twenty-two putative haplotypes were identified across the *LY96* gene and flanking regions. Fourteen of these putative haplotypes were comprised of three or more variants (range: 3 to 8 variants; named *LY96* Hap-1 to *LY96* Hap-14 in order of decreasing complexity/variant number), and the remaining eight putative haplotypes were each comprised of two variants (named *LY96* 2-SNP Hap-1 to *LY96* 2-SNP Hap-8). **Table 4.4** shows the composition of each putative haplotype, i.e., the variants in complete LD for each putative haplotype. Data were not available in the 1000 Genomes Project populations for two putative haplotypes, namely *LY96* 2-SNP Hap-7 and *LY96* 2-SNP Hap-8. Interestingly, the two-variant putative haplotype, *LY96* 2-SNP Hap-8, was comprised of two deletions (rs1406963683 CTT/C and rs1161163048 CCTTCCTTCCT/C) and was a fairly prevalent putative haplotype in the SA population (HF: 5.2%). The variants making up the *LY96* Hap-1, *LY96* Hap-7, *LY96* Hap-8, *LY96* Hap-12 and *LY96* 2-SNP Hap-6 putative haplotypes were absent or found in frequencies insufficient for linkage disequilibrium analysis in non-African populations (EUR, EAS, and SAS), however these putative haplotypes were present across the two African populations (YRI and LWK) with strong LD between the variants ($r^2 > 0.75$). All remaining putative haplotypes were conserved across all the 1000 Genomes Project populations (YRI. LWK, EUR, EAS, and SAS) showing strong LD ($r^2 > 0.80$) between variants for all the putative haplotypes except for *LY96* Hap-2, where the LD ($r^2=0.66$) was weaker between certain SNPs in the SAS population.

Table 4.3: Putative *LBP* gene haplotypes (r^2=1) in the South African population.

Haplotype ID (*LBP*)	List of variants in haplotype			
Hap-1	rs113464344	rs113966205	rs111874373	rs113489543
	rs113579225	rs112318889	rs113031135	rs111848745
	rs113734860	rs113203674	rs112685620	rs112098655
	rs113366471	**rs1289322656**[2]	rs112052603	rs112924468
	rs113207369[1]	**rs145688223**[2]	rs112314157	rs113915430
	rs113525611	rs112652150	rs112038028	rs533481732
Hap-2	rs6025188	rs6025195	rs11536978	rs6025198
	rs6014840			
Hap-3	rs2232617	rs6069931	rs11536993	**rs2232618**[4]
	rs11536994			
Hap-4	rs1631643	rs1739638	rs12624843	rs1780628
Hap-5	rs6123629	rs2232595	rs6127861	rs182432670
Hap-6	rs11906594	**rs137996210**[2]	rs6025098	
Hap-7	rs2232593	rs6127841	rs2232592	
Hap-8	rs139715955	rs11536963	rs140999190	
Hap-9	rs6014862	rs11086581	rs2232616	
Hap-10	rs2232609	rs5741818	rs6025196	
Hap-11	rs2232590	rs2232589	rs41332344	
Hap-12	rs147651643	rs116538384	rs145503460	
2-SNP Hap-1	rs6127831	rs6127837		
2-SNP Hap-2	rs1780627	rs1739640		
2-SNP Hap-3	rs6099236	**rs2232578**[3]		
2-SNP Hap-4	rs6014804	rs6014805		
2-SNP Hap-5	rs1780624	rs22325961		
2-SNP Hap-6	rs6099228	rs8120088		
2-SNP Hap-7	rs7267449	rs7267600		
2-SNP Hap-8	**rs11481047**[2]	rs6014861		
2-SNP Hap-9	rs16987261	rs116743201		
2-SNP Hap-10	rs140501730	rs11536951		
2-SNP Hap-11	rs140069602	rs11086565		
2-SNP Hap-12	rs6127849	rs6127855		
2-SNP Hap-13	rs11536949	**rs2232613**[4]		
2-SNP Hap-14	rs111067091	**rs2076840912**[1]		

[1]: deletion; [2]: insertion; [3]: 5' flanking region variant; [4]: exon variant (all indicated in bold font); all other variants are intronic.

Table 4.4: Putative *LY96* gene haplotypes (r^2=1) in the South African population.

Haplotype ID (*LY96*)	List of variants in haplotype			
Hap-1	rs112541809	**rs147765065**[1]	rs73335866	rs112793867
	rs11993601[3]	rs147251489	rs146530931	rs73335869
Hap-2	rs140425838	rs73335860	rs111931925	rs17226734
	rs11403178[2]	**rs4737403**[3]	rs187932458	rs4737399
Hap-3	rs7821865	rs7001511	rs4738412	rs7829369
	rs11465996[4]	rs11466001	rs62509331	rs72661869
Hap-4	**rs4738415**[3]	**rs7013457**[3]	rs6472813	rs2114169
	rs7838114	rs7007114	rs7819708	
Hap-5	rs62509332	rs56343979	rs12546519	**rs138140428**[1]
	rs12546552	rs7388563		
Hap-6	rs7820086	rs77754189	rs56230757	rs113066143
	rs55917243			
Hap-7	rs10088782	rs113459347	rs113548678	rs10092443
	rs10107496			
Hap-8	rs185952327	**rs10282832**[3]	rs190725508	rs73335834
	rs16938762			
Hap-9	rs4738414	**rs7013662**[3]	rs6993513	rs7839553
Hap-10	rs11786591	rs11783456	**rs34767907**[2]	rs4738413
Hap-11	rs4738411	rs7009353	rs55722404	rs10808798
Hap-12	rs11466000	rs77952155	rs116507106	rs115211571
Hap-13	**rs373090437**[2]	rs112589509	rs16938760	
Hap-14	**rs11409980**[2]	rs28718132	rs28435175	
2-SNP Hap-1	rs7841862	rs60515500		
2-SNP Hap-2	rs73335845	rs12542066		
2-SNP Hap-3	rs16938758	rs16938759		
2-SNP Hap-4	rs7822407	rs10086976		
2-SNP Hap-5	rs113532117	rs80343839		
2-SNP Hap-6	rs113382383	rs73335849		
2-SNP Hap-7	rs546722838	rs113037498		
2-SNP Hap-8	**rs1406963683**[1]	**rs1161163048**[1]		

[1]: deletion; [2]: insertion; [3]: 5' flanking region variant; [4]: exon variant (all indicated in bold font); all other variants are intronic.

4.3.3 Comparisons of the representation of the *LBP* and *LY96* variants between South African controls and HIV-1 controllers

4.3.3.1 *LBP*

Comparing the representation of all *LBP* variants identified between HIV-1 controllers and SA controls revealed several SNPs to be significantly different between the groups **(Table 4.5)**. Minor allele frequency (p=0.011, OR=34.6, CI: 2.887-414.699) and the heterozygosity (p=0.01, OR=43, CI: 3.189-579.773) for the rs6025023 SNP located in the 5' flanking region of the *LBP* gene were significantly higher in viraemic controllers compared to SA controls. In the 3' flanking region, MAF and heterozygosity for the rs111621628 SNP were higher in both the total group of HIV-1 controllers (MAF: p=0.017, OR=3.407, CI=1.312-8.843; heterozygosity: p=0.013, OR=3.879, CI: 1.416-10.626) and elite controllers (MAF: p=0.034, OR=3.458, CI: 1.130-10.580; heterozygosity: p=0.028, OR=3.95, CI: 1.197-13.035) compared to the SA controls.

In the coding region (exon 2) of the *LBP* gene, we found an underrepresentation of the heterozygous genotype of a synonymous SNP (rs1739654) in viraemic controllers compared to SA controls (p=0.031, OR=0.072, CI: 0.007-0.786). This variant was in strong LD (r^2=0.98) with an intronic SNP (rs1739655) which showed a similar association (**Table 4.5**). Minor allele frequency and heterozygosity for the nonsynonymous SNP in exon 13 (rs142724815, Lys439Thr) were higher in viraemic controllers compared to SA controls (MAF: p=0.034, OR=11.4, CI: 1.706-76.178; heterozygosity: p=0.032, OR=14, CI: 1.8-108.882). This variant showed strong LD (r^2 > 0.87) with three intronic SNPs (rs141862127, rs145588209, and rs140231834) which also showed similar associations (**Table 4.5**).

Comparison of the *LBP* intronic variants between the groups revealed an overrepresentation of the rs11536943 SNP (both MAF and heterozygosity) in the total group of HIV-1 controllers (p=0.029 for both) and in elite controllers (MAF: p=0.037; heterozygosity: p=0.036) compared to the SA controls which had zero representation of this SNP. The rs11536943 SNP is in LD (r^2>0.98) with two intronic SNPs (rs960451887 and rs113577957), thus showing similar associations (**Table 4.5**). Interestingly, the rs1250247980 SNP, which was absent in the SA control population was found in all three controller subgroups, and MAF was significantly higher in both the total group of HIV-1 controllers (p=0.003) and viraemic controllers (p=0.0002) compared to SA controls. Additionally, we observed an overrepresentation of the minor allele of the rs140069602 SNP in viraemic controllers compared to SA controls (p=0.024, OR=8, CI: 1.351-47.376). Two additional intronic SNPs showed minor allele

underrepresentation in the total group of HIV-1 controllers compared to the SA controls (rs6025098: p=0.036, OR=0.384, CI=1.153-0.964; rs6014805: p=0.024, OR=0.498, CI: 0.278-0.892). Minor allele homozygosity of the rs6014805 SNP (T/T) was also underrepresented in the total group of HIV-1 controllers (p=0.022) and elite controllers (p=0.045) compared to the SA controls. The rs6025098 SNP is in LD (r^2=1) with the intronic rs137996210 indel and rs11906594 SNP, respectively. These are the variants comprising *LBP* Hap-6 and thus all showed similar associations (**Table 4.5**).

No significant differences were observed in the representation of both allelic and genotypic frequencies of the *LBP* variants between the HIV-1 controller subgroups.

Table 4.5: Comparison of allelic and genotypic frequencies of *LBP* variants between SA controls and HIV-1 controllers.

			Minor allele / genotype frequency					HICs vs SA controls			ECs vs SA controls			VC vs SA controls			HVL LTNPs vs SA controls		
Location on gene	**Variant**	**Minor allele/genotype**	**SAs**	**HICs**	**ECs**	**VCs**	**HVL LTNPs**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**
5' flanking region	rs6025023	**A**	1(0.6)	3(3.8)	1(2.4)	2(16.7)	0	6.92	0.708-67.613	0.089	4.22	0.258-68.883	0.352	34.6	2.887-414.699	0.011	0	0-NaN	1
		G/A	1(1.1)	3(7.7)	1(4.8)	2(33.3)	0	7.167	0.721-71.226	0.088	4.3	0.258-71.726	0.353	43	3.189-579.773	0.01	0	0-NaN	1
Exon 2	rs1739654	A/G	37(42.5)	16(41.0)	11(52.4)	1(16.7)	4(33.3)	0.701	0.401-1.226	0.246	0.977	0.471-2.027	1	0.072	0.007-0.786	0.031	0.613	0.256-1.469	0.349
Intron 2	rs1739655	A/G	37(43.7)	17(43.6)	11(52.4)	1(16.7)	5(41.7)	0.522	0.152-1.788	0.339	2.023	0.224-18.291	1	0.061	0.006-0.679	0.023	0.461	0.074-2.863	0.59
Intron 3	rs11536943	**A**	0	3(3.8)	2(4.8)	0	1(4.2)	∞	NaN-∞	0.029	∞	NaN-∞	0.037	-	-	-	∞	NaN-∞	0.121
		G/A	0	3(7.1)	2(9.5)	0	1(8.3)	∞	NaN-∞	0.029	∞	NaN-∞	0.036	-	-	-	∞	NaN-∞	0.121
	rs960451887	**A**	0	3(3.8)	2(4.8)	0	1(4.2)	∞	NaN-∞	0.029	∞	NaN-∞	0.037	-	-	-	∞	NaN-∞	0.121
		G/A	0	3(7.7)	2(9.5)	0	1(8.3)	∞	NaN-∞	0.029	∞	NaN-∞	0.036	-	-	-	∞	NaN-∞	0.121
	rs6025098	**T**	31(17.8)	6(7.7)	4(9.5)	0	2(8.3)	0.384	0.153-0.964	0.036	0.486	0.161-1.460	0.246	0	0-NaN	0.222	0.419	0.094-1.877	0.381
	rs137996210	**TATAA**	31(17.8)	6(7.7)	4(9.5)	0	2(8.3)	0.384	0.153-0.964	0.036	0.486	0.161-1.460	0.2246	0	0-NaN	0.222	0.419	0.094-1.877	0.381
	rs6014805	**T**	74(42.5)	21(26.9)	11(26.1)	2(16.7)	8(33.3)	0.498	0.278-0.892	0.024	0.48	0.226-1.016	0.055	0.27	0.057-1.270	0.127	0.676	0.275-1.662	0.508
		T/T	13(14.9)	1(2.6)	0	0	0	0.105	0.0123-0.875	0.022	0	0-NaN	0.045	0	0-NaN	1	0	0-NaN	0.375
	rs1250247980	**C***	0	5(6.4)	1(2.4)	3(25.0)	1(4.2)	∞	NaN-∞	0.003	∞	NaN-∞	0.194	∞	NaN-∞	0.0002	∞	NaN-∞	0.121
	rs11906594	A/G	31(17.8)	6(7.7)	4(9.5)	0	2(8.3)	0.384	0.153-0.964	0.036	0.486	0.161-1.460	0.2246	0	0-NaN	0.222	0.419	0.094-1.877	0.381
Intron 5	rs140069602	G/C	17(19.5)	12(13.8)	5(23.8)	4(66.7)	3(25.0)	1.178	0.750-4.214	0.253	1.25	0.401-3.893	0.765	8	1.351-47.376	0.024	1.333	0.325-5.465	0.708
Intron 6	rs113577957	**G**	0	2(2.6)	2(4.8)	0	0	∞	NaN-∞	0.095	∞	NaN-∞	0.037	-	-	-	-	-	-
		G/A	0	2(5.1)	2(9.5)	0	0	∞	NaN-∞	0.094	∞	NaN-∞	0.036	-	-	-	-	-	-
Intron 8	rs141862127	**C**	3(1.7)	4(5.1	1(2.4)	2(16.7)	1(4.2)	3.081	0.672-14.110	0.208	1.39	0.141-13.711	0.581	11.4	1.706-76.178	0.034	2.478	0.247-24.835	0.406
		A/C	3(3.4)	4(10.3)	1(2.8)	2(33.3)	1(8.3)	3.2	0.681-15.05	0.202	1.4	0.138-14.177	1	14	1.800-108.882	0.032	2.545	0.243-26.660	0.409
Intron 10	rs145588209	**T**	3(1.7)	4(5.1	1(2.4)	2(16.7)	1(4.2)	3.081	0.672-14.110	0.208	1.39	0.141-13.711	0.581	11.4	1.706-76.178	0.034	2.478	0.247-24.835	0.406
		C/T	3(3.4)	4(10.3)	1(2.8)	2(33.3)	1(8.3)	3.2	0.681-15.05	0.202	1.4	0.138-14.177	1	14	1.800-108.882	0.032	2.545	0.243-26.660	0.409
Exon 13	rs142724815	**C**	3(1.7)	4(5.1)	1(2.4)	2(16.7)	1(4.2)	3.081	0.673-14.110	0.208	1.39	0.141-13.711	0.581	11.4	1.706-76.178	0.034	2.478	0.247-24.835	0.406
		A/C	3(3.4)	4(10.3)	1(2.8)	2(33.3)	1(8.3)	3.2	0.681-15.05	0.202	1.4	0.138-14.177	1	14	1.800-108.882	0.032	2.545	0.243-26.660	0.409
Intron 13	rs140231834	**T**	3(1.7)	4(5.1	1(2.4)	2(16.7)	1(4.2)	3.081	0.672-14.110	0.208	1.39	0.141-13.711	0.581	11.4	1.706-76.178	0.034	2.478	0.247-24.835	0.406
		A/T	3(3.4)	4(10.3)	1(2.8)	2(33.3)	1(8.3)	3.2	0.681-15.05	0.202	1.4	0.138-14.177	1	14	1.800-108.882	0.032	2.545	0.243-26.660	0.409
3' flanking region	rs111621628	**C**	8(4.6)	11(14.1)	6(14.2)	2(16.7)	3(12.5)	3.407	1.312-8.843	0.017	3.458	1.130-10.580	0.034	4.15	0.777-22.174	0.129	2.964	0.729-12.050	0.134
		A/C	8(9.2)	11(28.2)	6(28.6)	2(33.3)	2(16.7)	3.879	1.416-10.626	0.013	3.95	1.197-13.035	0.028	4.938	0.779-31.296	0.124	3.292	0.738-14.684	0.128

SAs: South African (SA) controls (n=87), **HICs**: HIV-1 controllers (n=39), **ECs**: elite controllers (n=21), VCs: viraemic controllers (n=6), and HVL LTNPs: high viral load long term non progressors (n=12). Variant IDs highlighted in the same colour represent variants showing linkage disequilibrium. Significantly different p values between respective groups being compared are indicated by red font. Minor allele nucleotides are indicated with bold font. * major allele is T.

4.3.3.2 *LY96*

Comparisons of all *LY96* gene variants between HIV-1 controllers and SA controls revealed no significant differences in the representation of the regulatory region and exonic variants. We however found differences in the representation of several intronic *LY96* gene variants (**Table 4.6**). Minor allele frequency and heterozygosity of an intronic insertion, rs534972696, was higher in HIV-1 controllers compared to SA controls (MAF: p=0.01, OR=4.857, CI: 1.417-16.652; heterozygosity: p=0.008, OR=5.354, CI: 1.505-19.052). This higher representation of the rs534972696 indel in HIV-1 controllers was driven by the HVL LTNPs, with the indel showing stronger significant higher representation in HVL LTNPs compared to SA controls (MAF: p=0.008, OR=8.5, CI: 1.971-36.649; heterozygosity: p=0.007, OR=10.375, CI: 2.171-49.589). In elite controllers we found a significant lower MAF of three SNPs, namely rs7822709 (p=0.034, OR=0.209, CI0.052-0.842), rs7841862 (p=0.185, OR=0.052-0.658) and rs60515500 (p=0.185, OR=0.052-0.658) compared to the SA controls. The rs7841862 and rs60515500 SNPs are in strong LD (r^2=0.96) and they were also both in LD with the rs7822709 SNP (r^2=0.72). Minor allele frequency of the rs7841862 (p=0.01, OR=0.235, CI=0.083-0.670) and rs60515500 (p=0.01, OR=0.255, CI: 0.091-0.717) SNPs was also significantly lower in the total group of HIV-1 controllers compared to the SA controls. Lastly, we found an overrepresentation of the minor allele and heterozygosity of the rs149605245 SNP in elite controllers compared to SA controls (MAF: p=0.007; heterozygosity: p=0.007) - this SNP was absent in the SA control population, HVL LTNPs and viraemic controllers, however MAF and heterozygosity for this SNP was still higher in the total group of HIV-1 controllers compared to SA controls (MAF: p=0.029; heterozygosity: p=0.028), due to the high representation in the elite controller subgroup.

No significant differences in the representation of the *LY96* gene variants were observed between the HIV-1 controller subgroups.

Table 4.6: Comparison of allelic and genotypic frequencies of *LY96* variants between SA controls and HIV-1 controllers.

		Minor allele / genotype frequency					HICs vs SA controls			ECs vs SA controls			VC vs SA controls			HVL LTNPs vs SA controls		
Variant	**Minor allele/**genotype	**SAs**	**HICs**	**ECs**	**VCs**	**HVL LTNPs**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**
rs534972696	**GT**	4(2.3)	8(10.3)	3(7.1)	1(8.3)	4(16.7)	4.857	1.417-16.652	0.01	3.269	0.703-15.202	0.135	3.863	0.397-37.570	0.286	8.5	1.971-36.649	0.008
	G/GT	4(4.6)	8(20.5)	3(14.3)	1(16.6)	4(33.3)	5.354	1.505-19.052	0.008	3.458	0.711-16.813	0.132	4.15	0.388-44.391	0.289	10.375	2.171-49.589	0.007
rs7822709[2]	A/G	47(54.0)	14(35.9)	7(33.3)	3(50.0)	4(33.3)	0.298	0.089-0.995	0.054	0.209	0.052-0.842	0.034	∞	NaN-∞	∞	0.298	0.046-1.940	0.218
rs7841862[1]	A/G	51(58.6)	12(30.8)	6(28.5)	3(50.0)	3(25.0)	0.235	0.083-0.670	0.01	0.185	0.052-0.658	0.01	0.647	0.061-6.820	0.561	0.215	0.038-1.214	0.097
rs60515500[1]	A/C	51(58.6)	13(33.3)	6(28.5)	3(50.0)	4(33.3)	0.255	0.091-0.717	0.01	0.185	0.052-0.658	0.01	0.647	0.061-6.820	0.561	0.288	0.056-1.472	0.142
rs149605245	**C**	0	3(3.8)	3(7.1)	0	0	∞	NaN-∞	0.029	inf	nan-inf	0.007	-	-	-	-	-	-
	G/C	0	3(7.7)	3(14.3)	0	0	∞	NaN-∞	0.028	inf	nan-inf	0.007	-	-	-	-	-	-

SAs: South African (SA) controls (n=87), **HICs**: HIV-1 controllers (n=39), **ECs:** elite controllers (n=21), **VCs:** viraemic controllers (n=6), and **HVL LTNPs:** high viral load long term non progressors (n=12), Variant IDs highlighted in the same colour represent variants showing linkage disequilibrium. **[1]:** SNPs in linkage disequilibrium (LD, r^2=0.96), **[2]:** SNP in LD with SNPs marked with 1 (r^2=0.72). Significantly different p values between respective groups being compared are indicated by red font. Minor allele nucleotides are indicated with bold font.

4.4 Discussion

The lipopolysaccharide binding proteins LBP and MD-2 are essential components of the TLR4 signalling pathway (Lu et al., 2008). Polymorphisms in genes encoding these two proteins have been associated with varying expression levels, LPS responsiveness, as well as a number of diseases, in different populations (Hamann et al., 2004, Gu et al., 2007, Zeng et al., 2012a, Eckert et al., 2013, Bank et al., 2014, Lu et al., 2018). However, *LBP* and *LY96* gene association studies are scarce in African populations. In this study, we have described the *LBP* and *LY96* genetic variation (SNPs and indels) and LD patterns in black South Africans (control population) and compared to select data available from 1000 Genomes Project. We have also assessed representation of *LBP* and *LY96* polymorphisms in black South African HIV-1 controllers relative to the South African control individuals.

Here we describe five and 10 putative novel variants in the *LBP* and *LY96* genes, respectively, and all but one *LY96* variant (MAF of 2.3%, *LY96* NI-9, 8:74017606) were identified in single individuals, and would thus need to be verified as true novel variants using other genotyping methods. The whole genome sequencing results reveal that the *LBP* and *LY96* genes are highly polymorphic with 405 and 344 variants identified, respectively – largely attributed to the large gene sizes and with most variants located in the intronic regions. Using data available from the 1000 Genomes Project showed that, as expected, the MAFs of both the *LBP* and *LY96* variants in the SA population were more similar to those of African YRI and LWK reference populations and showed extensive significant differences to the non-African populations (EUR, EAS and SAS). Variants located in the 5' and 3' flanking regions of these genes in the SA population that were predicted to have a category 1 RegulomeDb rank, suggestive of a variant located in a functional region and associated with gene expression (Dong et al., 2022), were either present across all the populations or exclusive to the African populations - i.e., none were found exclusively in the SA population.

Extensive LD across the *LBP* and *LY96* genes was seen with 26 and 22 complete LD putative haplotypes identified in these genes, respectively. The largest *LBP* putative haplotype identified (*LBP* Hap-1; HF: 6.9%), involving 24 intronic variants, appears to be absent in both the African and non-African 1000 Genomes Project populations studied. Although this putative haplotype is comprised of only intronic SNPs, given that introns have been shown to play an important role in enhancing gene expression (Rose, 2019), this haplotype could potentially impact *LBP* gene regulation in the SA population. A previously identified common *LBP* haplotype in the 5' flanking region of the gene, comprised of four SNPs (rs1780616,

rs5741812, rs2232571 and rs1780617), was found to associate with susceptibility to sepsis (Flores et al., 2009). Although these SNPs were present in our population, LD between these SNPs was low.

A haplotype comprised of two SNPs (rs1809442 and rs1809441), located in the 5' UTR promotor region of the *LY96* gene, has been reported to be associated with increased MD-2 expression and *Dermatophagoides pteronyssinus* group 2 (Der p 2) allergy in a Taiwanese population (Liao et al., 2015a, Liao et al., 2015b). These two SNPs were present in our population and showed strong but not complete LD (r^2=0.84) - given our criteria of only describing haplotypes with absolute LD in this chapter, this haplotype was thus not described in this study. Nonetheless, representation of the rs1809442 and rs1809441 SNPs in the African-based populations (SA, YRI and LWK populations) was lower than the non-African populations (EUR, EAS, and SAS). Furthermore, these two SNPs have a RegulomeDB rank of 1f, which suggests that they are likely to affect transcription factor binding and gene expression levels – the lower frequency of these polymorphisms/haplotype in African populations may be responsible for a difference in MD-2 expression between African and non-African individuals and warrants further investigation.

HIV-1 controllers are a subgroup of people living with HIV-1 who can naturally control the HIV-1 infection in the absence of antiretroviral therapy. Comparison of *LBP* variants between HIV-1 controllers and SA controls revealed that the representation of three *LBP* variants (rs2232571, rs2232613 and rs2232618) previously associated with either LBP expression levels, LPS responsiveness or sepsis (Chien et al., 2008, Zeng et al., 2012a, Eckert et al., 2013, Lu et al., 2018), did not differ between SA controls and HIV-1 controllers. Several other *LBP* polymorphisms however were found to be differentially represented between SA HIV-1 controllers and controls, thereby suggesting a potential role of this molecule in HIV-1 control. Minor allele representation as well heterozygosity for a 5' flanking region SNP (rs6025023) was significantly higher in viraemic controllers compared to SA controls. In elite controllers, and the total group of HIV-1 controllers we found minor allele overrepresentation of a 3' flanking region SNP (rs111621628). RegulomeDB analysis revealed both the rs6025023 and rs111621628 SNPs to have a RegulomeDB rank of 5, suggesting that there is minimal evidence for these variants to be in a functionally relevant region. However, given that these are rare variants and no data were available for the rs6025023 SNP in the 1000 Genomes Project, there could be other, yet to be identified, functional consequences of these variants in the SA population. SNPs located in the 3' and 5' flanking regions have been shown to affect RNA

secondary structures (Gu et al., 2015) and these structures play an important role in microRNA-mediated gene regulation and initiation of translation (Gu et al., 2014, Mayr, 2019).

Our findings also show a heterozygous genotype overrepresentation of two coding region SNPs (rs1739654 and rs142724815) in viraemic controllers compared to SA controls. The rs142724815 SNP is also in LD with three intronic SNPs (rs141862127, rs145588209, and rs140231834) also showing similar associations, while the rs1739655 SNP in LD with rs1739654 was also differentially represented. The rs1739654 SNP is a synonymous SNP, while the rs142724815 SNP is a nonsynonymous variant causing a lysine to threonine substitution at position 439 which is predicted to be "tolerated" and "benign" by SIFT and Polyphen-2, respectively. Interestingly, according to the GTEx project (https://gtexportal.org/), the rs1739654 SNP results in higher expression levels of the LBP in lungs, emphasizing that synonymous variants can affect gene expression levels (Vihinen, 2022). In addition, the rs1739654 SNP has been associated with susceptibility to type 2 diabetes (Takeuchi et al., 2007) We also identified two intronic SNPs (rs1250247980 and rs140069602) with minor allele overrepresentation in viraemic controllers compared to SA controls, with the rs1250247980 SNP also showing minor allele overrepresentation in the total group of HIV-1 controllers. It is however important to note that the sample size for viraemic controllers (n=6) is small, and the range of the confidence interval for odds ratio was wide, suggesting that this association and all associations involving viraemic controllers should be treated with caution and verified in larger cohorts. However, extreme phenotypes allow for identification of rare causal variants - this has been demonstrated in a previous study conducted in a small cohort of LTNPs (n=7) and elite controllers (n=4) that identified several rare human genetic variants in genes involved in the HIV replication and innate immune sensing which associate with HIV-1 control in LTNPs (Nissen et al., 2018).

Our results also indicated minor allele overrepresentation of the following intronic variants, rs11536943, rs113577957, rs960451887, rs6014805, rs6025098, rs137996210, and rs11906594, in the total group of controllers. The rs11536943 and rs6014805 SNPs were also overrepresented in elite controllers. The rs6025098 and rs11906594 SNPs, and rs137996210 indel are variants which comprise *LBP* Hap-6. According to the RegulomeDB, the rs11536943 SNP has a rank of 1b, while the rs6014805, rs6025098, rs137996210, and rs11906594 variants have a rank of 1f, suggesting that these variants are likely to affect binding and expression levels. Data from the GTEx project showed that the rs11536943 and rs6025098 SNPs result in lower expression levels of LBP in skeletal muscle and BPI (bactericidal permeability

increasing protein) in various tissues including whole blood. Bactericidal permeability increasing protein is an LPS binding protein produced by neutrophils and is structurally and functionally related to LBP (Hubacek et al., 1997), however, BPI suppresses inflammation by interfering with the transfer of LPS to CD14 for TLR4 activation (Weiss, 2003). According to GTEx, the rs6014805 SNP results in lower expression of DHX35 (DEAH-box helicase 35) in the aorta artery, while the rs137996210 and rs11906594 variants result in lower LBP expression levels in various tissues.

LBP has functions beyond the transferring of LPS to CD14 for initiation of innate immune responses - this molecule can at higher concentrations inhibit the inflammatory immune responses by transferring LPS to lipoproteins (Zweigner et al., 2001). Given this concentration depended dual role of LBP, variants which could affect expression levels and/or structure of LBP have the potential to influence the innate immune response. It is therefore not surprising that variants in the LBP genes have been associated with range of inflammation related diseases (Meng et al., 2021). With respect to HIV-1, plasma levels of LBP are considered an indirect marker of microbial translocation, therefore *LBP* variants could play a role on immune activation in people living with HIV-1. In people living with HIV-1, the *LBP* SNP (rs2232582; synonymous SNP) has been associated with HAART-associated lipodystrophy syndrome (Viladés et al., 2014). Interestingly, the rs2232582 SNP is a very prevalent SNP (MAF: 47.7%) in the SA population and significantly more prevalent compared to the LWK African population as well as all the non-African populations.

MD-2 is a key molecule in LPS sensing and activation of innate immune responses (Shimazu et al., 1999). Expression levels of this molecule have been shown to be elevated in people living with HIV-1 and in sepsis (Pugin et al., 2004, Sandanger et al., 2009). Previous studies have shown that *LY96* gene variants affect LPB expression levels and LPS responsiveness (Hamann et al., 2004, Gu et al., 2007). In this study, we did not detect the rare *LY96* nonsynonymous mutation (+103 A to G) which causes a threonine to alanine substitution at position 35 and results in reduced NF-κβ-activation and TNF-α production upon LPS stimulation (Hamann et al., 2004). We did however detect one synonymous SNP (rs1195824354) in the coding region of the *LY96* gene - the rs1195824354 SNP is a rare variant, only present in one individual in the SA population, and data for this SNP were not available in the 1000 Genomes Project for all populations. Although synonymous SNPs do not effect on protein function, it has been shown that they can alter local RNA structures which could potentially affect gene expression (Gu et al., 2015). Another important *LY96* variant is the rs11465996 SNP located in the

promoter region of the *LY96* gene, and has been shown to affect the promotor activity and associate with increased expression of *LY96* mRNA and TNF-α upon LPS stimulation (Gu et al., 2007). This variant has also been associated with a range of diseases (Zeng et al., 2012b, Bank et al., 2014). In the SA population, the rs11465996 SNP was a fairly prevalent variant (MAF: 13.8%), however MAF of this SNP was significantly lower compared to the non-African European and South Asian 1000 Genomes Project populations, possibly suggesting differences in the functional significance of this promotor variant in the different populations.

To the best of our knowledge, no previous studies have investigated the representation of *LY96* polymorphisms in HIV-1 controllers relative to progressors or to a control background population. In this study, although we failed to see any significant differences in the representation of *LY96* regulatory and coding region variants between SA controls and HIV-1 controllers, we did find several intronic variants showing differences representation. Minor allele and heterozygosity overrepresentation of the rs534972696 SNP in the total group of controllers and in HVL LTNPs compared to the SA controls was seen. Although this variant has a RegulomeDB rank of 6, suggesting minimal evidence of functional significance, Ensembl (https://www.ensembl.org/index.html) predicts that this variant impacts gene regulation. The minor alleles of two intronic SNPs (rs7841862 and rs60515500) in LD (r^2=0.96) were underrepresented in the total group of HIV-1 controllers and elite controllers compared to SA controls. These SNPs have a RegulomeDB rank of 1f suggesting that there are located in a region of functional significance. In addition, MAF of the SNP (rs7822709), in LD (r^2=0.72) with the rs7841862 and rs60515500 SNPs, was significantly lower in elite controllers compared to SA controls. Lastly, the rs149605245 SNP, found exclusively in elite controllers (MAF: 7.1%), showed both MAF and heterozygosity to be significantly overrepresented in elite controllers compared to SA controls (also resulted in the total HIV-1 controller group also showing significant overrepresentation compared to the SA controls). This variant has a RegulomeDB rank of 7, suggesting minimal evidence of functional significance, and no significant eQTLs were found for the variant in the GTEx project. However, given that the majority of GTEx project tissue donors are white individuals (84.6%), the lack of significant eQTLs for this variant could be due to ethnic differences, as this variant is absent in all three non-African 1000 Genomes Project populations (EUR, EAS and SAS) - the rs149605245 SNP thus seems to be exclusive to African populations and has a MAF of 4% in the total African population (1000 Genomes Project). It will be interesting to investigate whether this variant is indeed involved in HIV-1 elite control in African populations.

It is worth noting that we did not observe any significant association between *LBP* variants and HVL LTNPs in this study. In addition, no significant associations between *LY96* variants and viraemic controllers were observed. Although the lack of association could be due to low sample sizes in both populations, these results also suggest differences in mechanisms of control between HIV-1 controllers.

This study has several limitations. Firstly, the sample sizes for our study participants were low, as already mentioned, particularly the group of viraemic controllers (n=6), therefore, all significant association in viraemic controllers need to be verified in larger cohorts. Additionally, care should be taken in the interpretation of rare SNPs in our cohorts as these were small sample sizes and might not necessarily be representative of the bigger populations. We also did not correct for multiple comparisons due to the exploratory nature of this study. Findings from this study should form the basis for future studies aimed at investigating functional or regulatory *LBP* and *LY96* variants, which potentially affect immune responses to microbial translocation and/ or play a role in HIV-1 control in cohorts with the inclusion of HIV-1 progressors.

In conclusion, our results suggest that *LBP* and *LY96* polymorphisms and LD patterns in the black SA population more closely mirror those of other African populations compared to European and Asian populations - this emphasises the importance of population specific association studies and identification of population specific eQTLs. Overall, our findings contribute to the understanding of the extent of genetic variation in these important genes in the South African population. In addition, we have identified a number of *LBP* and *LY96* gene variants that are differentially represented in black South African HIV-1 controllers – using these results as a reference study, these associations can now be tested using larger replication cohorts.

CHAPTER 5

Plasma levels of REG3α and sCD14 distinguish extreme clinical phenotypes of black South African persons living with HIV-1

5.1 Introduction

Despite the success of antiretroviral therapy (ART), morbidity and mortality from both AIDS and non-AIDS defining illnesses are still higher in people living with HIV-1 (PLWH) than the general population (Sandler and Douek, 2012, de Coninck et al., 2018). These outcomes are related to the associated persistent immune activation and inflammation, even in PLWH on suppressive ART (Sandler et al., 2011, Hunt et al., 2014). Among other factors such as residual HIV replication and co-infections with other pathogens (Younas et al., 2016), microbial translocation has also been identified as a major factor fuelling the persistent inflammatory immune response (Brenchley et al., 2006).

The preferential depletion of CD4+ T cells from the gastrointestinal tract during the acute phase of HIV-1 infection (Brenchley et al., 2008) results in the disruption of the gut integrity, subsequently resulting in the translocation of microbial products such as lipopolysaccharide (LPS), from the gut into the circulatory system (Sandler and Douek, 2012). Lipopolysaccharide, a component of gram-negative bacteria, elicits innate immune responses through the activation of the toll-like receptor 4 (TLR4), in a complex pathway that involves the use of additional receptors and accessory molecules. These include CD14, lipopolysaccharide binding protein (LBP), and Myeloid differentiation factor 2 (MD-2) (Lu et al., 2008).

Plasma levels of LPS and bacterial 16S ribosomal RNA (16S rRNA) gene copy numbers have been used as direct markers of bacterial translocation; whereas, LBP, sCD14 (the soluble version of the CD14 surface receptor released upon monocyte activation), and endotoxin core antibodies (EndoCAb) are used as indirect markers of microbial translocation (Marchetti et al., 2013). Elevated plasma levels of LPS, LBP, sCD14, and bacterial 16S rRNA gene copy numbers in PLWH compared to HIV-1 uninfected controls have been reported (Lien et al., 1998, Brenchley et al., 2006, Jiang et al., 2009, Nowroozalizadeh et al., 2010, Younas et al., 2019). In addition, studies conducted predominantly in Caucasians found elevated plasma levels of LPS, LBP, sCD14, and reduced EndoCAb to be associated with heightened T cell activation, low CD4+ T cell count, poor CD4+ T cell reconstitution after ART, high viral load, and disease progression (Lien et al., 1998, Brenchley et al., 2006, Marchetti et al., 2008, Nowroozalizadeh et al., 2010, Marchetti et al., 2011, Reus et al., 2013, Pilakka-Kanthikeel et al., 2014). Microbial translocation studies in African population groups have often yielded conflicting results. In South Africa, Kenya, and Guinea-Bissau, associations between HIV-1

infection and elevated plasma levels of LPS have been reported (Lester et al., 2009, Cassol et al., 2010, Nowroozalizadeh et al., 2010). However, in Ugandan PLWH, circulating levels LPS, EndoCAb, and sCD14 did not associate with HIV-1 disease progression (Redd et al., 2009). Antiretroviral therapy has been shown to reduce levels of microbial translocation, with the levels being lower in PLWH 48 weeks post initiation of ART (Brenchley et al., 2006) or in ART-treated compared to ART-naïve PLWH (Cassol et al., 2010).

A subset of PLWH termed elite controllers are able to naturally suppress the HIV-1 viral load (VL) to undetectable levels (<50 RNA copies/ml) in the absence of ART. Whereas, another subset known as long term non-progressors (LTNPs) are able to tolerate high viral loads (HVL, generally >10,000 RNA copies/ml) while maintaining CD4+ T cell counts >500 cells/µl for extended periods of time (>7 years) in the absence of ART (Gurdasani et al., 2014). Such HVL LTNPs are a rare phenotype, which display similar characteristics to simian immunodeficiency virus (SIV)-infected nonhuman primates that do not progress to AIDS (Silvestri et al., 2003, Choudhary et al., 2007, Rotger et al., 2011, Koor et al., 2019).

Elite controllers are an important study model for understanding and identifying factors that underlie natural control that can inform strategies for attaining a functional cure. However, despite having undetectable viral loads, some elite controllers experience a progressive decline of CD4+ T cells (Hunt et al., 2008, Leon et al., 2015). Additionally, they are at higher risk of serious non-AIDS-defining illnesses than HIV-1 uninfected individuals (Pereyra et al., 2012). Markers of microbial translocation have been shown to independently predict clinical outcomes in HIV-1 controllers (Leon et al., 2015). Th17 cells play a crucial role in the maintenance of the integrity of the intestinal epithelial barrier (Liu et al., 2009). Despite the findings that elite controllers have a higher frequency of Th17 cells compared to healthy HIV-1 uninfected individuals and PLWH on ART (Kim et al., 2014, Caetano et al., 2020), evidence of microbial translocation in these individuals has been reported. Levels of sCD14 and LBP have been shown to be lower in elite controllers when compared to HIV-1 progressors, however, compared to HIV-1 uninfected individuals, only levels of sCD14 and levels of LPS have been shown to be higher (Brenchley et al., 2006). Elevated plasma levels of LPS have also been associated with high levels of CD8+ T cell activation and CD4+ T cell depletion in HIV-1 controllers (Hunt et al., 2008). However, no differences in levels of LPS and sCD14 were seen between elite controllers and healthy HIV-1 uninfected individuals (Kim et al., 2014), including a study in Uganda, which did not find differences in LPS, LBP, and EndoCAb levels between PLWH with different HIV-1 clinical phenotypes (Redd et al., 2009).

Soluble intestinal fatty acid binding protein (I-FABP), a commonly used marker of intestinal damage, is reportedly higher in HIV-1 progressors, including virologically suppressed ART-treated PLWH, compared to healthy HIV-1 uninfected individuals and elite controllers (Cheru et al., 2018, Isnard et al., 2020). In a separate study however, no difference in I-FABP levels between PLWH and HIV-1 uninfected controls was reported (Younas et al., 2019). Correlations between I-FABP levels and markers of microbial translocation and immune activation have been inconsistent (Romero-Sanchez et al., 2012, Hunt et al., 2014, Cheru et al., 2018, Isnard et al., 2020). Recently, a study found elevated plasma levels of regenerating islet-derived protein 3α (REG3α), a C-type lectin constitutively expressed in the Paneth cells in the small intestine, in PLWH including elite controllers (Isnard et al., 2020). Although I-FABP levels correlated positively with REG3α in this study, REG3α but not I-FABP, correlated with markers of HIV-1 disease progression, microbial translocation, immune activation, and inflammation (Isnard et al., 2020).

Compared to studies of microbial translocation in PLWH of European descent, there are only a handful of studies that have been conducted in African populations. Furthermore, given some of the inconsistent findings across studies, there is a need to gain better insight into the molecules that are touted as plasma markers associated with microbial translocation, their relevance as markers in different populations, and factors that may influence their expression. In this study, we have quantitated plasma levels of select innate immune molecules that are reported to be altered as a consequence of bacterial translocation (LBP, sCD14, REG3α), or important in interacting with TLR4 (MD-2), in ART-naïve black South African elite controllers, HVL LTNPs and progressors, as well as in healthy HIV-1 uninfected controls.

5.2 Methods

5.2.1 Study participants

Our study was conducted on four groups of individuals from the black South African population –HIV-uninfected controls (HUCs), elite controllers, HVL LTNPs, and progressors. This design allows us to compare PLWH who, in the absence of ART, spontaneously control their VLs to <50 RNA copies/ml, and a very rare group of individuals who maintain CD4+ T cell counts >500 cells/µl despite high VLs, and those who have progressive infection at the time of recruitment. These represent extremes in terms of clinical parameters of VL and CD4 counts. The study of extreme phenotypes is highly informative despite smaller sample sizes.

The absence of any known co-infections was a recruitment criterion of both PLWH and uninfected individuals. HIV-1 uninfected controls (n=17) were volunteers recruited from the National Institute for Communicable Diseases, Johannesburg between September 2019 and January 2020.

Elite controllers (n=44) were defined as individuals, who in the absence of ART, have the ability to suppress viral loads to <50 RNA copies/ml of plasma and have CD4+ T cell counts >500 cells/µl (with the exception of a small number (n=4) where CD4+ T cell counts had fallen below 500 cells/µl). The definition was not based on duration of viral suppression. The majority of elite controllers had at least two VL tests performed 6 months apart with VL <50 RNA copies/ml (n=39). Five elite controllers had only one VL test with VL <50 RNA copies/ml. Of the 44 elite controllers, 26 were recruited between 2011 and 2018 from different sites in Johannesburg (JHB), Gauteng province (Chris Hani Baragwanath Hospital, Helen Joseph Hospital, and Life Brenthurst Hospital), and the remaining 18 were recruited from two eThekwini primary health clinics (Lancers Road and Chesterville) in Durban, KwaZulu-Natal (KZN) province between 2015 and 2017. The HVL LTNPs (n=12), recruited from Johannesburg sites, were defined as PLWH with the ability to maintain CD4+ T cell counts >500 cells/µl in the presence of VL >10,000 RNA copies/ml of plasma for >7 years without the use of ART.

Progressors (n=24) were defined as PLWH with CD4+ T cell counts <200 cells/µl and requiring initiation of ART. Samples were collected prior to commencement of ART. Like elite controllers, progressors were also recruited from both the Johannesburg (n=12) and the Durban, KZN sites (n=12), in 2015 and 2017, respectively. Progressors were selected to match the viral loads of the HVL LTNPs (2 progressors, one each from JHB and KZN, for every one HVL LTNP). Since progressors were enrolled into the study on initial diagnosis there were no data available regarding the duration of their infection. Concerted effort was however made to enrol younger individuals to control for individuals that may have been infected for extended periods without ART (i.e., may have exhibited viral control or suppression).

Viral loads (HIV-1 RNA levels) in plasma were determined using the COBAS® AmpliPrep/COBAS® Taqman® HIV-1 test, v1.5 and v2.0 ultrasensitive tests (Roche Diagnostic Systems, Inc., New Jersey, USA). The FACSCount System from Becton Dickinson (San Jose, California, USA) was used to determine CD4+ T cell counts in whole blood.

5.2.2 Quantification of LBP, sCD14, MD-2, and REG3α plasma levels

EDTA-anticoagulated whole blood samples were collected and processed within 6 hours of collection through centrifugation to collect plasma, which was stored at – 80 °C until use. The fasting status of the participants at the time of blood collection was not determined. Enzyme-linked immunosorbent assay (ELISA) assays were used to quantify plasma levels of LBP (Hycult Biotech, Uden, the Netherlands), sCD14 (R&D Systems, Minneapolis, Minnesota, USA), MD-2 (Sigma-Aldrich, Saint Louis, Missouri, USA) and REG3α (Novus Biologicals, Centennial, Colorado, USA), as per the respective manufacturer's instructions. The limits of detection for the LBP, sCD14, MD-2, and REG3α ELISA kits were 4.4 ng/ml, 250 pg/ml, 1.6 ng/ml and 15.63 pg/ml, respectively. Samples were assayed in duplicate, and the Versamax microplate reader (Molecular Devices, San Jose, California, USA) was used to measure optical density (OD) at an absorbance of 450 nm and 540 nm, and results expressed as ng/ml.

5.2.3 Statistical analysis

GraphPad Prism version 8.01 for Windows (GraphPad Software, La Jolla, California, USA) and Statistica™ software (TIBCO Software, Palo Alto, California, USA) were used for statistical analyses. The Mann-Whitney U test was used for comparisons of plasma levels of LBP, sCD14, MD-2 and REG3α between two groups and the Kruskal-Wallis test was used for comparisons between three or more groups, with p values <0.05 considered statistically significant. The false discovery rate (FDR) approach was used to correct for multiple comparisons. FDR was calculated using the two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli and adjusted p values <0.05 were considered statistically significant. Both unadjusted and FDR-adjusted p values are reported when applicable. The Spearman correlation coefficient was used to assess relationships between age, CD4+ T cell count and viral load with plasma levels of LBP, sCD14 and REG3α and the relationships between the respective markers.

5.3 Results

5.3.1 Characteristics of the study participants

The characteristics of the HIV-1 uninfected controls, elite controllers, HVL LTNPs and progressors are shown in **Table 5.1**. The HIV-1 uninfected controls (52.9% females) had a median age of 32 [interquartile range (IQR): 27.5-44.5] years. Elite controllers (ECs), HVL LTNPs and progressors were predominantly female, with at least 70% in each group ($p>0.05$).

The median age of the elite controllers and HVL LTNPs was 39 (IQR: 31-44) years and 39.5 (IQR: 33.3-46.5) years, respectively. Progressors had a median age of 29.5 (IQR: 27-36.3) years. The HIV-1 uninfected controls were significantly younger than the JHB ECs (p=0.030) but similar in age to total elite controllers, HVL LTNPs and progressors. Progressors were significantly younger compared to elite controllers (p=0.029). Furthermore, elite controllers and progressors recruited from KZN were significantly younger than elite controllers and progressors recruited from JHB (p=0.001, p=0.020, respectively).

KZN ECs had significantly higher CD4+ T cell counts compared to the JHB ECs (p=0.002). All elite controllers had undetectable viral loads (<50 RNA copies/ml) at the time of conducting the assays, with the exception of one individual recruited from JHB (VL=165 RNA copies/ml). Progressors were selected to match the HVL LTNPs in terms of their VLs, hence levels were not statistically different.

Table 5.1: Characteristics of study participants.

Study group	N	Sex (% female)	Age (years) Median (IQR)[3]	Years since diagnosis Median (IQR)	CD4+ T cell count (cells/µl) Median (IQR)	Viral load (RNA copies/ml)
HIV-1 uninfected controls[5]	17	52.9	32 (27.5-44.5)	N/A	ND	N/A
Elite controllers	44	72.7	39 (31-44)	3 (2 – 6.5)	829 (625-1 048)	<50
JHB	26	73.1	42 (37-47)	6 (2.75-8.5)[6]	647 (551-848)	<50[1]
KZN	18	72.2	31 (26-39)	2 (2-3)[7]	980 (842-1 150)[4]	<50
HVL LTNPs	12	75	39.5 (33-46.5)	7 (7-8)	661 (611-780)	38 393 (12 393 – 79 573)
Progressors	24	70.8	29.5 (27-36)	ND	156 (92-173)	37 302 (13 375 – 80 928)
JHB	12	66.7	35 (27-41)	ND	163 (94-219)[2]	37 302 (12 952 – 85 661)
KZN	12	75	27 (26-31)	ND	151 (89-181)	38 900 (13 375 – 78 550)

[1]One JHB EC had a viral load of 165 RNA copies/ml at the time of assay. [2]No CD4+ T cell count for one JHB progressor. [3]Progressors were significantly younger than elite controllers (p=0.029), KZN ECs and progressors were significantly younger than JHB ECs and progressors (p=0.001 and p=0.020, respectively). [4]KZN ECs had significantly higher CD4+ T cell count than JHB ECs (p=0.002). [5]HIV-1 uninfected controls significantly younger than JHB ECs (p=0.030). [6]Data not available for 8 of the 26 JHB elite controllers. ND: not done.

5.3.2 Plasma levels of LBP, sCD14, MD-2, and REG3α

Plasma levels of LBP, sCD14, MD-2, and REG3α were quantitated in HIV-1 uninfected controls, elite controllers, HVL LTNPs, and progressors. The levels of LBP did not differ significantly between any groups, although progressors had a larger spread of data points (**Figure 5.1A**). Soluble CD14 was substantially higher in progressors compared to HIV-1 uninfected controls (p=0.0001, FDR-adjusted p=0.0002), elite controllers (p<0.0001, FDR-adjusted p<0.0001), and HVL LTNPs (p=0.0005, FDR-adjusted p=0.0005), with no

differences between HIV-1 uninfected controls, elite controllers and HVL LTNPs (**Figure 5.1B**). Only 28 of our total study participants (n=97) had detectable levels of MD-2. The proportion of individuals with undetectable MD-2 plasma levels was not significantly different between HIV-1 uninfected controls, elite controllers, HVL LTNPs, and progressors (Fisher's exact test, **Figure 5.2A**). Levels of MD-2 were not significantly different between the study groups when all individuals including those with undetectable levels were compared (**Figure 5.1C)**. However, there was a strong trend of lower levels of MD-2 in elite controllers compared to HIV-1 uninfected controls (p=0.062; **Figure 5.1C**), which was lost after controlling for FDR. When individuals with undetectable levels of MD-2 were excluded, elite controllers had significantly lower levels of MD-2 compared to HIV-1 uninfected controls (p=0.011, **Figure 5.2B),** however this difference was reduced to a trend after controlling for FDR (FDR-adjusted p=0.070). Plasma levels of REG3α were unexpectedly significantly lower in progressors than in elite controllers (p=0.007, FDR-adjusted p=0.036) and in HVL LTNPs (p=0.018, FDR-adjusted p=0.046, **Figure 5.1D**), the opposite of what has previously been reported (Isnard et al., 2020). Levels of REG3α in elite controllers and HVL LTNPs were not significantly different from HIV-1 uninfected controls (**Figure 5.1D**. Therefore, only sCD14 levels reflected the expected increased immune activation that occurs with more progressive infection and increased microbial translocation.

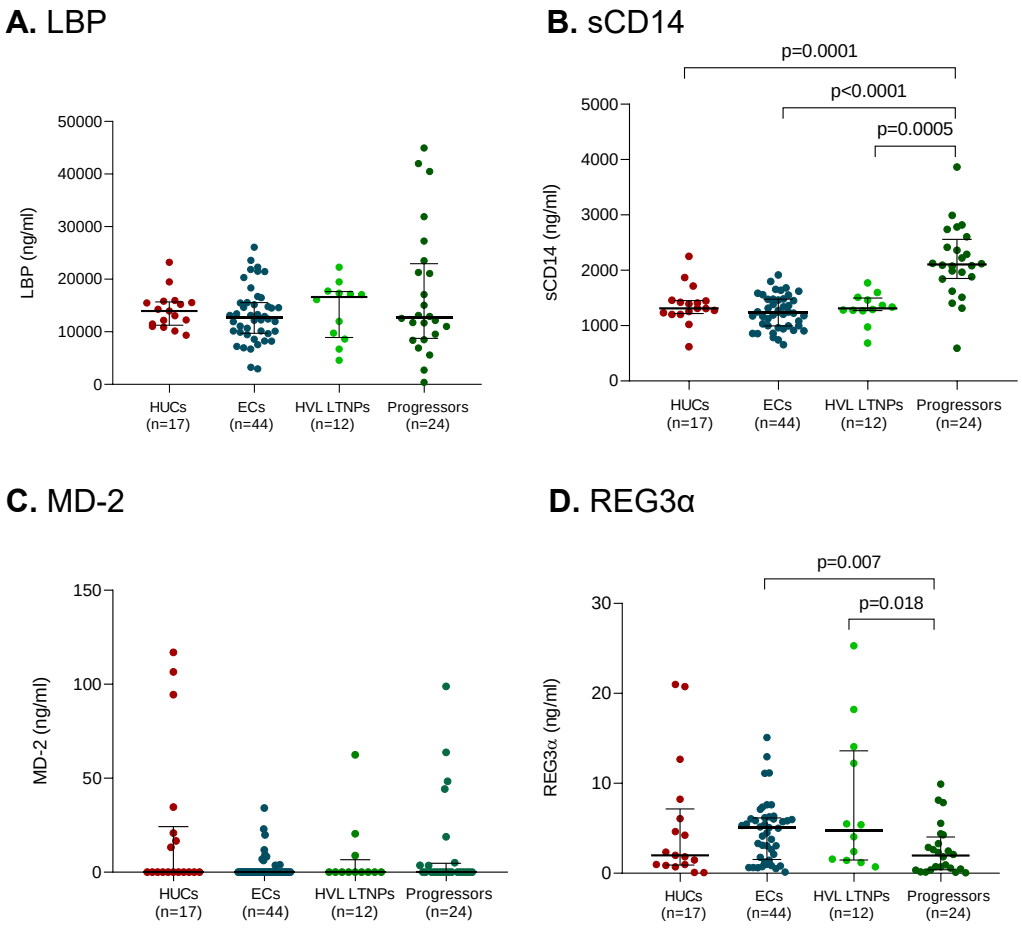


Figure 5.1: Comparisons of plasma levels of **(A)** LBP, **(B)** sCD14, **(C)** MD-2, and **(D)** REG3α between HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long-term non-progressors (HVL LTNPs), and progressors. Each dot represents a single individual, and the horizontal lines represent the median and interquartile ranges. Kruskal-Wallis test was used to test for statistical significance. FDR-adjusted p values for sCD14 comparison from left to right: p=0.0002, p<0.0001 and p=0.005. FDR-adjusted p values for REG3α comparison from left to right: p=0.036, and p=0.046.

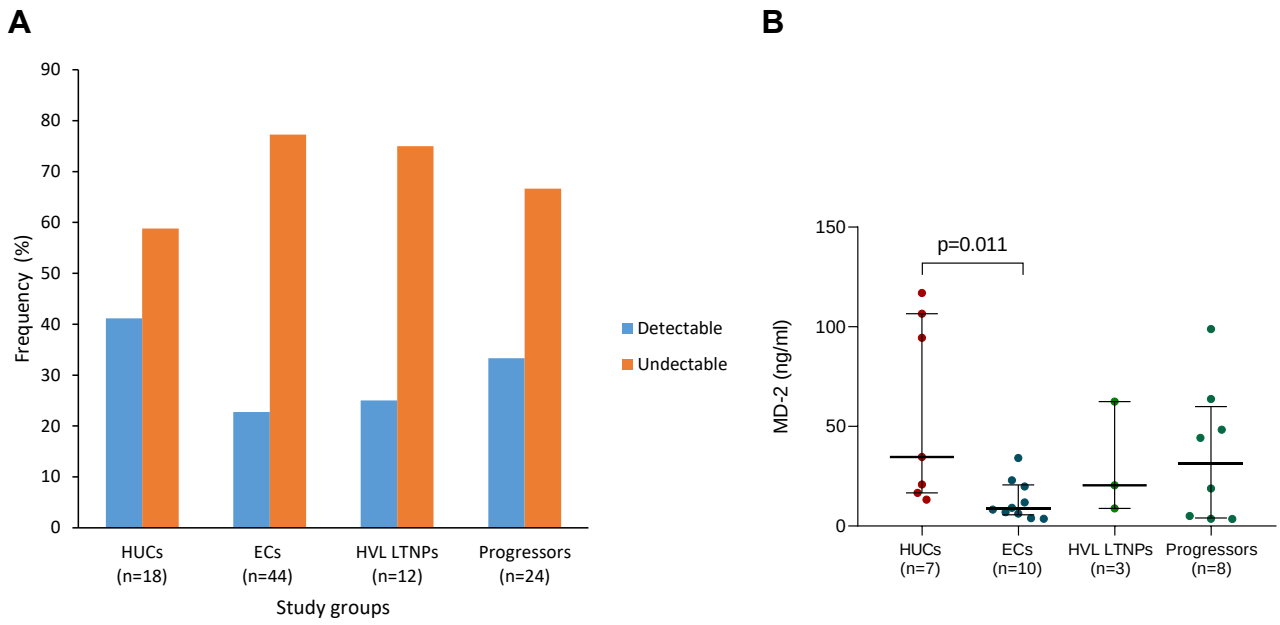


Figure 5.2: MD-2 plasma levels. **(A)** Frequency of individuals with detectable and undetectable plasma levels of MD-2 in HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long term non-progressors (HVL LTNPs), and progressors. **(B)** Comparisons of detectable plasma levels of MD-2 between HUCs, ECs, HVL LTNPs, and progressors. Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges.

5.3.3 Sex differences and plasma levels of LBP, sCD14, MD-2, and REG3α

Sex differences have been shown to influence constitutive expression levels of various genes in different tissues (Oliva et al., 2020) and impact innate and adaptive immunity (Klein and Flanagan, 2016). In PLWH, women are at higher risk of progression to AIDS for the same viral load as men (Farzadegan et al., 1998, Scully, 2018), however elite controllers are more likely to be female than male (Yang et al., 2017). We therefore compared plasma levels of LBP, sCD14, MD-2, and REG3α between males and females in our study groups. Females had a trend of higher LBP plasma levels compared to males in elite controllers (p=0.059; **Figure 5.3A**). No significant sex-specific differences were seen in levels of sCD14 (**Figure 5.3B**) or of MD-2 (**Figure 5.3C**). However, there was a trend of elevated MD-2 in female compared to male progressors (p=0.053**, Figure 5.3C**) when all individuals including those with undetectable levels were compared. This analysis however, could not be performed when individuals with undetectable levels were excluded, due to all male progressors having undetectable levels of MD-2. Interestingly, REG3α showed marked differences with respect to sex, with significantly higher levels in females compared to males in all the study groups (**Figure 5.4A,** HIV-1 uninfected controls and elite controllers, p<0.0001; HVL LTNPs, p=0.036; progressors, p=0.005).

Due to the sex differences observed in LBP (trend in elite controllers) and REG3α plasma levels (in all study groups), we also independently compared females and males across the groups. Plasma levels of LBP in males or females were not significantly different between any groups (**Figure 5.3A**). Regardless of the marked sex differences seen in plasma levels of REG3α, both male and female elite controllers (males, p=0.029; females, p=0.006) and HVL LTNPs (males, p=0.016; females, p=0.020) had significantly higher plasma levels of REG3α compared to male and female progressors, respectively (**Figure 5.4B**). When controlled for FDR, these significant differences were lost in males but not in females (elite controllers, FDR-adjusted p=0.020; HVL LTNPs, FDR-adjusted p=0.021). This was likely due to the smaller numbers of males in our study. The stratification by sex also revealed significantly higher REG3α plasma levels in HIV-1 uninfected controls (**Figure 5.4B**) compared to progressors in females (p=0.015; FDR-adjusted p=0.021), and a trend in males (p=0.051). Therefore, there was no evidence of increased systemic levels in more progressive infection.

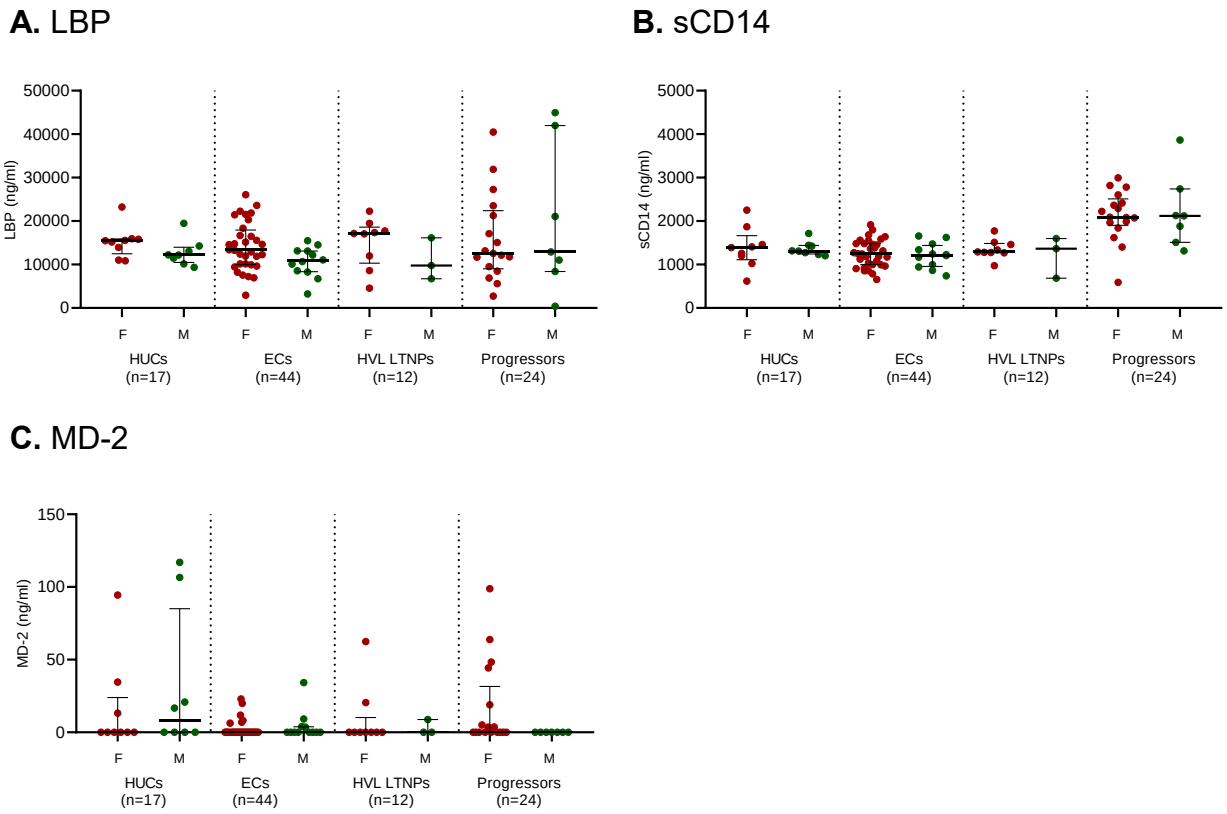


Figure 5.3: Comparisons of plasma levels of (**A**) LBP, (**B**) sCD14, and (**C**) MD-2 between males (M) and females (F) in HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long term non progressors (HVL LTNPs), and progressors. Mann-Whitney U-tests were conducted to test for significance. Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges.

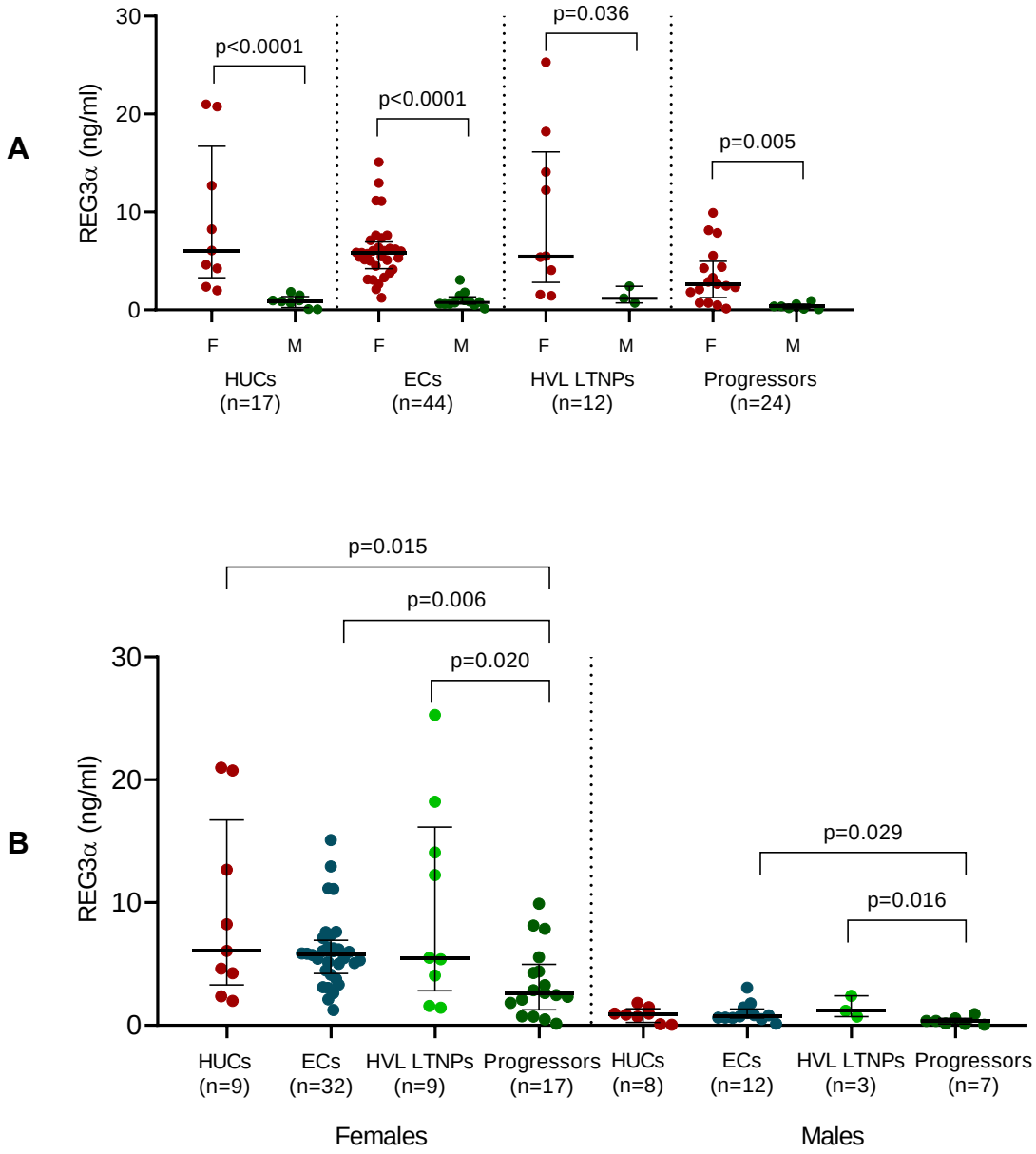


Figure 5.4: Comparisons of plasma levels of REG3α **(A)** between males (M) and females (F) in HIV-1 uninfected controls (HUCs), elite controllers (ECs), high viral load long-term non progressors (HVL LTNPs), and progressors. Mann-Whitney U-tests were conducted to test for significance. P values <0.05 are considered statistically significant. **(B)** Comparisons of REG3α plasma levels between female HUCs, ECs, HVL LTNPs and progressors and comparisons between male HUCs, ECs, HVL LTNPs and progressors. Kruskal-Wallis tests were conducted to test for significance. FDR-adjusted p values from left to right in females: p=0.021, p=0.020 and p=0.021, in males all FDR-adjusted p>0.05. Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges.

5.3.4 Plasma level differences (LBP and sCD14) among elite controllers according to site (JHB, KZN)

Elite controllers are a heterogeneous group of PLWH exhibiting differences in virological, immunological, and clinical characteristics (Navarrete-Muñoz et al., 2020). Since our elite controllers and progressors were recruited from two different provinces and considering that the KZN EC group was significantly younger and had higher CD4+ T cell counts compared to the JHB ECs, we therefore, in addition, stratified the elite controllers and progressors according to site to compare the plasma levels of the study biomarkers. Compared to the combined groups comparisons (**Figure 5.1**), all significant relationships identified were maintained in both JHB and KZN ECs and progressors when analysed separately (data not shown), with one exception and some additional differences that were unmasked. No significant differences in plasma levels of LBP, sCD14, MD-2, and REG3α were observed between JHB and KZN progressors. However, levels of both LBP (**Figure 5.5A**) and sCD14 (**Figure 5.5B**) were significantly higher in KZN ECs compared to JHB ECs (p=0.009 and p=0.02, respectively). On the other hand, levels of MD-2 (even when individuals with undetectable levels were excluded), and REG3α were similar (**Figure 5.5C and 5.5D**), although the KZN ECs had a broader range of REG3α. Additionally, the JHB ECs had significantly reduced LBP (p=0.035) and sCD14 (p=0.023) levels compared to HIV-1 uninfected controls, and reduced sCD14 levels compared to HVL LTNPs (p=0.041, **Figure 5.5E and 5.5F**). This was not unexpected given the differences in LBP and sCD14 levels observed between JHB and KZN ECs (**Figure 5.5A and 5.5B**). The differences in sCD14 plasma levels in JHB ECs compared to HIV-1 uninfected controls and HVL LTNPs were lost after correction for FDR.

5.3.5 Correlations between LBP, sCD14, and REG3α plasma levels

Correlations between plasma levels of LBP, sCD14, REG3α, and with age and CD4+ T cell count in the combined study groups, and when stratified by sex and recruitment site, are shown in **Figure 5.6**. In the total group of HIV-1 uninfected controls, there were no significant correlations between the levels of any of the plasma markers. However, when stratified by sex, a significant negative correlation between LBP and sCD14 was seen in males (r=-0.738, p=0.037). This relationship was however opposite in the presence of HIV-1 infection. LBP positively correlated with sCD14 in elite controllers (r=0.593, p<0.0001), a relationship predominantly due to elite controllers recruited from JHB (r=0.616, p=0.005), and in total progressors (r=0.483, p=0.017). LBP also showed a very strong positive correlation with REG3α in HVL LTNPs (r=0.888, p=0.0003) and a more moderate correlation in elite

controllers (r=0.440, p=0.003), which was also maintained when the elite controllers were stratified by recruitment site (JHB ECs, r=0.398, p=0.044; KZN ECs, r=0.494, p=0.037). These relationships were strongly driven by female sex (HVL LTNPs, r=0.867, p=0.005; elite controllers, r=0.399, p=0.024). The KZN female elite controllers (r=0.571, p=0.045) were the main contributors to the latter correlation, with significance lost in the JHB female elite controllers.

No significant correlations were observed between REG3α and sCD14, in the subgroups of PLWH regardless of the stratification by the site of recruitment or sex.

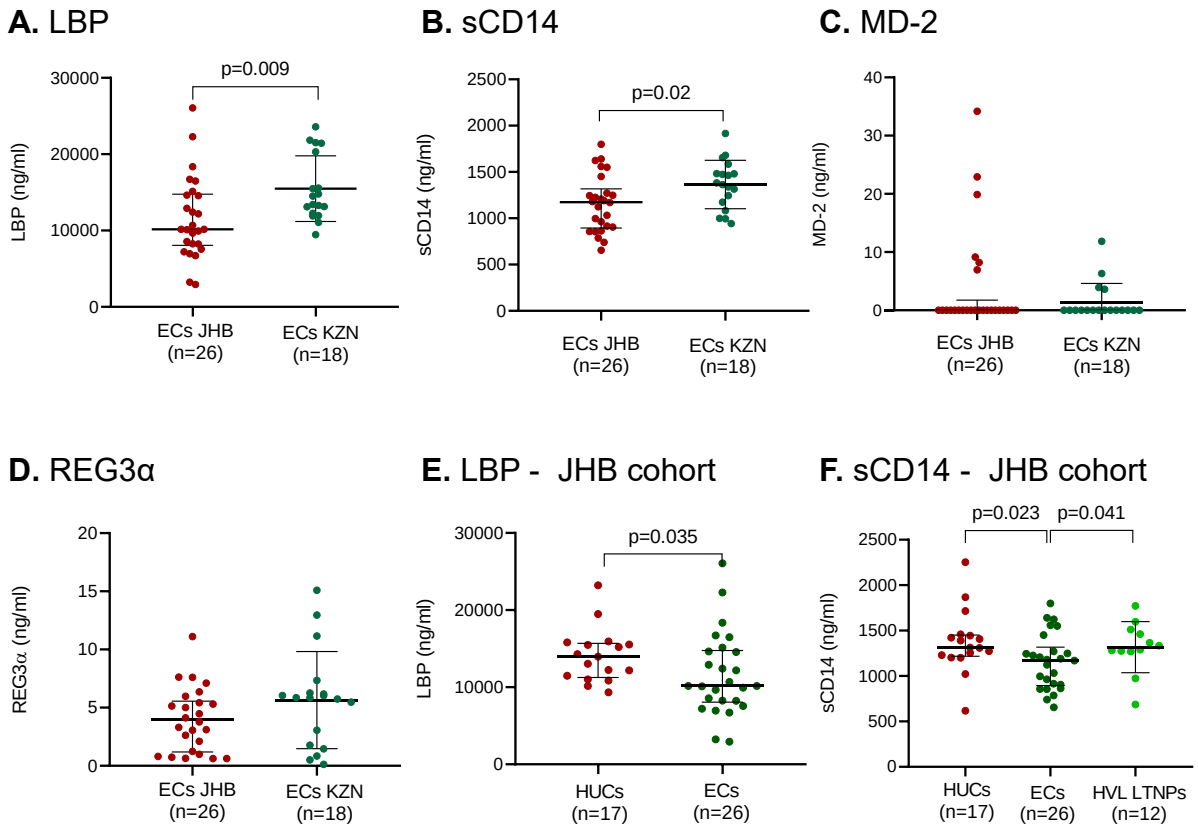


Figure 5.5: Comparisons of plasma levels of **(A)** LBP, **(B)** sCD14, **(C)** MD-2, and **(D)** REG3α between Johannesburg (JHB) elite controllers (ECs) and KwaZulu-Natal (KZN) ECs, **(E)** LBP plasma levels compared between HIV-1 uninfected controls (HUCs) and ECs from the JHB cohort, **(F)** soluble CD14 plasma levels compared between HIV-1 uninfected controls (HUCs), ECs and high viral load long-tern non-progressors (HVL LTNPs) in the JHB cohort (Kruskal-Wallis test – FDR-adjusted p > 0.05). Each dot represents a single individual and the horizontal lines represent the median and interquartile ranges. P values <0.05 are considered statistically significant.

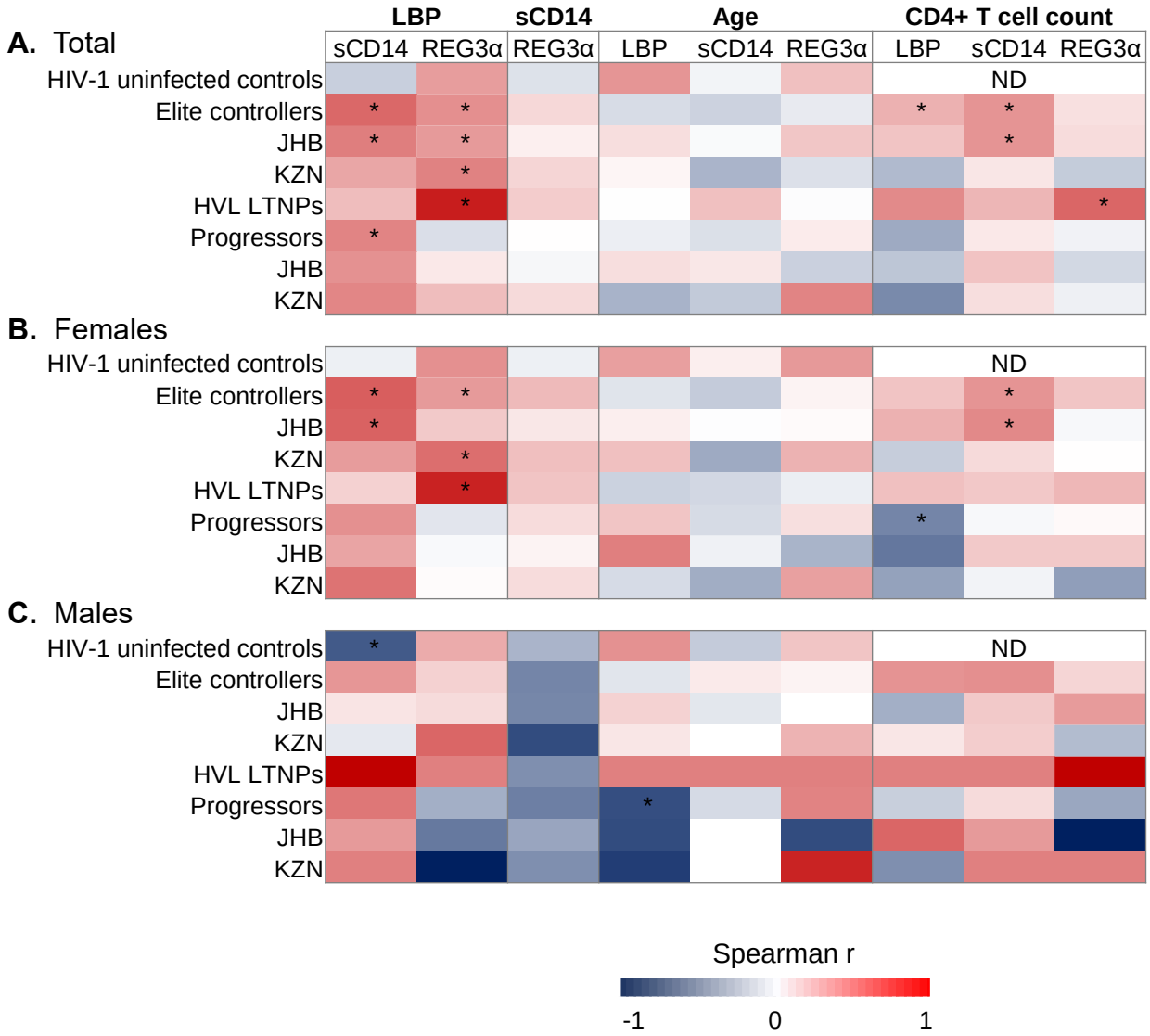


Figure 5.6: Relationships between plasma levels of LBP, sCD14, and REG3α, and with age and CD4+ T cell count. Correlations were done by Spearman's rank correlation coefficient. **(A)** Total HIV-1 uninfected controls, elite controllers, Johannesburg (JHB) ECs, KwaZulu-Natal (KZN) ECs, high viral load long-term non-progressors (HVL LTNPs), progressors, JHB progressors and KZN progressors. **(B)** Females and **(C)** Males. The colours represent the direction and strength of the correlation with blue representing a negative correlation ($r<0$), red representing a positive correlation ($r>0$) and white representing a lack of correlation ($r=0$). *$p<0.05$. ND: CD4+ T cell counts not done for HIV-1 uninfected controls.

5.3.6 Correlations between age, CD4+ T cell count, viral load, and LBP, sCD14, and REG3α plasma levels

Given the significant difference seen in age between elite controllers and progressors, and CD4+ T cell counts in JHB and KZN ECs, we also investigated the relationship between age and CD4+ T cell counts with the plasma levels of LBP, sCD14, MD-2, and REG3α (**Figure 5.6)**. LBP plasma levels negatively correlated with age in male progressors (p=-0.793, p=0.033). No correlations were seen between age and plasma levels of REG3α in all groups, but a weak negative correlation between plasma levels of sCD14 and age was seen in the total group of PLWH (r=-0.255, p=0.023, **Figure 5.7A**).

Levels of LBP (r=0.306, p=0.043) and sCD14 (r=0.423, p=0.004) positively correlated with CD4+ T cell counts in elite controllers, also present when elite controllers were stratified by the site of recruitment in the JHB ECs (r=0.420, p=0.033) (**Figure 5.6**). When stratified by sex, plasma levels of LBP negatively correlated with CD4+ T cell counts in female progressors (r=-0.544, p=0.029) and the sCD14 correlations with CD4+ T cell counts remained in females in the total group of elite controllers (r=0.421, p=0.01) and in the JHB ECs (r=0.465, p=0.045), but they were less significant. These correlations between LBP and sCD14 plasma levels with CD4+ T cell counts in elite controllers and sCD14 with age in the total group of PLWH could explain why we observed differences in the plasma levels of LBP and sCD14 between JHB and KZN ECs. Furthermore, REG3α plasma levels positively correlated with CD4+ T cell counts in HVL LTNPs (r=0.601, p=0.043), this correlation was also present in the total group of PLWH (r=0.338, p=0.002, **Figure 5.7**Error! Reference source not found.**B**).

There were no significant correlations observed between viral load and the plasma levels of LBP, sCD14, and REG3α in either HVL LTNPs or progressors even when analysed according to sex (**Figure 5.7C**).

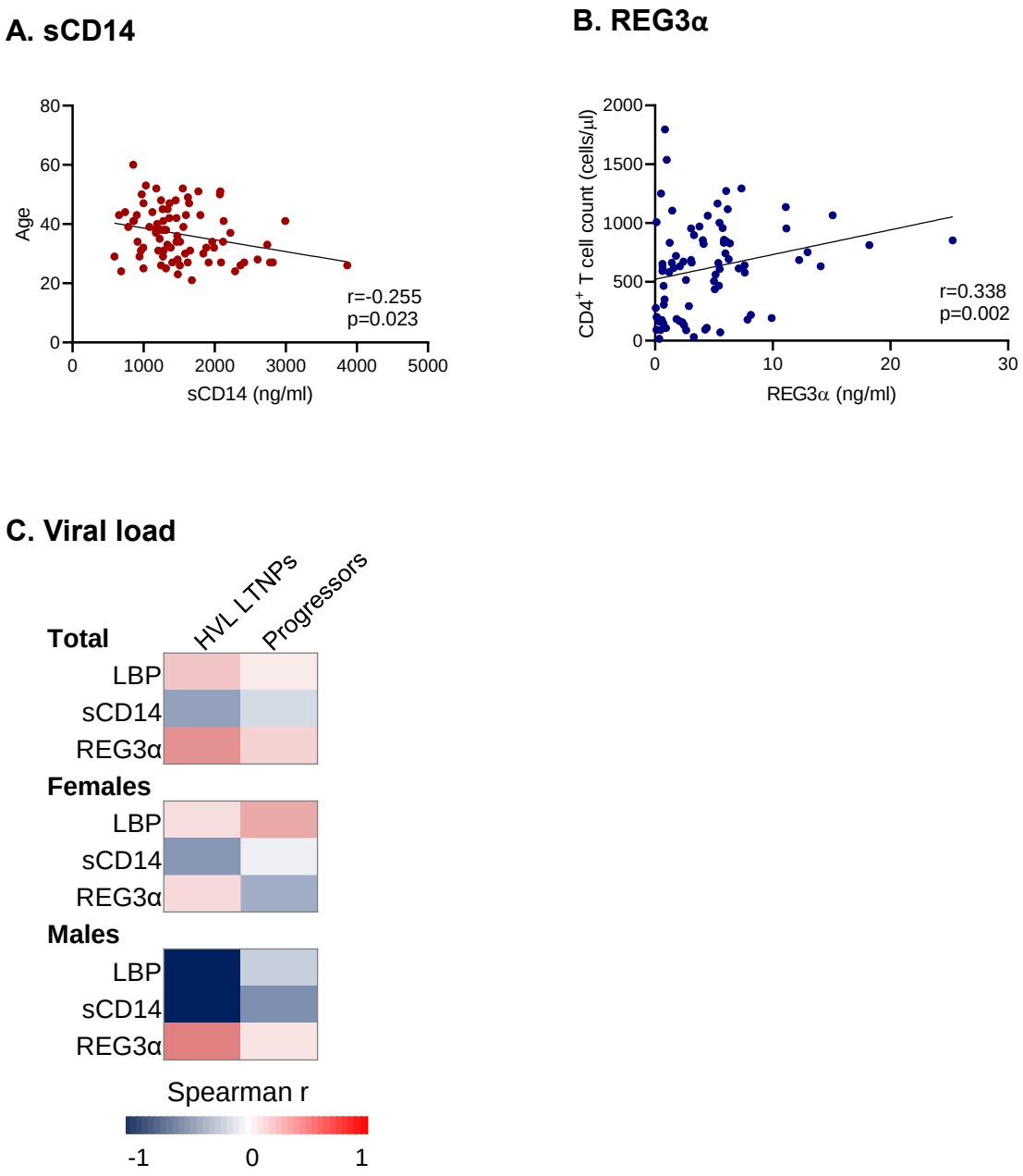


Figure 5.7: Correlations between **(A)** age and plasma levels of sCD14 in the total group of people living with HIV (PLWH), **(B)** $CD4^+$ T cell count and plasma levels of REG3α in the total group of PLWH, and **(C)** viral load and plasma levels of LBP, sCD14, and REG3α in the total group of high viral load long term non progressors (HVL LTNPs) and progressors, and in females and males. The colours represent the direction and strength of the correlation with blue representing a negative correlation (r<0), red representing a positive correlation (r>0) and white representing a lack of correlation (r=0).

5.4 Discussion

Immune activation is a predictor of HIV-1 disease progression (Giorgi et al., 1999, Giorgi et al., 2002) and the role of microbial translocation in immune activation in PLWH is well established (Brenchley et al., 2006, Marchetti et al., 2008, Nowroozalizadeh et al., 2010, Marchetti et al., 2011, Reus et al., 2013, Pilakka-Kanthikeel et al., 2014). Although ART has been shown to decrease levels of microbial translocation, the levels remain higher in PLWH on ART compared to HIV-1 uninfected controls (Brenchley et al., 2006, Cassol et al., 2010). While there have been numerous studies on microbial translocation in PLWH, studies in elite controllers and long-term non-progressors who maintain CD4+ T cell counts >500 cells/µl despite persistently high viral loads (called HVL LTNPs in our study) are limited, especially in African populations. We, therefore, quantified select innate immune molecules reported to be altered in plasma as a consequence of microbial translocation (LBP, sCD14, REG3α), or important in interacting with TLR4 (MD-2) – in black South African elite controllers, HVL LTNPs, progressors and healthy HIV-1 uninfected controls.

Soluble CD14 is a marker of monocyte activation, produced particularly in response to LPS or other microbial products binding to cell surface CD14 (Landmann et al., 1996, Shive et al., 2015). In PLWH, plasma levels of sCD14 independently predict mortality and disease progression (Marchetti et al., 2011, Sandler et al., 2011, Leon et al., 2015). Consistent with previous studies (Lien et al., 1998, Brenchley et al., 2006, Marchetti et al., 2011), we found progressive HIV-1 infection associated with significantly higher sCD14 plasma levels. Similar sCD14 levels in elite controllers and HIV-1 uninfected controls have been previously reported (Kim et al., 2014) and were also seen in the current study, while higher levels have been reported in elite controllers compared to controls in a separate study (Brenchley et al., 2006). Interestingly, when elite controllers were stratified by site of recruitment, the JHB elite controllers had significantly less sCD14 than KZN elite controllers, a feature that could be explained by the younger age of the latter group (as sCD14 correlated negatively with age in PLWH). Collectively, findings across all these studies including ours, suggest that the extent of monocyte activation in elite controllers represents a spectrum, ranging from those with less activation than HIV-1 uninfected controls to those with more elevated levels. Significantly elevated T cell activation is demonstrable in many elite controllers (Hunt et al., 2008, Pereyra et al., 2012, Picton et al., 2021), however, some studies have found no differences in T cell activation relative to HIV-1 uninfected controls (Sandler et al., 2011, Vesterbacka et al., 2017). A recent study of ours, studying a small group of HIV-1 controllers (elite controllers, viraemic

controllers with VLs <2000 RNA molecules/ml, LTNPs) showed significant T cell activation in these individuals but with no concomitant monocyte activation as measured by sCD14 plasma levels (Picton et al., 2021).

High VL LTNPs (also referred to as long-term asymptomatic or LTA-HVL; or viraemic non-progressors or VNPs) are a very rare phenotype in adults LWH. As a result, all studies on this subgroup of PLWH have small sample sizes (Choudhary et al., 2007, Rotger et al., 2011, Klatt et al., 2014, Negi et al., 2017). To the best of our knowledge, we have the largest group of HVL LTNPs described in current literature (n=11) (Gornalusse et al., 2015, Paximadis et al., 2017, Picton et al., 2017, Koor et al., 2019), however definitions do differ slightly. Our earlier studies did not include an additional case that we have included in the total group here (n=12). The host mechanisms that underlie preservation of CD4+ T cells in the presence of high viral loads remain to be fully elucidated. Natural nonhuman primate host species of SIV who despite high viral loads, preserve their CD4+ T cells, have no chronic immune activation and do not progress to simian AIDS (Brenchley et al., 2006, Choudhary et al., 2007). Importantly, they have no evidence of microbial translocation (Sandler et al., 2011). Our HVL LTNPs were viral load matched to progressors, yet their plasma levels of sCD14 were significantly lower and did not significantly differ from elite controllers or HIV-1 uninfected controls, supporting findings from previous studies that HVL LTNPs have low immune activation despite high viral loads (Choudhary et al., 2007). A study, which investigated a small group of extreme viraemic nonprogressors (VL >100,000 RNA copies/ml and CD4 >350 cells/µl for more than 3 years without treatment), found sCD14 plasma levels significantly lower in three of the four viraemic nonprogressors compared to rapid progressors (Rotger et al., 2011). However, another study conducted with Indian participants found that sCD14 plasma levels did not differ between rapid progressors, slow progressors (CD4+ T cells >500 cells/µl for >10 years with VL <10,000 RNA copies/ml) and viraemic slow progressors (n=5, CD4+ T cells >500 cells/µl for >7 years with VL >10,000 RNA copies/ml) (Negi et al., 2017). The latter group in this study (Negi et al., 2017) is closely aligned with our definition of HVL LTNPs. It is important to note that there are differences in ethnicity and environments across these studies and could in part account for some of the differences observed.

Soluble CD14 plasma levels in elite controllers and progressors, but not in HIV-1 uninfected controls and HVL LTNPs, correlated positively with plasma levels of LBP, which were not significantly different between HIV-1 controllers, progressors and HIV-1 uninfected controls. The correlation between LBP and sCD14 is indicative of microbial translocation in elite

controllers and progressors, suggesting that although microbial translocation may still be occurring in elite controllers, elite controllers have the ability to limit their levels of monocyte activation.

Although MD-2 is not considered a marker of microbial translocation, it is an essential molecule required for TLR4-mediated LPS responsiveness (Shimazu et al., 1999). MD-2 is either secreted directly as a soluble active protein or directly binds to TLR4 in the endoplasmic reticulum before the complex migrates to the cell surface (Visintin et al., 2001). Plasma levels of MD-2 have been reported to be elevated in PLWH (Sandanger et al., 2009, Trøseid et al., 2013). In contrast, we found lower levels of MD-2 in elite controllers compared to HIV-1 uninfected controls when individuals with undetectable levels were excluded. In addition, the comparable levels of MD-2 between HIV-1 controllers and progressors was in keeping with similar findings found in another report (Trøseid et al., 2013). It is nonetheless intriguing that MD-2 levels show this dichotomy of detectable/high vs undetectable in different individuals, regardless of whether living with HIV-1 or not - why this might be remains to be explained. This pattern of detectable vs undetectable MD-2 plasma levels has been observed in individuals with and without Sickle cell diseases in a study conducted in America (Zhang et al., 2021).

A recent study demonstrated that REG3α, used as a marker of gut damage, is elevated in the plasma of PLWH and highest in untreated progressive infection (Isnard et al., 2020). Elite controllers had higher levels compared to HIV-1 uninfected controls, but these were lower compared to progressors (Isnard et al., 2020). Contrary to their findings, we found the opposite - REG3α plasma levels were significantly higher in elite controllers and HVL LTNPs compared to progressors. Our findings were unexpected. The positive correlation between REG3α and LBP plasma levels in elite controllers and HVL LTNPs may suggest a link between microbial translocation and gut damage in these individuals. However, elite controllers have been shown to have lower levels of gut damage, measured by I-FABP when compared to progressors (Cheru et al., 2018). Isnard et al. (2020) also found a negative correlation between REG3α plasma levels and CD4+ T cell count, whereas we found a positive correlation between REG3α plasma levels and CD4+ T cell count in the total group of PLWH, mainly driven by HVL LTNPs, suggesting that higher REG3α plasma levels in our HIV-1 controller groups might be related to the high CD4+ T cell count in this group. Given that the REG3α plasma levels did not differ between HIV-1 uninfected controls and progressors, our findings could also suggest that there might be an upregulation of REG3α expression in elite controllers and HVL LTNPs, and that progressors are deficient in this regard. Elite controllers have been shown to have a

higher frequency of Th17 cells compared to HIV-1 uninfected controls and PLWH on ART (Kim et al., 2014, Caetano et al., 2020). Th17 cells are crucial in the maintenance of the integrity of the gut barrier, through the production of IL-17 and IL-22. The cytokines IL-17 and IL-22 are involved in epithelial regeneration, neutrophil recruitment, and the production of antimicrobial peptides (Liu et al., 2009). IL-17 and IL-22 are involved in the regulation of REG3α expression (Zheng et al., 2008, Lai et al., 2012).

A major difference between our study and the study that reported elevated plasma levels of REG3α in PLWH, including elite controllers compared to HIV-1 uninfected controls, is the distribution of males and female study participants (Isnard et al., 2020). In the Isnard et al. (2020) study, the participants were predominately male (>80%), while our HIV-1 controllers and progressors were predominantly female (>70%), while our HIV-1 uninfected controls had an equal distribution of males and females. This prompted us to investigate potential sex differences in the plasma levels of REG3α and the other soluble markers. Contrary to previous findings in HIV-1 uninfected controls, sCD14 (Reiner et al., 2013, Olson et al., 2020, Stanislawski et al., 2021) and MD-2 plasma levels did not show any sex differences. In contrast to the findings of Gonzalez-Quintela et al. (2013) that higher LBP plasma levels tended to be associated with males, we found no sex differences in LBP plasma levels in HIV-1 uninfected individuals, but female elite controllers tended to have higher LBP plasma levels.

We found in all groups that females had marked and significantly higher levels of REG3α compared to males. This is in keeping with higher levels of REG3α found in the small intestine tissues in females (n=67) compared to males (n=120) reported in the GTEx (The Genotype Tissue Expression project) portal (https://www.gtexportal.org). Our findings suggest that homeostatic levels of REG3α in the circulation are higher in females reflecting the same sex-difference evident in intestinal tissue. We propose that these higher basal levels may, in fact, provide protection from excessive activation and counter the effects of microbial translocation in elite controllers and the HVL LTNPs, and may also serve to explain why these groups are constituted by more females. Nonetheless, what is apparent is that, irrespective of sex, the same relationships held in both sexes in that progressors had reduced levels compared to elite controllers and HVL LTNPs. However, significance was lost in males when adjusted for multiple comparisons. This was perhaps not unsurprising given the smaller numbers of males in our study.

Our findings raise important questions about the functional role of REG3α in the context of HIV-1 infection. It is a multi-functional protein and aside from its antimicrobial properties, is involved in tissue repair, proliferation, apoptosis, and has an anti-inflammatory role in disease (Shin and Seeley, 2019, Wang et al., 2022). Our findings suggest that REG3α is not a reliable marker of gut damage in our populations of PLWH. We propose that progressors are unable to produce sufficient levels of this molecule in the gut (constitutively or in response to inflammatory stimuli), even in the presence of microbial translocation, as evidenced by increased levels of systemic activation (as measured by sCD14 levels). It may be the sufficient levels produced in the two controller groups, which did not differ from uninfected controls, that effectively counter immune activation in infection even in the presence of high viral loads in the HVL LTNP group. Whether the differences we see compared to the Isnard et al. (2020) study can in part be explained by ethnic differences between cohorts studied, or other environmental factors such as diet and exposure to pathogens, warrants further investigation. An additional point to keep in mind is that the ELISA kits used in our study and in the Isnard et al. (2020) study were from different manufacturers (Novus Biologicals vs R&D Systems, respectively) and of different type (ready-to-use vs DuoSet development assays, respectively). These differences in the ELISA kits used could result in differences in the sensitivity and specificity of the assay, however, our range of REG3α plasma levels in males was comparable to the range of REG3α plasma levels (>80% males) in the Isnard et al. (2020) study, suggesting similar performance of these assays.

Elite controllers are a heterogeneous group that shows differences in various factors including but not limited to levels of residual HIV-1 replication, loss of CD4+ T cells and duration of viral control (Navarrete-Muñoz et al., 2020). Our two groups of elite controllers, recruited from different sites (JHB and KZN), differed in age and CD4+ T cell count. Taken together with the positive correlation between both LBP and sCD14 with CD4+ T cells counts in elite controllers, and the negative correlation between age and sCD14 in PLWH, this could explain our finding of higher LBP and sCD14 plasma levels in KZN ECs. The importance of CD4+ T cells in differentiating between HIV-1 controllers has been shown in a study that stratified their elite controllers by the longitudinal data of CD4 percentage (Bansal et al., 2015). They found that elite controllers with a high percentage of CD4+ T cells, had lower expression levels of markers of immune activation, exhaustion, and immunosenescence on CD8+ T cells compared to elite controllers with a low percentage of CD4+ T cells and progressors, but the expression levels of the markers were not different compared to HIV-1 uninfected controls (Bansal et al., 2015).

The heterogeneity in elite controllers could in part explain the non-consensus findings with regards to T cell activation in elite controllers as well as the different findings in the levels of markers of microbial translocation in this group. For example, when we excluded the KZN ECs and only compared the JHB ECs to HIV-1 uninfected controls, the latter had significantly lower levels of both LBP and sCD14, suggesting that this cohort of elite controllers had monocyte activation levels lower than the HIV-1 uninfected controls.

The present study has some limitations worthy of mention. The duration of infection, which could influence levels of microbial translocation, was unknown for our cohorts of PLWH. Another limitation of this study is that we did not have information on comorbidities and lifestyle diseases which potentially impact the immune status and expression levels in our study participants. Additionally, we focused on indirect markers of microbial translocation (sCD14 and LBP), whose plasma levels could be influenced by other factors besides LPS. The inclusion of a direct marker of microbial translocation would have helped us better interpret our findings. We also did not include an additional well-established marker of intestinal damage along with REG3α, which could have provided additional insights into the extent of intestinal damage in our study groups. Regardless, our findings suggest that REG3α is not a useful marker of gut damage in our populations.

In conclusion, only plasma levels of sCD14 reflected the increased monocyte activation that occurs as a consequence of microbial translocation or other factors that occur with more progressive HIV-1 infection. Our findings have raised a number of important points that need to be considered with respect to levels of markers indicative of microbial translocation. Firstly, there may be differences in baseline levels in different populations that can influence the measured outcomes in the presence of HIV-1 infection. These include ethnicity (host genetics), and/or geographical location/environment. This was elegantly shown in the Redd et al. (2009) study, which highlighted many baseline differences in plasma levels of cytokines between uninfected individuals from the United States (predominantly African Americans) compared to Ugandan individuals. EndoCAb levels were substantially higher in the uninfected African individuals, while sCD14 levels were lower. Overall, in the context of HIV-1 infection, they showed no relationship between microbial translocation and disease progression using markers LPS, LBP and EndoCAb (Redd et al., 2009). Secondly, similarly as for population difference, sex differences in baseline levels cannot be ignored. We show that females have higher plasma levels of REG3α compared to males. Lastly, our findings have highlighted the importance of gaining a better understanding of the functions of the innate immune molecules themselves,

which clearly serve as markers of microbial translocation in some but not all studies. Our findings strongly emphasize this point for REG3α, a multi-functional protein with many attributes that could contribute towards a protective and anti-inflammatory role. It may be induced production and functionality of this molecule that is impaired in progressive HIV-1 infection, at least in our populations. Understanding the underlying mechanisms involved in the innate immune response in PLWH, particularly in rare groups such as elite controllers and the HVL LTNPs, will inform therapeutic strategies to reduce virus replication as well as the hazards of persistent immune activation.

CHAPTER 6

Summarising discussion and conclusion

Although antiretroviral therapy (ART) has significantly increased the life expectancy of people living with HIV-1 (PLWH), morbidity and mortality from both AIDS and non-AIDS defining illnesses remain higher in PLWH than in the general population (Sandler and Douek, 2012, de Coninck et al., 2018). Higher morbidity and mortality have been associated with immune activation and inflammation that persist even in those on suppressive ART (Hunt et al., 2016). Microbial translocation has been identified, among other factors, as a major cause of persistent immune activation and inflammation in PLWH (Brenchley et al., 2006). This phenomenon, however is not limited to HIV-1 infection, microbial translocation has also been implicated in obesity, diabetes, inflammatory bowel diseases (IBD), hepatitis B and C virus infections, fatty liver disease, pancreatitis, and cardiovascular diseases (Creely et al., 2007, Pussinen et al., 2011, Brenchley and Douek, 2012).

A small subset of PLWH can naturally control HIV-1 infection without the use of ART and are collectively known as HIV-1 controllers. These individuals comprise elite controllers (ECs) who naturally suppress the HIV-1 viral load (VL) to <50 RNA copies/ml (undetectable), viraemic controllers (VCs) who suppress the VL to <2000 RNA copies/ml, and high viral load long term non-progressors (HVL LTNPs) who can maintain CD4+ T cell counts >500 cells/µl in the presence of VL>10 000 RNA copies/ml for more than 7 years. Some studies have found no differences in immune activation levels between elite controllers and healthy HIV-1 uninfected individuals (Sandler et al., 2011, Vesterbacka et al., 2017). Other studies, however have reported higher levels of immune activation in elite controllers than ART-treated PLWH and healthy HIV-1 uninfected individuals (Hunt et al., 2008, Pereyra et al., 2012, Picton et al., 2021), suggesting that these individuals also remain at a higher risk of non-AIDS-defining illnesses (Pereyra et al., 2012). Elite controllers are a heterogeneous group; among them, some experience inconsistent suppressive control with viral load "blips" (Farmer et al., 2016), while others are considered exceptional elite controllers as they achieve viral control and no disease progression for more than 25 years (Casado et al., 2020). Nonetheless, HIV-1 controllers, specifically elite controllers represent a model for a functional cure, and thus an understanding of the contributors to immune activation in these rare individuals has important implications. Although evidence of microbial translocation has been seen in HIV-1 controllers (Brenchley et al., 2006), these studies are predominantly conducted in Caucasian-based population groups, and microbial translocation studies in HIV-1 controllers from African populations are scarce.

Microbial translocation involves the passage of microbial products such as lipopolysaccharide (LPS), lipoteichoic acid (LTA), ribosomal DNA, peptidoglycan, flagellin, and unmethylated

CpG-containing DNA from the gastrointestinal tract into the circulatory system (Alexander et al., 1990, Sandler and Douek, 2012). These microbial products elicit innate immune responses through the activation of toll-like receptors (TLR1 to TLR10) resulting in the production of pro-inflammatory cytokines such as tumour necrosis factor (TNF), type I interferons, interleukin-1β (IL-1β), and IL-6, which can induce inflammation (Sandler and Douek, 2012, Kawasaki and Kawai, 2014). The TLR4 signalling pathway, which recognises LPS, and requires the use of CD14, LBP, and MD-2 as coreceptors, has been the most extensively studied TLR pathway (Lu et al., 2008). In addition to LPS activation via the TLR4 signalling pathway, damage-associated molecular patterns (DAMPs) derived from tissue injury can activate the TLR4 receptor (Lee and Seong, 2009). The TLR4 signalling pathway, therefore, plays an essential role in inflammatory responses in both infectious and non-infectious diseases (Molteni et al., 2016).

Substantial evidence supports that genetic variation and expression levels of the molecules required for the activation of the TLR4 receptor by LPS (i.e, LBP, CD14, and MD-2) are associated with various immune-related diseases (Hubacek et al., 1999, Opal et al., 1999, Kondo et al., 2003, Cook et al., 2004, Gu et al., 2007, Simiantonaki et al., 2007, González-Reyes et al., 2010, Noreen et al., 2012, Liao et al., 2015a, Vidyant et al., 2019, Xu et al., 2019, Meng et al., 2021). Such association studies, however have been predominantly conducted in European, American, and Asian populations. Little is known about both the baseline expression levels and genetic variation in these molecules in African populations. This provides an important rationale for characterising the baseline expression levels of the LPS recognition and signalling molecules and the genetic variation in genes encoding these molecules in African populations.

In this thesis, we measured constitutive expression levels of the cell surface TLR4 and CD14 (mCD14) receptors as well as plasma levels of soluble CD14, LBP, and MD-2, in black and white South African populations and investigated the influence of ethnicity, sex, and age on these expression levels. We next sought to determine if TLR4 expression on monocytes and neutrophils might be differentially regulated post-stimulation with LPS and LTA according to population group and sex. Given that African populations exhibit greater genetic variability compared with non-African populations (Campbell and Tishkoff, 2008, Tucci and Akey, 2019), we characterised genetic variation and linkage disequilibrium patterns of the *LBP*, *CD14*, *TLR4*, and *LY96* genes in only the black South African population and compared to select populations from the 1000 Genomes Project – and assessed the representation of the

genetic variants in HIV-1 controllers (elite controllers, viraemic controllers, and HVL LTNPs). Finally, we measured and compared plasma levels of innate immune molecules considered as markers of microbial translocation (LBP, sCD14), gut damage (Reg3α), and a TLR4 signalling molecule (MD-2) in black South African elite controllers, HVL LTNPs, progressors, and healthy HIV-1 uninfected controls.

We found that only plasma levels of LBP exhibited population differences, with significantly higher levels in black South Africans. Soluble CD14 plasma levels, although not significant, showed the opposite of LBP, with a trend of lower sCD14 in black compared to white South Africans, consistent with findings from previous studies of lower levels reported in black compared to white individuals (Reiner et al., 2013, De Voeght et al., 2016, Olson et al., 2020). Due to the lack of a significant positive correlation between the plasma levels of LBP and sCD14 (a marker of monocyte activation) in the total group of black South Africans, we postulate that the LBP population differences could be due to other factors such as genetic differences, environmental and/or lifestyle factors that can affect background inflammation levels. To the best of our knowledge, population differences in plasma levels of LBP have not been previously reported. In serum of healthy Spanish individuals LBP levels correlate positively with levels of the pro-inflammatory cytokines IL-6 and IL-8 (Gonzalez-Quintela et al., 2013). Levels of systemic inflammation markers including, IL-6, CRP, E-selectin, and Fibrinogen are higher in healthy black compared to white individuals (Stepanikova et al., 2017), suggesting that background inflammation may be higher in black individuals. Furthermore, high levels of LBP in the serum have been associated with type 2 diabetes as well as risk factors for diabetes such as being overweight, obesity, the presence of metabolic syndrome, and low insulin sensitivity in different population groups (Moreno-Navarrete et al., 2012, Gonzalez-Quintela et al., 2013, Liu et al., 2014, Tilves et al., 2015, Kim et al., 2016). The prevalence of obesity and diabetes is higher in black Americans compared to their white counterparts (Ogden et al., 2015, Cheng et al., 2019). Likewise, studies conducted in South Africa have shown that black South Africans have lower insulin sensitivity, and a higher risk of obesity and diabetes compared with white individuals (Goedecke et al., 2009, Peer et al., 2018). These differences observed between black and white South Africans could be a cause or a consequence of the higher LBP plasma levels in black South Africans observed in this study. A limitation of the current study is that we did not obtain demographic information such as BMI, weight, and the fasting status at the time of blood collection. These factors influence the plasma levels of LBP (Gonzalez-Quintela et al., 2013), and would have

helped elucidate whether the differences observed in the LBP expression levels were influenced by differences in the demographics, as opposed to genetic differences – naturally it could be a combination of the two.

Interestingly, stimulation with LPS revealed differences in the regulation of TLR4 expression between black and white South African individuals. In black South Africans, a decrease in TLR4 expression on monocytes and neutrophils was observed, but an increase was observed in white South Africans. Among other factors, these differences in response to LPS stimulation could be related to the differences in the baseline expression levels of LBP - emphasising that baseline differences in expression levels of any molecule involved in LPS responsiveness could affect the functionality of the TLR4 receptor. A similar phenomenon was also seen in white South Africans, where we observed sex differences in baseline TLR4 expression levels, with white females having significantly higher TLR4 expression levels than males. With the stratification of white South Africans by sex, the upregulation of TLR4 expression on monocytes, and neutrophils by LPS was only present in females, suggesting that there are sex differences in TLR4 functionality related to the differences in the baseline TLR4 expression levels. These results require future studies to measure cytokines post-stimulation with LPS to determine if the population and sex differences observed in the TLR4 expression after LPS stimulation translate to immune response differences. Sex differences in TLR4 expression (Aomatsu et al., 2013) and response to LPS stimulation have been described (Aulock et al., 2006, Lefèvre et al., 2012, Aomatsu et al., 2013). Contrary to our findings, a previous study found higher TLR4 expression on neutrophils in males compared with females (Aomatsu et al., 2013). These data taken together show that there are population differences and population-specific sex differences, which have potential implications for innate immune responses and could result in different outcomes in diseases that are associated with dysregulation of the TLR4 signalling pathway.

Our study is the first to fully describe genetic variation and linkage disequilibrium patterns in the LPS recognition and signalling molecules (LBP, CD14, TLR4, and MD-2) in a black South Africa population (Chapters 3 and 4). We have identified several novel genetic variants across these genes - one, eight, four, and 10 novel variants in the *CD14*, *TLR4*, *LBP*, and *LY96* genes, respectively. However, most of these novel variants were present in single individuals and thus will require verification as true novel variants using other genotyping assays. Although we have shown that the variation and LD patterns in these genes, as expected, more closely resemble those of the YRI (West Africa) and LWK (East Africa) reference African populations

than the European (EUR) and Asian reference populations from the 1000 Genomes Project, there were still some significant differences between the SA black population and the African reference populations. The South African white population resembles the EUR population from the 1000 Genomes Project in other genes studied in our research unit (Paximadis et al., 2021). Given the genetic differences observed between the SA black population and the EUR 1000 Genomes Project population, it is reasonable to speculate that there would be genetic differences in the *LBP-CD14-TLR4-LY96* gene complex that result in differences in gene regulation of the LPS recognition molecules. Having demonstrated population-based differences in baseline LBP plasma levels in chapter 2, raises the question whether these differences may in part be attributed to differences in gene regulation between the populations. Thus, future studies to investigate the influence of genetic variation in these genes with expression levels of the respective molecules in healthy individuals are needed - these studies would inform us of the functional significance of these variants in the African context. Furthermore, genetic variants that are shared by African populations and distinctly different to non-African populations, can offer valuable insights into the risk and severity of various diseases, which are associated with the LPS recognition and signalling molecules, between these population groups.

Given that polymorphisms in the LPS recognition and signalling molecules have been associated with various diseases, we also investigated the representation of the genetic variants in these genes in a cohort of black SA HIV-1 controllers and compared them to a background HIV-1-unifected SA population (SA controls). To our knowledge, no previous studies have investigated the representation of these variants in HIV-1 controllers. We have demonstrated that the representations of several variants in these genes are different between HIV-1 controllers and SA controls. For *CD14*, we only found minor allele frequency (MAF) overrepresentation of a rare regulatory region SNP (rs186291587) in elite controllers compared with SA controls, with no differences seen in the other HIV-1 controller subgroups. This CD14 SNP is not a known expression quantitative trait loci (eQTL) based on the RegulomeDb rank of 4. With regard to *TLR4*, however, several variants were differentially represented between HIV-1 controllers and SA controls. Although these differential representations were observed in the total group of HIV-1 controllers and elite controllers relative to the SA controls, the most significant differences observed were when HVL LTNPs were compared to SA controls, suggesting a role of TLR4 in HIV-1 control in the presence of high viraemia. A frameshift mutation in the TLR4 gene, of SIV-infected sooty mangabeys, has been suggested to be a factor

contributing to the lower immune activation seen in these nonhuman primates despite high levels of viral replication (Palesch et al., 2018). The phenotype of HVL LTNPs is similar to SIV-infected sooty mangabeys. The most significant differences observed in this study were of two SNPs in complete linkage disequilibrium, a newly identified intronic variant (*TLR4* NI-2) and a 3' UTR SNP rs113017335. Data for the rs113017335 SNP were not available in the 1000 Genomes Project; however, data from the Genome Aggregation Database (gnomAD; https://registry.opendata.aws/broad-gnomad) on Ensembl shows this SNP to be a very rare SNP across all populations with a MAF of 0.02% in African populations. Further investigation of this variant in larger cohorts is required; however, given the scarcity of HVL LTNPs, it would be difficult to do so. According to the miRNASNP-v3 prediction tool, the rs113017335 SNP is predicted to result in a gain of binding sites for two microRNAs (miRNAs) and a loss of binding sites for four miRNAs. Whether or not the loss and gain of these miRNA binding sites results in lower or higher TLR4 expression requires experimental determination.

Contrary to what we saw in the representation of *TLR4* variants, for *LBP*, the differentially represented variants were restricted to the total group of controllers, elite controllers, and viraemic controllers, with no differences in HVL LTNPs, suggesting potential differences in the role of these variants in different modes of HIV-1 control. The most significant difference observed was the MAF overrepresentation of the rs1250247980 intronic SNP in the total group of HIV-1 controllers and viraemic controllers relative to the SA controls. Interestingly, this SNP was absent in the SA control population and present in the HIV-1 controllers, with the highest MAF (50%) in viraemic controllers. Although our sample size of viraemic controllers was small (n=6), the high MAF of the rs1250247980 SNP suggests that this variant may be important in viraemic control of HIV-1. Although no data are available for this SNP in the 1000 Genomes Project, according to gnomAD on Ensembl, this SNP is a very rare SNP across all populations with a MAF of 0.1% in African populations. The high prevalence in the viraemic controllers is thus worthy of further investigation using larger cohorts in addition to investigating its prevalence in individuals with more progressive HIV-1 infection.

Comparisons of *LY96* genetic variants revealed differentially represented variants in the total group of HIV-1 controllers, elite controllers, and HVL LTNPs compared with SA controls, with no differences seen in viraemic controllers. The intronic rs149605245 SNP showed the most significant association, with MAF overrepresentation in elite controllers compared with SA controls - this variant was found only in the elite controller subgroup of HIV-1 controllers (MAF: 7.1%) and absent in the SA controls. In the 1000 Genomes Project, this variant is

restricted to African populations with 4% MAF in the total African population. Again, this variant could have functional significance in HIV-1 elite control and is worthy of further investigation.

To evaluate if select genetic variants in the *LBP-CD14-TLR4-LY96* gene complex are associated with natural control of HIV-1, a more informative comparison would be to compare to HIV-1 progressing individuals (i.e., individuals failing to control infection), however, WGS data for such individuals were not available in this study. Nonetheless, HIV-1 controllers are rare individuals, accounting for less than 5% of PLWH, and elite controllers and HVL LTNPs represent less than 1% of PLWH. Taking these percentages into account, we can predict that among our SA controls, we are likely to have more individuals who would be HIV-1 progressors than controllers upon HIV-1 infection. Therefore, variants that are over- or underrepresented in HIV-1 controllers relative to the SA controls should not be ignored, and should be investigated in larger replication cohorts, ideally including HIV-1 progressors.

The activation of the TLR4 receptor on the cell surface results in the activation of two intracellular signalling pathways (Vaure and Liu, 2014) - the myeloid differentiation factor 88 (My88) dependent pathway, which results in the early activation of the NF-κβ transcription factor, and the TRIF-dependent pathway, which activates the interferon regulatory factor-3 (IRF3) transcription factor leading to the late activation of the NF-κβ (Maeshima and Fernandez, 2013, Vaure and Liu, 2014, Kuzmich et al., 2017). Thus, in addition to the genes encoding the receptors and accessory molecules (LBP, CD14, TLR4 and MD-2) required for the LPS recognition and signalling playing a role in LPS responsiveness, variants in the downstream signalling molecules can also affect innate immune responses to LPS and should also be considered in future studies that aim to understand the contribution of genetic variation in different populations to this complex yet important pathway.

Understanding the mechanisms of control in HIV-1 controllers, especially elite controllers will help inform future strategies for developing an HIV-1 functional cure. Elite controllers have lower levels of LPS, sCD14, and LBP compared with HIV-1 progressors (Brenchley et al., 2006). However, studies have yielded conflicting results when levels of microbial translocation in elite controllers are compared with healthy HIV-1 uninfected controls. Some studies have shown higher levels of LPS, sCD14, and LBP in elite controllers compared with HIV-1 uninfected controls (Brenchley et al., 2006); however, others did not find differences in LPS and sCD14 plasma levels between elite controllers and HIV-1 uninfected controls (Kim et al.,

2014). In this study, we quantified select markers of microbial translocation (LBP, sCD14), intestinal damage (REG3α), and a molecule important in interacting with TLR4 (MD-2) in black South African elite controllers, HVL LTNPs, progressors, and healthy HIV-1 uninfected controls. Plasma levels of LBP and MD-2 did not differ between any groups compared. The lack of differences in the plasma level of LBP between elite controllers, HVL LTNPs, progressors, and healthy HIV-1 uninfected controls, suggests that LBP may not be a good marker of microbial translocation in the South African population.

In agreement with previous studies (Lien et al., 1998, Brenchley et al., 2006, Choudhary et al., 2007, Marchetti et al., 2011), plasma levels of sCD14 were higher in progressors compared to healthy HIV-1 uninfected controls and HIV-1 controllers (elite controllers, HVL LTNPs). This result is consistent with high plasma levels of sCD14 associated with progressive HIV-1 infection. Previous studies have found sCD14 plasma levels to be independent predictors of mortality and disease progression in PLWH (Marchetti et al., 2011, Sandler et al., 2011, Leon et al., 2015). Our finding of no difference in sCD14 plasma levels between elite controllers and HIV-1 uninfected controls is consistent with previous findings (Kim et al., 2014). However, direct comparisons of studies in HIV-1 controllers, specifically elite controllers and HVL LTNPs, are difficult due to the heterogeneity of HIV-1 controllers, differences in HIV-1 control definitions, and differences in lengths of follow-up. In this study, the heterogeneity of elite controllers was evident when elite controllers recruited from different sites in South Africa (Johannesburg and KwaZulu-Natal) showed differences in plasma levels of LBP and sCD14. Although we could relate the differences to the differences in age and CD4+ T cell count between the two groups of elite controllers, these results demonstrate that elite controls can differ, not just between populations but also within the same populations. It is noteworthy in this regard that previous studies have shown conflicting results concerning levels of T cell activation in elite controllers relative to HIV-1 uninfected controls (Hunt et al., 2008, Pereyra et al., 2012, Vesterbacka et al., 2017, Picton et al., 2021). Therefore, when comparing or extrapolating findings from HIV-1 controllers, the heterogeneity in these individuals should be considered.

In our HVL LTNP subgroup of controllers, regardless of the high viral loads, which normally associate with elevated levels of microbial translocation, plasma levels of sCD14 (an indirect marker of microbial translocation and marker of monocyte activation) did not differ significantly compared to HIV-1 uninfected controls and elite controllers, but were significantly lower compared with viral load-matched HIV-1 progressors. This important result

suggests that HVL LTNPs have the means to limit microbial translocation-related monocyte activation. The HVL LTNP phenotype resembles that of SIV-infected sooty mangabeys. In these nonhuman primates that do not progress to AIDS despite high levels of viral replication, chronic immune activation and microbial translocation are absent (Brenchley et al., 2006, Choudhary et al., 2007). Polymorphisms in the *LBP*, *CD14*, *TLR4*, and *LY96* genes may modulate immune responses to translocated LPS, and the *TLR4* polymorphisms found to be differentially represented in the HVL LTNPs relative to healthy controls in Chapter 3 could in part be influence the innate immune responses to microbial translocation in HVL LTNPs.

A study published in 2020, reported for the first time that plasma levels of REG3α, an antimicrobial peptide were higher in people living with HIV-1, including elite controllers compared with HIV-1 uninfected individuals, however, elite controllers had lower levels compared with progressors (Isnard et al., 2020). Unexpectedly, in our study (Chapter 5) we found higher plasma levels of REG3α in elite controllers and HVL LTNPs, compared with progressors, but the plasma levels did not differ significantly when elite controllers, HVL LTNPs and progressors were compared with HIV-1 uninfected controls. These differences were maintained in both males and females when our study participants were stratified by sex, suggesting that the differences in the distribution of males and females in ours and the Isnard et al. (2020) study did not account for the differences in our findings. Another major finding in our study was that REG3α plasma levels were significantly higher in females compared to males, regardless of the HIV-1 infection status. Given that elite controllers have been shown to have lower levels of gut damage compared to progressors as measured by I-FABP, which is a well-established marker of intestinal (enterocyte) damage, our results suggest that REG3α might not be a good marker of gut damage in the black South African population. In fact, the lack of differences in the plasma levels of REG3α between HIV-1 uninfected controls and progressors suggests that the REG3α expression might be upregulated in elite controllers and HVL LTNPs – seemingly as a protective response that is otherwise impaired in individuals with progressive infection. Although population differences and environmental factors such as diet and exposure to pathogens could account for the differences observed between our study and the Isnard et al. (2020) study, this remains to be investigated. In addition, a limitation of this study is that we did not include other markers of gut damage, which would provide additional insights into the extent of intestinal damage in our study participants. Therefore, future studies to determine the extent of gut damage in these same cohorts are required. Furthermore, incorporating studies assessing the influence of genetic variants in the *LBP*-

CD14-TLR4-LY96 gene complex on the plasma levels of these soluble markers in HIV-1 controllers, progressors as well as uninfected controls would be highly informative. Given our intriguing findings with REG3α, taking a similar approach as illustrated in Chapters 3 and 4, exploring the genetic variation of *REG3α* may provide important insights into this molecule and its expression in different populations, and in the context of HIV-1 infection.

This study had a number of limitations. Our sample sizes were small, especially for black and white South Africans when investigating the influence of sex on the constitutive expression levels, and for analyses involving viraemic controllers. Regardless of the small numbers, our findings form a strong basis for future investigations. In addition, we did not correct for multiple testing when assessing the representation of *LBP*, *CD14*, *TLR4*, and *LY96* genetic variants in HIV-1 controllers. Given the high number of genetic variants present in each gene, none of our significant associations would have been maintained. However, given the exploratory nature of this study, these results are still informative. The addition of HIV-1 progressors in the cohorts used for the genetic association comparisons would have provided additional information on the *LBP*, *CD14*, *TLR4*, and *LY96* genetic variants, particularly variants under- or over-represented in this group relative to other groups would identify further variants that will not necessarily have come to the fore in comparisons of HIV-1 controller groups with the SA controls in the context of natural control of HIV-1 infection. Finally, we did not include direct markers of microbial translocation or additional markers of intestinal damage, these would have given more insights into the state of microbial translocation and gut damage in our PLWH and HIV-1 uninfected controls.

In conclusion, findings from this thesis have substantially expanded our understanding of the baseline expression levels and genetic variation in *LBP*, *CD14*, *TLR4*, and *LY96* genes in the South African context. This work is also the first to show differences in the representation of the *LBP*, *CD14*, *TLR4*, and *LY96* genetic variants in black South African HIV-1 controllers. In addition, our findings highlight the importance of taking ethnicity, sex, and age into consideration when measuring levels of these biological markers. Differences in expression levels of the TLR4 signalling molecules have implications in the risk and severity of various immune-related diseases. Any outcomes of infection/disease reflect a combination of an individual's genotypes and environmental factors which influence phenotypic expression of these genotypes. Further understanding of the complex and combinatory interactions that underlie regulation of the TLR4 immune axis with its signalling molecules, will help inform

strategies to curb persistent immune activation in many disease contexts including that of HIV-1 infection.

REFERENCES

Abad-Fernández, M., Vallejo, A., Hernández-Novoa, B., Díaz, L., Gutiérrez, C., et al. 2013. Correlation between different methods to measure microbial translocation and its association with immune activation in long-term suppressed HIV-1-infected individuals. *Journal of acquired immune deficiency syndromes* 64**,** 149-153.

Abeles, R. D., McPhail, M. J., Sowter, D., Antoniades, C. G., Vergis, N., et al. 2012. CD14, CD16 and HLA-DR reliably identifies human monocytes and their subsets in the context of pathologically reduced HLA-DR expression by CD14(hi) /CD16(neg) monocytes: Expansion of CD14(hi) /CD16(pos) and contraction of CD14(lo) /CD16(pos) monocytes in acute liver failure. *Cytometry A,* 81**,** 823-834.

Abreu, M. T., Vora, P., Faure, E., Thomas, L. S., Arnold, E. T., et al. 2001. Decreased expression of Toll-like receptor-4 and MD-2 correlates with intestinal epithelial cell protection against dysregulated proinflammatory gene expression in response to bacterial lipopolysaccharide. *Journal of immunology,* 167**,** 1609-1616.

Adland, E., Paioni, P., Thobakgale, C., Laker, L., Mori, L., et al. 2015. Discordant Impact of HLA on Viral Replicative Capacity and Disease Progression in Pediatric and Adult HIV Infection. *PLoS pathogens,* 11**,** e1004954-e1004954.

Adriaanse, M. P., Tack, G. J., Passos, V. L., Damoiseaux, J. G., Schreurs, M. W., et al. 2013. Serum I-FABP as marker for enterocyte damage in coeliac disease and its relation to villous atrophy and circulating autoantibodies. *Alimentary pharmacology & therapeutics,* 37**,** 482-490.

Agnese, D. M., Calvano, J. E., Hahm, S. J., Coyle, S. M., Corbett, S. A., et al. 2002. Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with an increased risk of gram-negative infections. *The Journal of infectious diseases,* 186**,** 1522-1525.

Al-Saffar, A. K., Meijer, C. H., Gannavarapu, V. R., Hall, G., Li, Y., et al. 2017. Parallel Changes in Harvey-Bradshaw Index, TNFα, and Intestinal Fatty Acid Binding Protein in Response to Infliximab in Crohn's Disease. *Gastroenterology Research and Practice,* 2017**,** 1745918.

Alexander, J. W., Boyce, S. T., Babcock, G. F., Gianotti, L., Peck, M. D., et al. 1990. The process of microbial translocation. *Annals of surgery,* 212**,** 496-512.

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., et al. 2018. Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global health action,* 11**,** 1507133.

Alter, G., Malenfant, J. M., Delabre, R. M., Burgett, N. C., Yu, X. G., et al. 2004. Increased natural killer cell activity in viremic HIV-1 infection. *Journal of immunology,* 173**,** 5305-5311.

Ancuta, P., Kamat, A., Kunstman, K. J., Kim, E. Y., Autissier, P., et al. 2008. Microbial translocation is associated with increased monocyte activation and dementia in AIDS patients. *PLoS One,* 3**,** e2516.

Andrade, A., Bailey, J. R., Xu, J., Philp, F. H., Quinn, T. C., et al. 2008. CD4+ T cell depletion in an untreated HIV type 1-infected human leukocyte antigen-B*5801-positive patient with an undetectable viral load. *Clinical infectious diseases,* 46**,** e78-82.

Aomatsu, M., Kato, T., Kasahara, E. & Kitagawa, S. 2013. Gender difference in tumor necrosis factor-α production in human neutrophils stimulated by lipopolysaccharide and interferon-γ. *Biochemical and biophysical research communications,* 441**,** 220-225.

Appay, V. & Sauce, D. 2008. Immune activation and inflammation in HIV-1 infection: causes and consequences. *The Journal of pathology,* 214**,** 231-241.

Arbour, N. C., Lorenz, E., Schutte, B. C., Zabner, J., Kline, J. N., et al. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nature genetics,* 25**,** 187-191.

Arts, E. J. & Hazuda, D. J. 2012. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine,* 2**,** a007161-a007161.

Aulock, S. V., Deininger, S., Draing, C., Gueinzius, K., Dehus, O., et al. 2006. Gender difference in cytokine secretion on immune stimulation with LPS and LTA. *Journal of interferon & cytokine research,* 26**,** 887-892.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., et al. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

Baker, A. R., Qiu, F., Randhawa, A. K., Horne, D. J., Adams, M. D., et al. 2012. Genetic variation in TLR genes in Ugandan and South African populations and comparison with HapMap data. *PLoS One,* 7**,** e47597.

Balagopal, A., Ray, S. C., De Oca, R. M., Sutcliffe, C. G., Vivekanandan, P., et al. 2009. Kupffer cells are depleted with HIV immunodeficiency and partially recovered with antiretroviral immune reconstitution. *AIDS* 23**,** 2397-2404.

Baldini, M., Lohman, I. C., Halonen, M., Erickson, R. P., Holt, P. G., et al. 1999. A Polymorphism* in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *American journal of respiratory cell and molecular biology,* 20**,** 976-983.

Bank, S., Skytt Andersen, P., Burisch, J., Pedersen, N., Roug, S., et al. 2014. Polymorphisms in the inflammatory pathway genes TLR2, TLR4, TLR9, LY96, NFKBIA, NFKB1, TNFA, TNFRSF1A, IL6R, IL10, IL23R, PTPN22, and PPARG are associated with susceptibility of inflammatory bowel disease in a Danish cohort. *PloS one,* 9**,** e98815-e98815.

Bansal, A., Sterrett, S., Erdmann, N., Westfall, A. O., Dionne-Odom, J., et al. 2015. Normal T-cell activation in elite controllers with preserved CD4+ T-cell counts. *AIDS,* 29**,** 2245-2254.

Baroncelli, S., Galluzzo, C. M., Pirillo, M. F., Mancini, M. G., Weimer, L. E., et al. 2009. Microbial translocation is associated with residual viral replication in HAART-treated HIV+ subjects with <50copies/ml HIV-1 RNA. *Journal of Clinical Virology,* 46**,** 367-370.

Beamer, L. J., Carroll, S. F. & Eisenberg, D. 1998. The BPI/LBP family of proteins: a structural analysis of conserved regions. *Protein science* 7**,** 906-914.

Beaulieu, E., Green, L., Elsby, L., Alourfi, Z., Morand, E. F., et al. 2011. Identification of a novel cell type-specific intronic enhancer of macrophage migration inhibitory factor (MIF) and its regulation by mithramycin. *Clinical and experimental immunology,* 163**,** 178-188.

Beenakker, K. G. M., Westendorp, R. G. J., de Craen, A. J. M., Chen, S., Raz, Y., et al. 2020. Men Have a Stronger Monocyte-Derived Cytokine Production Response upon Stimulation with the Gram-Negative Stimulus Lipopolysaccharide than Women: A Pooled Analysis Including 15 Study Populations. *Journal of Innate Immunity,* 12**,** 142-153.

Bell, A. C., West, A. G. & Felsenfeld, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell,* 98**,** 387-396.

Biswas, S. K. & Lopez-Collazo, E. 2009. Endotoxin tolerance: new mechanisms, molecules and clinical significance. *Trends in immunology,* 30**,** 475-487.

Boufassa, F., Lechenadec, J., Meyer, L., Costagliola, D., Hunt, P. W., et al. 2014. Blunted response to combination antiretroviral therapy in HIV elite controllers: an international HIV controller collaboration. *PLoS One,* 9**,** e85516.

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., et al. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research,* 22**,** 1790-1797.

Brenchley, J. M. & Douek, D. C. 2012. Microbial translocation across the GI tract. *Annual review of immunology,* 30**,** 149-173.

Brenchley, J. M., Paiardini, M., Knox, K. S., Asher, A. I., Cervasi, B., et al. 2008. Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections. *Blood,* 112**,** 2826-2835.

Brenchley, J. M., Price, D. A., Schacker, T. W., Asher, T. E., Silvestri, G., et al. 2006. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nature medicine,* 12**,** 1365-1371.

Broadley, S. P., Plaumann, A., Coletti, R., Lehmann, C., Wanisch, A., et al. 2016. Dual-Track Clearance of Circulating Bacteria Balances Rapid Restoration of Blood Sterility with Induction of Adaptive Immunity. *Cell host & microbe,* 20**,** 36-48.

Bruunsgaard, H., Pedersen, A. N., Schroll, M., Skinhoj, P. & Pedersen, B. K. 1999. Impaired production of proinflammatory cytokines in response to lipopolysaccharide (LPS) stimulation in elderly humans. *Clinical and experimental immunology,* 118**,** 235-241.

Burton, T., Liang, B., Dibrov, A. & Amara, F. 2002. Transforming growth factor-beta-induced transcription of the Alzheimer beta-amyloid precursor protein gene involves interaction between the CTCF-complex and Smads. *Biochemical and biophysical research communications,* 295**,** 713-723.

Caetano, D. G., de Paula, H. H. S., Bello, G., Hoagland, B., Villela, L. M., et al. 2020. HIV-1 elite controllers present a high frequency of activated regulatory T and Th17 cells. *PLoS One,* 15**,** e0228745.

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

Canouï, E., Lécuroux, C., Avettand-Fenoël, V., Gousset, M., Rouzioux, C., et al. 2017. A Subset of Extreme Human Immunodeficiency Virus (HIV) Controllers Is Characterized by a Small HIV Blood Reservoir and a Weak T-Cell Activation Level. *Open forum infectious diseases,* 4**,** ofx064.

Carty, M., Goodbody, R., Schroder, M., Stack, J., Moynagh, P. N., et al. 2006. The human adaptor SARM negatively regulates adaptor protein TRIF-dependent Toll-like receptor signaling. *Nature immunology,* 7**,** 1074-1081.

Casado, C., Galvez, C., Pernas, M., Tarancon-Diez, L., Rodriguez, C., et al. 2020. Permanent control of HIV-1 pathogenesis in exceptional elite controllers: a model of spontaneous cure. *Scientific Reports,* 10**,** 1902.

Cassol, E., Malfeld, S., Mahasha, P., van der Merwe, S., Cassol, S., et al. 2010. Persistent microbial translocation and immune activation in HIV-1-infected South Africans receiving combination antiretroviral therapy. *The Journal of infectious diseases,* 202**,** 723-733.

Chattopadhyay, A. & Lu, T. P. 2020. Overcoming the challenges of imputation of rare variants in a Taiwanese cohort. *Translational cancer research,* 9**,** 4065-4069.

Cheng, Y. J., Kanaya, A. M., Araneta, M. R. G., Saydah, S. H., Kahn, H. S., et al. 2019. Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016. *The Journal of the American Medical Association,* 322**,** 2389-2398.

Cheru, L. T., Park, E. A., Saylor, C. F., Burdo, T. H., Fitch, K. V., et al. 2018. I-FABP Is Higher in People With Chronic HIV Than Elite Controllers, Related to Sugar and Fatty Acid Intake and Inversely Related to Body Fat in People With HIV. *Open forum infectious diseases,* 5**,** ofy288.

Chien, J. W., Boeckh, M. J., Hansen, J. A. & Clark, J. G. 2008. Lipopolysaccharide binding protein promoter variants influence the risk for Gram-negative bacteremia and mortality after allogeneic hematopoietic cell transplantation. *Blood,* 111**,** 2462-2469.

Choudhary, S. K., Vrisekoop, N., Jansen, C. A., Otto, S. A., Schuitemaker, H., et al. 2007. Low immune activation despite high levels of pathogenic human immunodeficiency virus type 1 results in long-term asymptomatic disease. *Journal of virology,* 81**,** 8838-8842.

Clavel, F. & Hance, A. J. 2004. HIV Drug Resistance. *New England Journal of Medicine,* 350**,** 1023-1035.

Coffin, J. & Swanstrom, R. 2013. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine,* 3**,** a012526-a012526.

Cohen, I. R. & Norins, L. C. 1966. Natural human antibodies to gram-negative bacteria: immunoglobulins G, A, and M. *Science,* 152**,** 1257-1259.

Connolly, N., Riddler, S. & Rinaldo, C. 2005. Proinflammatory cytokines in HIV disease - A review and rationale for new therapeutic approaches. *AIDS reviews,* 7**,** 168-180.

Cook, D. N., Pisetsky, D. S. & Schwartz, D. A. 2004. Toll-like receptors in the pathogenesis of human disease. *Nature immunology,* 5**,** 975-979.

Creely, S. J., McTernan, P. G., Kusminski, C. M., Fisher f, M., Da Silva, N. F., et al. 2007. Lipopolysaccharide activates an innate immune system response in human adipose tissue in obesity and type 2 diabetes. *American journal of physiology. Endocrinology and metabolism,* 292**,** E740-747.

da Rocha, J. E. B., Othman, H., Botha, G., Cottino, L., Twesigomwe, D., et al. 2021. The Extent and Impact of Variation in ADME Genes in Sub-Saharan African Populations. *Frontiers in Pharmacology,* 12.

Dauphinee, S. M. & Karsan, A. 2006. Lipopolysaccharide signaling in endothelial cells. *Laboratory Investigation,* 86**,** 9-22.

Davey, R. T., Bhat, N., Yoder, C., Chun, T.-W., Metcalf, J. A., et al. 1999. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proceedings of the National Academy of Sciences,* 96**,** 15109-15114.

de Coninck, Z., Hussain-Alkhateeb, L., Bratt, G., Ekström, A. M., Gisslén, M., et al. 2018. Non-AIDS Mortality Is Higher Among Successfully Treated People Living with HIV Compared with Matched HIV-Negative Control Persons: A 15-Year Follow-Up Cohort Study in Sweden. *AIDS Patient Care STDS,* 32**,** 297-305.

De Maeyer, R. P. H. & Chambers, E. S. 2021. The impact of ageing on monocytes and macrophages. *Immunology Letters,* 230**,** 1-10.

De Voeght, A., Maes, N. & Moutschen, M. 2016. sCD14 is not a bona-fide biomarker of microbial translocation in HIV-1-infected Africans living in Belgium. *AIDS,* 30**,** 921-924.

Deeks, S. G., Archin, N., Cannon, P., Collins, S., Jones, R. B., et al. 2021. Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy 2021. *Nature Medicine,* 27**,** 2085-2098.

Deeks, S. G., Lewin, S. R. & Havlir, D. V. 2013. The end of AIDS: HIV infection as a chronic disease. *The Lancet,* 382**,** 1525-1533.

Deeks, S. G. & Walker, B. D. 2007. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity,* 27**,** 406-416.

Dias, B. D. C., Paximadis, M., Martinson, N., Chaisson, R. E., Ebrahim, O., et al. 2020. The impact of bone marrow stromal antigen-2 (BST2) gene variants on HIV-1 control in black South African individuals. *Infection, genetics and evolution,* 80**,** 104216.

Dong, S., Zhao, N., Spragins, E., Kagda, M. S., Li, M., et al. 2022. Annotating and prioritizing human non-coding variants with RegulomeDB. *bioRxiv***,** 2022.2010.2018.512627.

Durack, D. T. 1981. Opportunistic infections and Kaposi's sarcoma in homosexual men. *The New England journal of medicine,* 305**,** 1465-1467.

Dziarski, R. & Gupta, D. 2005. Peptidoglycan recognition in innate immunity. *Journal of endotoxin research,* 11**,** 304-310.

Dziarski, R., Wang, Q., Miyake, K., Kirschning, C. J. & Gupta, D. 2001. MD-2 enables Toll-like receptor 2 (TLR2)-mediated responses to lipopolysaccharide and enhances TLR2-mediated responses to Gram-positive and Gram-negative bacteria and their cell wall components. *Journal of immunology,* 166**,** 1938-1944.

Eckert, J. K., Kim, Y. J., Kim, J. I., Gurtler, K., Oh, D. Y., et al. 2013. The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity. *Immunity,* 39**,** 647-660.

Eggena, M. P., Barugahare, B., Jones, N., Okello, M., Mutalya, S., et al. 2005. Depletion of Regulatory T Cells in HIV Infection Is Associated with Immune Activation. *The Journal of Immunology,* 174**,** 4407-4414.

Eggleton JS, Nagalli S. Highly Active Antiretroviral Therapy (HAART) [Updated 2020 Feb 1]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK554533.

El-Zayat, S. R., Sibaii, H. & Mannaa, F. A. 2019. Toll-like receptors activation, signaling, and targeting: an overview. *Bulletin of the National Research Centre,* 43**,** 187.

El Kamari, V., Sattar, A. & McComsey, G. A. 2019. Brief Report: Gut Structural Damage: an Ongoing Process in Chronically Untreated HIV Infection. *Journal of acquired immune deficiency syndromes,* 80**,** 242-245.

Elsbach, P. & Weiss, J. 1995. Prospects for use of recombinant BPI in the treatment of gram-negative bacterial infections. *Infectious agents and disease,* 4**,** 102-109.

Eng, H. L., Wang, C. H., Chen, C. H., Chou, M. H., Cheng, C. T., et al. 2004. A CD14 promoter polymorphism is associated with CD14 expression and Chlamydia-stimulated TNF alpha production. *Genes and immunity,* 5**,** 426-430.

Epple, H. J., Allers, K., Troger, H., Kuhl, A., Erben, U., et al. 2010. Acute HIV infection induces mucosal infiltration with CD4+ and CD8+ T cells, epithelial apoptosis, and a mucosal barrier defect. *Gastroenterology,* 139**,** 1289-1300.

Fairweather, D. & Rose, N. R. 2004. Women and autoimmune diseases. *Emerging infectious diseases,* 10**,** 2005-2011.

Farmer, A., Wang, X., Ganesan, A., Deiss, R. G., Agan, B. K., et al. 2016. Factors associated with HIV viral load "blips" and the relationship between self-reported adherence and efavirenz blood levels on blip occurrence: a case–control study. *AIDS Research and Therapy,* 13**,** 16.

Farzadegan, H., Hoover, D. R., Astemborski, J., Lyles, C. M., Margolick, J. B., et al. 1998. Sex differences in HIV-1 viral load and progression to AIDS. *The Lancet,* 352**,** 1510-1514.

Ferwerda, B., Kibiki, G. S., Netea, M. G., Dolmans, W. M. & van der Ven, A. J. 2007. The toll-like receptor 4 Asp299Gly variant and tuberculosis susceptibility in HIV-infected patients in Tanzania. *AIDS,* 21**,** 1375-1377.

Fitzgerald, F. C., Lhomme, E., Harris, K., Kenny, J., Doyle, R., et al. 2018. Microbial translocation does not drive immune activation in Ugandan children with HIV. *Journal of Infectious Diseases*.

Flores, C., Perez-Mendez, L., Maca-Meyer, N., Muriel, A., Espinosa, E., et al. 2009. A common haplotype of the LBP gene predisposes to severe sepsis. *Critical care medicine,* 37**,** 2759-2766.

Frey, E. A., Miller, D. S., Jahr, T. G., Sundan, A., Bazil, V., et al. 1992. Soluble CD14 participates in the response of cells to lipopolysaccharide. *The Journal of experimental medicine,* 176**,** 1665-1671.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., et al. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature,* 397**,** 436-441.

Gegner, J. A., Ulevitch, R. J. & Tobias, P. S. 1995. Lipopolysaccharide (LPS) signal transduction and clearance. Dual roles for LPS binding protein and membrane CD14. *The Journal of biological chemistry,* 270**,** 5320-5325.

Giorgi, J. V., Hultin, L. E., McKeating, J. A., Johnson, T. D., Owens, B., et al. 1999. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. *The Journal of infectious diseases,* 179**,** 859-870.

Giorgi, J. V., Lyles, R. H., Matud, J. L., Yamashita, T. E., Mellors, J. W., et al. 2002. Predictive Value of Immunologic and Virologic Markers After Long or Short Duration of HIV-1 Infection. *Journal of Acquired Immune Deficiency Syndromes,* 29**,** 346-355.

Glass, D., Viñuela, A., Davies, M. N., Ramasamy, A., Parts, L., et al. 2013. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome biology,* 14**,** R75-R75.

Goedecke, J. H., Dave, J. A., Faulenbach, M. V., Utzschneider, K. M., Lambert, E. V., et al. 2009. Insulin response in relation to insulin sensitivity: an appropriate beta-cell response in black South African women. *Diabetes Care,* 32**,** 860-865.

Gonzalez-Quintela, A., Alonso, M., Campos, J., Vizcaino, L., Loidi, L., et al. 2013. Determinants of Serum Concentrations of Lipopolysaccharide-Binding Protein (LBP) in the Adult Population: The Role of Obesity. *PLOS ONE,* 8**,** e54600.

González-Reyes, S., Marín, L., González, L., González, L. O., del Casar, J. M., et al. 2010. Study of TLR3, TLR4 and TLR9 in breast carcinomas and their association with metastasis. *BMC Cancer,* 10**,** 665.

Gornalusse, G. G., Mummidi, S., Gaitan, A. A., Jimenez, F., Ramsuran, V., et al. 2015. Epigenetic mechanisms, T-cell activation, and CCR5 genetics interact to regulate T-cell expression of CCR5, the major HIV-1 coreceptor. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** E4762-4771.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., et al. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *The New England journal of medicine,* 305**,** 1425-1431.

Goyert, S. M., Ferrero, E., Rettig, W. J., Yenamandra, A. K., Obata, F., et al. 1988. The CD14 monocyte differentiation antigen maps to a region encoding growth factors and receptors. *Science,* 239**,** 497-500.

Gray, P., Michelsen, K. S., Sirois, C. M., Lowe, E., Shimada, K., et al. 2010. Identification of a novel human MD-2 splice variant that negatively regulates Lipopolysaccharide-induced TLR4 signaling. *Journal of Immunology,* 184**,** 6359-6366.

Gray, P. W., Corcorran, A. E., Eddy, R. L., Byers, M. G. & Shows, T. B. 1993. The Genes for the Lipopolysaccharide Binding Protein (LBP) and the Bactericidal Permeability Increasing Protein (BPI) Are Encoded in the Same Region of Human Chromosome 20. *Genomics,* 15**,** 188-190.

Groves, K. C., Bibby, D. F., Clark, D. A., Isaksen, A., Deayton, J. R., et al. 2012. Disease Progression in HIV-1-Infected Viremic Controllers. *Journal of acquired immune deficiency syndromes* 61**,** 407-416.

Grube, B. J., Cochane, C. G., Ye, R. D., Green, C. E., McPhail, M. E., et al. 1994. Lipopolysaccharide binding protein expression in primary human hepatocytes and HepG2 hepatoma cells. *The Journal of biological chemistry,* 269**,** 8477-8482.

Grund, B., Baker, J. V., Deeks, S. G., Wolfson, J., Wentworth, D., et al. 2016. Relevance of Interleukin-6 and D-Dimer for Serious Non-AIDS Morbidity and Death among HIV-Positive Adults on Suppressive Antiretroviral Therapy. *PloS one,* 11**,** e0155100-e0155100.

Gu, W., Gurguis, C. I., Zhou, J. J., Zhu, Y., Ko, E. A., et al. 2015. Functional and Structural Consequence of Rare Exonic Single Nucleotide Polymorphisms: One Story, Two Tales. *Genome biology and evolution,* 7**,** 2929-2940.

Gu, W., Shan, Y.-a., Zhou, J., Jiang, D.-p., Zhang, L., et al. 2007. Functional significance of gene polymorphisms in the promoter of myeloid differentiation-2. *Annals of surgery,* 246**,** 151-158.

Gu, W., Xu, Y., Xie, X., Wang, T., Ko, J. H., et al. 2014. The role of RNA structure at 5' untranslated region in microRNA-mediated gene regulation. *RNA,* 20**,** 1369-1375.

Guinane, C. M. & Cotter, P. D. 2013. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therapeutic advances in gastroenterology,* 6**,** 295-308.

Gurdasani, D., Iles, L., Dillon, D. G., Young, E. H., Olson, A. D., et al. 2014. A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS* 28**,** 149-162.

Hamann, L., Kumpf, O., Muller, M., Visintin, A., Eckert, J., et al. 2004. A coding mutation within the first exon of the human MD-2 gene results in decreased lipopolysaccharide-induced signaling. *Genes & Immunity* 5**,** 283-288.

Han, S. H., Kim, J. H., Martin, M., Michalek, S. M. & Nahm, M. H. 2003. Pneumococcal lipoteichoic acid (LTA) is not as potent as staphylococcal LTA in stimulating Toll-like receptor 2. *Infection and immunity,* 71**,** 5541-5548.

Hartigan-O'Connor, D. J., Hirao, L. A., McCune, J. M. & Dandekar, S. 2011. Th17 cells and regulatory T cells in elite control over HIV and SIV. *Current opinion in HIV and AIDS,* 6**,** 221-227.

Hazenberg, M. D., Otto, S. A., van Benthem, B. H. B., Roos, M. T., Coutinho, R. A., et al. 2003. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *AIDS,* 17.

Hazenberg, M. D., Stuart, J. W., Otto, S. A., Borleffs, J. C., Boucher, C. A., et al. 2000. T-cell division in human immunodeficiency virus (HIV)-1 infection is mainly due to immune activation: a longitudinal analysis in patients before and during highly active antiretroviral therapy (HAART). *Blood,* 95**,** 249-255.

Heesen, M., Bloemeke, B., Heussen, N. & Kunz, D. 2002. Can the interleukin-6 response to endotoxin be predicted? Studies of the influence of a promoter polymorphism of the interleukin-6 gene, gender, the density of the endotoxin receptor CD14, and inflammatory cytokines. *Critical care medicine,* 30**,** 664-669.

Hellerstein, M., Hanley, M. B., Cesar, D., Siler, S., Papageorgopoulos, C., et al. 1999. Directly measured kinetics of circulating T lymphocytes in normal and HIV-1-infected humans. *Nature medicine,* 5**,** 83-89.

Henkens, M. T. H. M., Raafs, A. G., Verdonschot, J. A. J., Linschoten, M., van Smeden, M., et al. 2022. Age is the main determinant of COVID-19 related in-hospital mortality with minimal impact of pre-existing comorbidities, a retrospective cohort study. *BMC Geriatrics,* 22**,** 184.

Hertz, D. & Schneider, B. 2019. Sex differences in tuberculosis. *Seminars in Immunopathology,* 41**,** 225-237.

Hishida, A., Matsuo, K., Goto, Y., Mitsuda, Y., Hiraki, A., et al. 2009. Toll-like receptor 4 +3725 G/C polymorphism, Helicobacter pylori seropositivity, and the risk of gastric atrophy and gastric cancer in Japanese. *Helicobacter,* 14**,** 47-53.

Huang, X., Chen, H., Chen, H., Li, W., Li, H., et al. 2012. Precise determination of time to reach viral load set point after acute HIV-1 infection. *Journal of acquired immune deficiency syndromes* 61**,** 448-454.

Hubacek, J. A., Büchler, C., Aslanidis, C. & Schmitz, G. 1997. The Genomic Organization of the Genes for Human Lipopolysaccharide Binding Protein (LBP) and Bactericidal Permeability Increasing Protein (BPI) Is Highly Conserved. *Biochemical and Biophysical Research Communications,* 236**,** 427-430.

Hubacek, J. A., Rothe, G., Pit'ha, J., Skodova, Z., Stanek, V., et al. 1999. C(-260)-->T polymorphism in the promoter of the CD14 monocyte receptor gene as a risk factor for myocardial infarction. *Circulation,* 99**,** 3218-3220.

Hunt, P. W. 2007. Role of immune activation in HIV pathogenesis. *Current HIV/AIDS reports,* 4**,** 42-47.

Hunt, P. W. 2009. Natural Control of HIV-1 Replication and Long-Term Nonprogression: Overlapping but Distinct Phenotypes. *The Journal of Infectious Diseases,* 200**,** 1636-1638.

Hunt, P. W., Brenchley, J., Sinclair, E., McCune, J. M., Roland, M., et al. 2008. Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy. *The Journal of infectious diseases,* 197**,** 126-133.

Hunt, P. W., Cao, H. L., Muzoora, C., Ssewanyana, I., Bennett, J., et al. 2011. Impact of CD8+ T-cell activation on CD4+ T-cell recovery and mortality in HIV-infected Ugandans initiating antiretroviral therapy. *AIDS* 25**,** 2123-2131.

Hunt, P. W., Lee, S. A. & Siedner, M. J. 2016. Immunologic Biomarkers, Morbidity, and Mortality in Treated HIV Infection. *The Journal of infectious diseases,* 214 Suppl 2**,** S44-50.

Hunt, P. W., Martin, J. N., Sinclair, E., Bredt, B., Hagos, E., et al. 2003. T Cell Activation Is Associated with Lower CD4+ T Cell Gains in Human Immunodeficiency Virus-Infected Patients with Sustained Viral Suppression during Antiretroviral Therapy. *The Journal of Infectious Diseases,* 187**,** 1534-1543.

Hunt, P. W., Sinclair, E., Rodriguez, B., Shive, C., Clagett, B., et al. 2014. Gut epithelial barrier dysfunction and innate immune activation predict mortality in treated HIV infection. *The Journal of infectious diseases,* 210**,** 1228-1238.

Husebye, H., Halaas, Ø., Stenmark, H., Tunheim, G., Sandanger, Ø., et al. 2006. Endocytic pathways regulate Toll-like receptor 4 signaling and link innate and adaptive immunity. *The EMBO journal,* 25**,** 683-692.

Isnard, S., Ramendra, R., Dupuy, F. P., Lin, J., Fombuena, B., et al. 2020. Plasma Levels of C-Type Lectin REG3alpha and Gut Damage in People With Human Immunodeficiency Virus. *The Journal of infectious diseases,* 221**,** 110-121.

Jain, V., Hartogensis, W., Bacchetti, P., Hunt, P. W., Hatano, H., et al. 2013. Antiretroviral therapy initiated within 6 months of HIV infection is associated with lower T-cell activation and smaller HIV reservoir size. *The Journal of infectious diseases,* 208**,** 1202-1211.

Jang, J., Kim, W., Kim, K., Chung, S.-I., Shim, Y. J., et al. 2015. Lipoteichoic acid upregulates NF-κB and proinflammatory cytokines by modulating β-catenin in bronchial epithelial cells. *Molecular Medicine Reports* 12**,** 4720-4726.

Jenabian, M. A., El-Far, M., Vyboh, K., Kema, I., Costiniuk, C. T., et al. 2015. Immunosuppressive Tryptophan Catabolism and Gut Mucosal Dysfunction Following Early HIV Infection. *The Journal of infectious diseases,* 212**,** 355-366.

Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K. B., et al. 2020. Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature,* 585**,** 261-267.

Jiang, W., Lederman, M. M., Hunt, P., Sieg, S. F., Haley, K., et al. 2009. Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in persons with antiretroviral-treated HIV infection. *The Journal of infectious diseases,* 199**,** 1177-1185.

Jung, A. C. & Paauw, D. S. 1998. Diagnosing HIV-related disease: using the CD4 count as a guide. *Journal of general internal medicine,* 13**,** 131-136.

Kagan, J. C. & Medzhitov, R. 2006. Phosphoinositide-mediated adaptor recruitment controls Toll-like receptor signaling. *Cell,* 125**,** 943-955.

Kang, J. Y. & Lee, J. O. 2011. Structural biology of the Toll-like receptor family. *Annual review of biochemistry,* 80**,** 917-941.

Kassutto, S. & Rosenberg, E. S. 2004. Primary HIV type 1 infection. *Clinical infectious diseases* 38**,** 1447-1453.

Kawai, T. & Akira, S. 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunology,* 11**,** 373-384.

Kawasaki, T. & Kawai, T. 2014. Toll-Like Receptor Signaling Pathways. *Frontiers in Immunology,* 5.

Kedzierska, K., Azzam, R., Ellery, P., Mak, J., Jaworowski, A., et al. 2003. Defective phagocytosis by human monocyte/macrophages following HIV-1 infection: underlying mechanisms and modulation by adjunctive cytokine therapy. *Journal of clinical virology,* 26**,** 247-263.

Kennedy, M. N., Mullen, G. E., Leifer, C. A., Lee, C., Mazzoni, A., et al. 2004. A complex of soluble MD-2 and lipopolysaccharide serves as an activating ligand for Toll-like receptor 4. *The Journal of biological chemistry,* 279**,** 34698-34704.

Kim, C. J., Kovacs, C., Chun, T.-W., Kandel, G., Osborne, B. J. W., et al. 2014. Antiretroviral therapy in HIV-infected elite controllers: impact on gut immunology, microbial translocation, and biomarkers of serious non-AIDS conditions. *Journal of acquired immune deficiency syndromes (1999),* 67**,** 514-518.

Kim, J. I., Lee, C. J., Jin, M. S., Lee, C. H., Paik, S. G., et al. 2005. Crystal structure of CD14 and its implications for lipopolysaccharide signaling. *The Journal of biological chemistry,* 280**,** 11347-11351.

Kim, K. E., Cho, Y. S., Baek, K. S., Li, L., Baek, K.-H., et al. 2016. Lipopolysaccharide-binding protein plasma levels as a biomarker of obesity-related insulin resistance in adolescents. *Korean Journal of Pediatrics,* 59**,** 231 - 238.

Kitchens, R. L., Thompson, P. A., Viriyakosol, S., O'Keefe, G. E. & Munford, R. S. 2001. Plasma CD14 decreases monocyte responses to LPS by transferring cell-bound LPS to plasma lipoproteins. *The Journal of clinical investigation,* 108**,** 485-493.

Kitchens, R. L., Wolfbauer, G., Albers, J. J. & Munford, R. S. 1999. Plasma lipoproteins promote the release of bacterial lipopolysaccharide from the monocyte cell surface. *Journal of Biological Chemistry,* 274**,** 34116-34122.

Klatt, N. R., Bosinger, S. E., Peck, M., Richert-Spuhler, L. E., Heigele, A., et al. 2014. Limited HIV Infection of Central Memory and Stem Cell Memory CD4+ T Cells Is Associated with Lack of Progression in Viremic Individuals. *PLOS Pathogens,* 10**,** e1004345.

Klatt, N. R., Funderburg, N. T. & Brenchley, J. M. 2013. Microbial translocation, immune activation, and HIV disease. *Trends in microbiology,* 21**,** 6-13.

Klein, S. L. & Flanagan, K. L. 2016. Sex differences in immune responses. *Nature Reviews Immunology,* 16**,** 626-638.

Kondo, T., Ohno, M., Shimokata, K., Iino, S., Inden, Y., et al. 2003. CD14 promoter polymorphism is associated with acute myocardial infarction resulting from insignificant coronary artery stenosis. *Heart* 89**,** 931-932.

Koor, G. W., Paximadis, M., Picton, A. C. P., Karatas, F., Loubser, S. A., et al. 2019. Cis-regulatory genetic variants in the CCR5 gene and natural HIV-1 control in black South Africans. *Clinical immunology,* 205**,** 16-24.

Kresfelder, T. L., Janssen, R., Bont, L., Pretorius, M. & Venter, M. 2011. Confirmation of an association between single nucleotide polymorphisms in the VDR gene with respiratory syncytial virus related disease in South African children. *Journal of medical virology,* 83**,** 1834-1840.

Krishnan, S., Wilson, E. M. P., Sheikh, V., Rupert, A., Mendoza, D., et al. 2014. Evidence for innate immune system activation in HIV type 1-infected elite controllers. *The Journal of infectious diseases,* 209**,** 931-939.

Kulkarni, V. & Ruprecht, R. M. 2017. Mucosal IgA Responses: Damaged in Established HIV Infection-Yet, Effective Weapon against HIV Transmission. *Frontiers in immunology,* 8**,** 1581-1581.

Kuzmich, N. N., Sivak, K. V., Chubarev, V. N., Porozov, Y. B., Savateeva-Lyubimova, T. N., et al. 2017. TLR4 Signaling Pathway Modulators as Potential Therapeutics in Inflammation and Sepsis. *Vaccines (Basel),* 5.

Lai, Y., Li, D., Li, C., Muehleisen, B., Radek, K. A., et al. 2012. The antimicrobial protein REG3A regulates keratinocyte proliferation and differentiation after skin injury. *Immunity,* 37**,** 74-84.

Laine, M. L., Morré, S. A., Murillo, L. S., van Winkelhoff, A.-J. & Peña, A. S. 2005. CD14 and TLR4 Gene Polymorphisms in Adult Periodontitis. *Journal of Dental Research,* 84**,** 1042-1046.

Landmann, R., Knopf, H. P., Link, S., Sansano, S., Schumann, R., et al. 1996. Human monocyte CD14 is upregulated by lipopolysaccharide. *Infection and immunity,* 64**,** 1762-1769.

Lane, H. C., Masur, H., Edgar, L. C., Whalen, G., Rook, A. H., et al. 1983. Abnormalities of B-cell activation and immunoregulation in patients with the acquired immunodeficiency syndrome. *The New England journal of medicine,* 309**,** 453-458.

Lavreys, L., Baeten, J. M., Chohan, V., McClelland, R. S., Hassan, W. M., et al. 2006. Higher Set Point Plasma Viral Load and More-Severe Acute HIV Type 1 (HIV-1) Illness Predict Mortality among High-Risk HIV-1–Infected African Women. *Clinical Infectious Diseases,* 42**,** 1333-1339.

Le, T., Wright, E. J., Smith, D. M., He, W., Catano, G., et al. 2013. Enhanced CD4+ T-cell recovery with earlier HIV-1 antiretroviral therapy. *The New England journal of medicine,* 368**,** 218-230.

Lee, K. M. & Seong, S. Y. 2009. Partial role of TLR4 as a receptor responding to damage-associated molecular pattern. *Immunology Letters* 125**,** 31-39.

Lefèvre, N., Corazza, F., Duchateau, J., Desir, J. & Casimir, G. 2012. Sex differences in inflammatory cytokines and CD99 expression following in vitro lipopolysaccharide stimulation. *Shock,* 38**,** 37-42.

Lenoir, C., Sapin, C., Broquet, A. H., Jouniaux, A. M., Bardin, S., et al. 2008. MD-2 controls bacterial lipopolysaccharide hyporesponsiveness in human intestinal epithelial cells. *Life sciences,* 82**,** 519-528.

Leon, A., Leal, L., Torres, B., Lucero, C., Inciarte, A., et al. 2015. Association of microbial translocation biomarkers with clinical outcome in controllers HIV-infected patients. *AIDS* 29**,** 675-681.

Leon, A., Perez, I., Ruiz-Mateos, E., Benito, J. M., Leal, M., et al. 2016. Rate and predictors of progression in elite and viremic HIV-1 controllers. *AIDS,* 30**,** 1209-1220.

Lester, R. T., Yao, X. D., Ball, T. B., McKinnon, L. R., Kaul, R., et al. 2008. Toll-like receptor expression and responsiveness are increased in viraemic HIV-1 infection. *AIDS,* 22**,** 685-694.

Lester, R. T., Yao, X. D., Ball, T. B., McKinnon, L. R., Omange, W. R., et al. 2009. HIV-1 RNA dysregulates the natural TLR response to subclinical endotoxemia in Kenyan female sex-workers. *PLoS One,* 4**,** e5644.

LeVan, T., Bloom, J., Bailey, T., Karp, C., Halonen, M., et al. 2001. A Common Single Nucleotide Polymorphism in the CD14 Promoter Decreases the Affinity of Sp Protein Binding and Enhances Transcriptional Activity. *Journal of immunology,* 167**,** 5838-5844.

Levan, T. D., Michel, O., Dentener, M., Thorn, J., Vertongen, F., et al. 2008. Association between CD14 polymorphisms and serum soluble CD14 levels: effect of atopy and endotoxin inhalation. *The Journal of allergy and clinical immunology,* 121**,** 434-440.e431.

Liao, E.-C., Chang, C.-Y., Wu, C.-C., Wang, G.-J. & Tsai, J.-J. 2015a. Association of Single Nucleotide Polymorphisms in the MD-2 Gene Promoter Region With Der p 2 Allergy. *Allergy, Asthma & Immunology Research,* 7**,** 249-255.

Liao, E.-C., Hsieh, C.-W., Chang, C.-Y., Yu, S.-J., Sheu, M.-L., et al. 2015b. Enhanced Allergic Inflammation of Der p 2 Affected by Polymorphisms of MD-2 Promoter. *Allergy, Asthma & Immunology Research* 7**,** 497-506.

Lichtfuss, G. F., Hoy, J., Rajasuriar, R., Kramski, M., Crowe, S. M., et al. 2011. Biomarkers of immune dysfunction following combination antiretroviral therapy for HIV infection. *Biomarkers in medicine,* 5**,** 171-186.

Lien, E., Aukrust, P., Sundan, A., Müller, F., Frøland, S. S., et al. 1998. Elevated Levels of Serum-Soluble CD14 in Human Immunodeficiency Virus Type 1 (HIV-1) Infection: Correlation to Disease Progression and Clinical Events. *Blood,* 92**,** 2084-2092.

Lin, J., Yao, Y. M., Yu, Y., Chai, J. K., Huang, Z. H., et al. 2007. Effects of CD14-159 C/T polymorphism on CD14 expression and the balance between proinflammatory and anti-inflammatory cytokines in whole blood culture. *Shock,* 28**,** 148-153.

Liu, C. J., Fu, X., Xia, M., Zhang, Q., Gu, Z., et al. 2021. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic acids research,* 49**,** D1276-d1281.

Liu, J. Z., Pezeshki, M. & Raffatellu, M. 2009. Th17 cytokines and host-pathogen interactions at the mucosa: dichotomies of help and harm. *Cytokine,* 48**,** 156-160.

Liu, L., Zhang, Q., Xiong, X. Y., Gong, Q. W., Liao, M. F., et al. 2018. TLR4 gene polymorphisms rs11536889 is associated with intracranial aneurysm susceptibility. *Journal of clinical neuroscience,* 53**,** 165-170.

Liu, X., Lu, L., Yao, P., Ma, Y., Wang, F., et al. 2014. Lipopolysaccharide binding protein, obesity status and incidence of metabolic syndrome: a prospective study among middle-aged and older Chinese. *Diabetologia,* 57**,** 1834-1841.

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., et al. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene,* 5**,** 1743-1753.

Long, H., O'Connor, B. P., Zemans, R. L., Zhou, X., Yang, I. V., et al. 2014. The Toll-like receptor 4 polymorphism Asp299Gly but not Thr399Ile influences TLR4 signaling and function. *PLoS One,* 9**,** e93550.

Lorenz, E., Mira, J. P., Frees, K. L. & Schwartz, D. A. 2002. Relevance of Mutations in the TLR4 Receptor in Patients With Gram-Negative Septic Shock. *Archives of Internal Medicine,* 162**,** 1028-1032.

Lu, H. X., Sun, J. H., Wen, D. L., Du, J., Zeng, L., et al. 2018. LBP rs2232618 polymorphism contributes to risk of sepsis after trauma. *World journal of emergency surgery,* 13**,** 52.

Lu, Y.-C., Yeh, W.-C. & Ohashi, P. S. 2008. LPS/TLR4 signal transduction pathway. *Cytokine,* 42**,** 145-151.

Macpherson, A. J., Geuking, M. B. & McCoy, K. D. 2012. Homeland security: IgA immunity at the frontiers of the body. *Trends in immunology,* 33**,** 160-167.

Maeshima, N. & Fernandez, R. 2013. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. *Frontiers in Cellular and Infection Microbiology,* 3.

Makamure, M. T., Reddy, P., Chuturgoon, A., Naidoo, R. N., Mentz, G., et al. 2017. Interaction between ambient pollutant exposure, CD14 (-159) polymorphism and respiratory outcomes among children in Kwazulu-Natal, Durban. *Human & experimental toxicology,* 36**,** 238-246.

Manček, M., Pristovšek, P. & Jerala, R. 2002. Identification of LPS-Binding Peptide Fragment of MD-2, a Toll-Receptor Accessory Protein. *Biochemical and Biophysical Research Communications,* 292**,** 880-885.

Marafini, I., Di Sabatino, A., Zorzi, F., Monteleone, I., Sedda, S., et al. 2014. Serum regenerating islet-derived 3-alpha is a biomarker of mucosal enteropathies. *Alimentary pharmacology & therapeutics,* 40**,** 974-981.

Marchetti, G., Bellistri, G. M., Borghi, E., Tincati, C., Ferramosca, S., et al. 2008. Microbial translocation is associated with sustained failure in CD4+ T-cell reconstitution in HIV-infected patients on long-term highly active antiretroviral therapy. *AIDS,* 22**,** 2035-2038.

Marchetti, G., Cozzi-Lepri, A., Merlini, E., Bellistri, G. M., Castagna, A., et al. 2011. Microbial translocation predicts disease progression of HIV-infected antiretroviral-naive patients with high CD4+ cell count. *AIDS,* 25**,** 1385-1394.

Marchetti, G., Tincati, C. & Silvestri, G. 2013. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clinical microbiology reviews,* 26**,** 2-18.

Mathison, J. C. & Ulevitch, R. J. 1979. The clearance, tissue distribution, and cellular localization of intravenously injected lipopolysaccharide in rabbits. *Journal of immunology,* 123**,** 2133-2143.

Mayr, C. 2017. Regulation by 3′-Untranslated Regions. *Annual Review of Genetics,* 51**,** 171-194.

Mayr, C. 2019. What Are 3' UTRs Doing? *Cold Spring Harbor perspectives in biology,* 11.

McMichael, A. J., Borrow, P., Tomaras, G. D., Goonetilleke, N. & Haynes, B. F. 2010. The immune response during acute HIV-1 infection: clues for vaccine development. *Nature reviews. Immunology,* 10**,** 11-23.

Means, T. K., Lien, E., Yoshimura, A., Wang, S., Golenbock, D. T., et al. 1999. The CD14 ligands lipoarabinomannan and lipopolysaccharide differ in their requirement for Toll-like receptors. *Journal of immunology,* 163**,** 6748-6755.

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J. A., et al. 1996. Prognosis in HIV-1 Infection Predicted by the Quantity of Virus in Plasma. *Science,* 272**,** 1167-1170.

Mendoza, D., Johnson, S. A., Peterson, B. A., Natarajan, V., Salgado, M., et al. 2012. Comprehensive analysis of unique cases with extraordinary control over HIV replication. *Blood,* 119**,** 4645-4655.

Meng, L., Song, Z., Liu, A., Dahmen, U., Yang, X., et al. 2021. Effects of Lipopolysaccharide-Binding Protein (LBP) Single Nucleotide Polymorphism (SNP) in Infections, Inflammatory Diseases, Metabolic Disorders and Cancers. *Frontiers in immunology,* 12**,** 681810.

Merrill, J. E., Koyanagi, Y. & Chen, I. S. 1989. Interleukin-1 and tumor necrosis factor alpha can be induced from mononuclear phagocytes by human immunodeficiency virus type 1 binding to the CD4 receptor. *Journal of virology,* 63**,** 4404-4408.

Molodecky, N. A., Soon, I. S., Rabi, D. M., Ghali, W. A., Ferris, M., et al. 2012. Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review. *Gastroenterology,* 142**,** 46-54.e42.

Molteni, M., Gemma, S. & Rossetti, C. 2016. The Role of Toll-Like Receptor 4 in Infectious and Noninfectious Inflammation. *Mediators of inflammation,* 2016**,** 6978936.

Montaner, J. S. G., Lima, V. D., Harrigan, P. R., Lourenço, L., Yip, B., et al. 2014. Expansion of HAART Coverage Is Associated with Sustained Decreases in HIV/AIDS Morbidity, Mortality and HIV Transmission: The "HIV Treatment as Prevention" Experience in a Canadian Setting. *PLOS ONE,* 9**,** e87872.

Moreno-Navarrete, J. M., Ortega, F., Serino, M., Luche, E., Waget, A., et al. 2012. Circulating lipopolysaccharide-binding protein (LBP) as a marker of obesity-related insulin resistance. *International Journal of Obesity,* 36**,** 1442-1449.

Muenchhoff, M., Adland, E., Karimanzira, O., Crowther, C., Pace, M., et al. 2016. Nonprogressing HIV-infected children share fundamental immunological features of nonpathogenic SIV infection. *Science translational medicine,* 8**,** 358ra125-358ra125.

Munthe-Kaas, M. C., Torjussen, T. M., Gervin, K., Lødrup Carlsen, K. C., Carlsen, K. H., et al. 2010. CD14 polymorphisms and serum CD14 levels through childhood: A role for gene methylation? *Journal of Allergy and Clinical Immunology,* 125**,** 1361-1368.

Muzio, M., Bosisio, D., Polentarutti, N., D'amico, G., Stoppacciaro, A., et al. 2000. Differential Expression and Regulation of Toll-Like Receptors (TLR) in Human Leukocytes: Selective Expression of TLR3 in Dendritic Cells. *The Journal of Immunology,* 164**,** 5998-6004.

Nabih, E. S., Kamel, H. F. & Kamel, T. B. 2016. Association Between CD14 Polymorphism (-1145G/A) and Childhood Bronchial Asthma. *Biochemical genetics,* 54**,** 50-60.

Naeger, D. M., Martin, J. N., Sinclair, E., Hunt, P. W., Bangsberg, D. R., et al. 2010. Cytomegalovirus-specific T cells persist at very high levels during long-term antiretroviral treatment of HIV disease. *PLoS One,* 5**,** e8886.

Nagai, Y., Akashi, S., Nagafuku, M., Ogata, M., Iwakura, Y., et al. 2002. Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nature immunology,* 3**,** 667-672.

Nair, V., Sankaranarayanan, R. & Vasavada, A. R. 2021. Deciphering the association of intronic single nucleotide polymorphisms of crystallin gene family with congenital cataract. *Indian Journal of Ophthalmology,* 69**,** 2064-2070.

Navarrete-Muñoz, M. A., Restrepo, C., Benito, J. M. & Rallón, N. 2020. Elite controllers: A heterogeneous group of HIV-infected patients. *Virulence,* 11**,** 889-897.

Nazli, A., Chan, O., Dobson-Belaire, W. N., Ouellet, M., Tremblay, M. J., et al. 2010. Exposure to HIV-1 directly impairs mucosal epithelial barrier integrity allowing microbial translocation. *PLoS pathogens,* 6**,** e1000852.

Negi, N., Singh, R., Sharma, A., Das, B. K. & Vajpayee, M. 2017. Comparative evaluation of microbial translocation products (LPS, sCD14, IgM Endocab) in HIV-1 infected Indian individuals. *Microbial pathogenesis,* 111**,** 331-337.

Newport, M. J., Allen, A., Awomoyi, A. A., Dunstan, S. J., McKinney, E., et al. 2004. The toll-like receptor 4 Asp299Gly variant: no influence on LPS responsiveness or susceptibility to pulmonary tuberculosis in The Gambia. *Tuberculosis,* 84**,** 347-352.

Nissen, S. K., Christiansen, M., Helleberg, M., Kjær, K., Jørgensen, S. E., et al. 2018. Whole Exome Sequencing of HIV-1 long-term non-progressors identifies rare variants in genes encoding innate immune sensors and signaling molecules. *Scientific Reports,* 8**,** 15253.

Nockher, W. A., Bergmann, L. & Scherberich, J. E. 1994. Increased soluble CD14 serum levels and altered CD14 expression of peripheral blood monocytes in HIV-infected patients. *Clinical and experimental immunology,* 98**,** 369-374.

Noel, N., Lerolle, N., Lecuroux, C., Goujard, C., Venet, A., et al. 2015. Immunologic and Virologic Progression in HIV Controllers: The Role of Viral "Blips" and Immune Activation in the ANRS CO21 CODEX Study. *PLoS One,* 10**,** e0131922.

Nomura, F., Akashi, S., Sakao, Y., Sato, S., Kawai, T., et al. 2000. Cutting edge: endotoxin tolerance in mouse peritoneal macrophages correlates with down-regulation of surface toll-like receptor 4 expression. *Journal of immunology,* 164**,** 3476-3479.

Noreen, M., Shah, M. A., Mall, S. M., Choudhary, S., Hussain, T., et al. 2012. TLR4 polymorphisms and disease susceptibility. *Inflammation research* 61**,** 177-188.

Nowroozalizadeh, S., Mansson, F., da Silva, Z., Repits, J., Dabo, B., et al. 2010. Microbial translocation correlates with the severity of both HIV-1 and HIV-2 infections. *The Journal of infectious diseases,* 201**,** 1150-1154.

Nyamweya, S., Hegedus, A., Jaye, A., Rowland-Jones, S., Flanagan, K. L., et al. 2013. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Reviews in Medical Virology,* 23**,** 221-240.

O'Brien, S. J. & Hendrickson, S. L. 2013. Host genomic influences on HIV/AIDS. *Genome biology,* 14**,** 201-201.

Offner, P. J., Moore, E. E. & Biffl, W. L. 1999. Male gender is a risk factor for major infections after surgery. *Archives of surgery,* 134**,** 935-938; discussion 938-940.

Ogden, C. L., Carroll, M. D., Fryar, C. D. & Flegal, K. M. 2015. Prevalence of obesity among adults and youth: United States, 2011-2014. *National Center for Health Statistics data brief***,** 1-8.

Ohnishi, T., Muroi, M. & Tanamoto, K. 2003. MD-2 is necessary for the toll-like receptor 4 protein to undergo glycosylation essential for its translocation to the cell surface. *Clin Diagn Lab Immunol,* 10**,** 405-410.

Ohta, S., Bahrun, U., Tanaka, M. & Kimoto, M. 2004. Identification of a novel isoform of MD-2 that downregulates lipopolysaccharide signaling. *Biochemical and biophysical research communications,* 323**,** 1103-1108.

Ohto, U., Yamakawa, N., Akashi-Takamura, S., Miyake, K. & Shimizu, T. 2012. Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *The Journal of biological chemistry,* 287**,** 40611-40617.

Okulicz, J. F., Grandits, G. A., Weintrob, A. C., Landrum, M. L., Ganesan, A., et al. 2010. CD4 T cell count reconstitution in HIV controllers after highly active antiretroviral therapy. *Clinical infectious diseases,* 50**,** 1187-1191.

Okulicz, J. F. & Lambotte, O. 2011. Epidemiology and clinical characteristics of elite controllers. *Current opinion in HIV and AIDS,* 6**,** 163-168.

Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., et al. 2020. The impact of sex on gene expression across human tissues. *Science* 369**,** eaba3066.

Olson, N. C., Koh, I., Reiner, A. P., Judd, S. E., Irvin, M. R., et al. 2020. Soluble CD14, Ischemic Stroke, and Coronary Heart Disease Risk in a Prospective Study: The REGARDS Cohort. *Journal of the American Heart Association,* 9**,** e014241-e014241.

Opal, S. M., Scannon, P. J., Vincent, J. L., White, M., Carroll, S. F., et al. 1999. Relationship between plasma levels of lipopolysaccharide (LPS) and LPS-binding protein in patients with severe sepsis and septic shock. *The Journal of infectious diseases,* 180**,** 1584-1589.

Özdemir, B. C. & Dotto, G.-P. 2017. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? *Trends in cancer,* 3**,** 181-197.

Palesch, D., Bosinger, S. E., Tharp, G. K., Vanderford, T. H., Paiardini, M., et al. 2018. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature,* 553**,** 77-81.

Pantaleo, G. & Fauci, A. S. 1996. Immunopathogenesis of HIV infection. *Annual review of microbiology,* 50**,** 825-854.

Papadopoulos, A. I., Ferwerda, B., Antoniadou, A., Sakka, V., Galani, L., et al. 2010. Association of Toll-Like Receptor 4 Asp299Gly and Thr399Ile Polymorphisms with Increased Infection Risk in Patients with Advanced HIV-1 Infection. *Clinical Infectious Diseases,* 51**,** 242-247.

Papasavvas, E., Azzoni, L., Foulkes, A., Violari, A., Cotton, M. F., et al. 2011. Increased microbial translocation in ≤ 180 days old perinatally human immunodeficiency virus-positive infants as compared with human immunodeficiency virus-exposed uninfected infants of similar age. *The Pediatric infectious disease journal,* 30**,** 877-882.

Park, B. S., Song, D. H., Kim, H. M., Choi, B. S., Lee, H., et al. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature,* 458**,** 1191-1195.

Paximadis, M., Ngqobe, R. N., Chaisson, R. E., Martinson, N. A. & Tiemessen, C. T. 2017. RICH2 is implicated in viraemic control of HIV-1 in black South African individuals. *Infection, genetics and evolution,* 49**,** 78-87.

Paximadis, M., Picton, A. C. P., Sengupta, D., Ramsay, M., Puren, A., et al. 2021. Interleukin-8 genetic diversity, haplotype structure and production differ in two ethnically distinct South African populations. *Cytokine,* 143**,** 155489.

Peer, N., Balakrishna, Y., de Villiers, A. & Naidoo, P. 2018. Differential associations of cardio-metabolic diseases by population group, gender and adiposity in South Africa. *PLOS ONE,* 13**,** e0202899.

Pereyra, F., Lo, J., Triant, V. A., Wei, J., Buzon, M. J., et al. 2012. Increased coronary atherosclerosis and immune activation in HIV-1 elite controllers. *AIDS,* 26**,** 2409-2412.

Perez-Vilar, J. & Hill, R. L. 1999. The structure and assembly of secreted mucins. *The Journal of biological chemistry,* 274**,** 31751-31754.

Pernas, M., Tarancón-Diez, L., Rodríguez-Gallego, E., Gómez, J., Prado, J. G., et al. 2018. Factors Leading to the Loss of Natural Elite Control of HIV-1 Infection. *Journal of Virology,* 92**,** e01805-01817.

Peters, M. J., Joehanes, R., Pilling, L. C., Schurmann, C., Conneely, K. N., et al. 2015. The transcriptional landscape of age in human peripheral blood. *Nature communications,* 6**,** 8570-8570.

Picton, A. C. P., Paximadis, M., Chaisson, R. E., Martinson, N. A. & Tiemessen, C. T. 2017. CXCR6 gene characterization in two ethnically distinct South African populations and association with viraemic disease control in HIV-1-infected black South African individuals. *Clinical immunology,* 180**,** 69-79.

Picton, A. C. P., Paximadis, M., Koor, G. W., Bharuthram, A., Shalekoff, S., et al. 2021. Reduced CCR5 Expression and Immune Quiescence in Black South African HIV-1 Controllers. *Frontiers in immunology,* 12**,** 781263.

Pilakka-Kanthikeel, S., Kris, A., Selvaraj, A., Swaminathan, S. & Pahwa, S. 2014. Immune activation is associated with increased gut microbial translocation in treatment-naive, HIV-infected children in a resource-limited setting. *Journal of acquired immune deficiency syndromes* 66**,** 16-24.

Pine, S. O., McElrath, M. J. & Bochud, P.-Y. 2009. Polymorphisms in toll-like receptor 4 and toll-like receptor 9 influence viral load in a seroincident cohort of HIV-1-infected individuals. *AIDS* 23**,** 2387-2395.

Pugin, J., Heumann, I. D., Tomasz, A., Kravchenko, V. V., Akamatsu, Y., et al. 1994. CD14 is a pattern recognition receptor. *Immunity,* 1**,** 509-516.

Pugin, J., Schürer-Maly, C. C., Leturcq, D., Moriarty, A., Ulevitch, R. J., et al. 1993. Lipopolysaccharide activation of human endothelial and epithelial cells is mediated by lipopolysaccharide-binding protein and soluble CD14. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 2744-2748.

Pugin, J., Stern-Voeffray, S., Daubeuf, B., Matthay, M. A., Elson, G., et al. 2004. Soluble MD-2 activity in plasma from patients with severe sepsis and septic shock. *Blood,* 104**,** 4071-4079.

Pussinen, P. J., Havulinna, A. S., Lehto, M., Sundvall, J. & Salomaa, V. 2011. Endotoxemia is associated with an increased risk of incident diabetes. *Diabetes care,* 34**,** 392-397.

Raetz, C. R. H. & Whitfield, C. 2002. Lipopolysaccharide endotoxins. *Annual review of biochemistry,* 71**,** 635-700.

Rallabhandi, P., Bell, J., Boukhvalova, M. S., Medvedev, A., Lorenz, E., et al. 2006. Analysis of TLR4 polymorphic variants: new insights into TLR4/MD-2/CD14 stoichiometry, structure, and signaling. *Journal of immunology,* 177**,** 322-332.

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., et al. 2016. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global health, epidemiology and genomics,* 1**,** e20.

Redd, A. D., Dabitao, D., Bream, J. H., Charvat, B., Laeyendecker, O., et al. 2009. Microbial translocation, the innate cytokine response, and HIV-1 disease progression in Africa. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 6718-6723.

Reeves, R. K., Gillis, J., Wong, F. E., Yu, Y., Connole, M., et al. 2010. CD16- natural killer cells: enrichment in mucosal and secondary lymphoid tissues and altered function during chronic SIV infection. *Blood,* 115**,** 4439-4446.

Reiner, A. P., Lange, E. M., Jenny, N. S., Chaves, P. H. M., Ellis, J., et al. 2013. Soluble CD14: genomewide association analysis and relationship to cardiovascular risk and mortality in older adults. *Arteriosclerosis, thrombosis, and vascular biology,* 33**,** 158-164.

Resman, N., Vasl, J., Oblak, A., Pristovsek, P., Gioannini, T. L., et al. 2009. Essential roles of hydrophobic residues in both MD-2 and toll-like receptor 4 in activation by endotoxin. *The Journal of biological chemistry,* 284**,** 15052-15060.

Reus, S., Portilla, J., Sanchez-Paya, J., Giner, L., Frances, R., et al. 2013. Low-level HIV viremia is associated with microbial translocation and inflammation. *Journal of acquired immune deficiency syndromes,* 62**,** 129-134.

Rigau, M., Juan, D., Valencia, A. & Rico, D. 2019. Intronic CNVs and gene expression variation in human populations. *PLOS Genetics,* 15**,** e1007902.

Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. 1998. A family of human receptors structurally related to Drosophila Toll. *Proceedings of the National Academy of Sciences,* 95**,** 588-593.

Romero-Sanchez, M., Gonzalez-Serna, A., Pacheco, Y. M., Ferrando-Martinez, S., Machmach, K., et al. 2012. Different biological significance of sCD14 and LPS in HIV-infection: importance of the immunovirology stage and association with HIV-disease progression markers. *The Journal of infection,* 65**,** 431-438.

Rosas-Taraco, A. G., Revol, A., Salinas-Carmona, M. C., Rendon, A., Caballero-Olin, G., et al. 2007. CD14 C(-159)T polymorphism is a risk factor for development of pulmonary tuberculosis. *The Journal of infectious diseases,* 196**,** 1698-1706.

Rose, A. B. 2019. Introns as Gene Regulators: A Brick on the Accelerator. *Frontiers in Genetics,* 9.

Rotger, M., Dalmau, J., Rauch, A., McLaren, P., Bosinger, S. E., et al. 2011. Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *The Journal of clinical investigation,* 121**,** 2391-2400.

Rudorf, D. C. & Krikorian, S. A. 2005. Adverse Effects Associated With Antiretroviral Therapy and Potential Management Strategies. *Journal of Pharmacy Practice,* 18**,** 258-277.

Sandanger, Ø., Ryan, L., Bohnhorst, J., Iversen, A. C., Husebye, H., et al. 2009. IL-10 enhances MD-2 and CD14 expression in monocytes and the proteins are increased and correlated in HIV-infected patients. *Journal of immunology,* 182**,** 588-595.

Sandler, N. G. & Douek, D. C. 2012. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. *Nature reviews. Microbiology,* 10**,** 655-666.

Sandler, N. G., Wand, H., Roque, A., Law, M., Nason, M. C., et al. 2011. Plasma levels of soluble CD14 independently predict mortality in HIV infection. *The Journal of Infectious Diseases,* 203**,** 780-790.

Santiago, M. L., Range, F., Keele, B. F., Li, Y., Bailes, E., et al. 2005. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *Journal of virology,* 79**,** 12515-12527.

Sato, K., Yoshimura, A., Kaneko, T., Ukai, T., Ozaki, Y., et al. 2012. A single nucleotide polymorphism in 3'-untranslated region contributes to the regulation of Toll-like receptor 4 translation. *The Journal of biological chemistry,* 287**,** 25163-25172.

Schacker, T., Collier, A. C., Hughes, J., Shea, T. & Corey, L. 1996. Clinical and epidemiologic features of primary HIV infection. *Annals of internal medicine,* 125**,** 257-264.

Schröder, J., Kahlke, V., Staubach, K.-H., Zabel, P. & Stüber, F. 1998. Gender differences in human sepsis. *Archives of surgery,* 133**,** 1200-1205.

Schroder, N. W., Morath, S., Alexander, C., Hamann, L., Hartung, T., et al. 2003. Lipoteichoic acid (LTA) of Streptococcus pneumoniae and Staphylococcus aureus activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide-binding protein (LBP), and CD14, whereas TLR-4 and MD-2 are not involved. *The Journal of biological chemistry,* 278**,** 15587-15594.

Schumann, R., Leong, Flaggs, G., Gray, P., Wright, S., et al. 1990. Structure and function of lipopolysaccharide binding protein. *Science,* 249**,** 1429-1431.

Schumann, R. R., Kirschning, C. J., Unbehaun, A., Aberle, H. P., Knope, H. P., et al. 1996. The lipopolysaccharide-binding protein is a secretory class 1 acute-phase protein whose gene is transcriptionally activated by APRF/STAT/3 and other cytokine-inducible nuclear proteins. *Molecular and cellular biology,* 16**,** 3490-3503.

Schumann, R. R. & Zweigner, J. 1999. A Novel Acute-Phase Marker: Lipopolysaccharide Binding Protein (LBP). *Clinical chemistry and laboratory medicine,* 37**,** 271-274.

Scully, E. P. 2018. Sex Differences in HIV Infection. *Current HIV/AIDS reports,* 15**,** 136-146.

Sedaghat, A. R., Rastegar, D. A., O'Connell, K. A., Dinoso, J. B., Wilke, C. O., et al. 2009. T cell dynamics and the response to HAART in a cohort of HIV-1-infected elite suppressors. *Clinical infectious diseases,* 49**,** 1763-1766.

Sereti, I., Krebs, S. J., Phanuphak, N., Fletcher, J. L., Slike, B., et al. 2017. Persistent, Albeit Reduced, Chronic Inflammation in Persons Starting Antiretroviral Therapy in Acute HIV Infection. *Clinical infectious diseases,* 64**,** 124-131.

Serrano-Villar, S., Perez-Elias, M. J., Dronda, F., Casado, J. L., Moreno, A., et al. 2014. Increased risk of serious non-AIDS-related events in HIV-infected subjects on antiretroviral therapy associated with a low CD4/CD8 ratio. *PLoS One,* 9**,** e85798.

Shimada, K., Watanabe, Y., Mokuno, H., Iwama, Y., Daida, H., et al. 2000. Common polymorphism in the promoter of the CD14 monocyte receptor gene is associated with acute myocardial infarction in Japanese men. *The American journal of cardiology,* 86**,** 682-684, a688.

Shimazu, R., Akashi, S., Ogata, H., Nagai, Y., Fukudome, K., et al. 1999. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *Journal of Experimental Medicine,* 189**,** 1777-1782.

Shin, J. H. & Seeley, R. J. 2019. Reg3 Proteins as Gut Hormones? *Endocrinology,* 160**,** 1506-1514.

Shive, C. L., Jiang, W., Anthony, D. D. & Lederman, M. M. 2015. Soluble CD14 is a nonspecific marker of monocyte activation. *AIDS* 29**,** 1263-1265.

Silvestri, G., Sodora, D. L., Koup, R. A., Paiardini, M., O'Neil, S. P., et al. 2003. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. *Immunity,* 18**,** 441-452.

Simiantonaki, N., Kurzik-Dumke, U., Karyofylli, G., Jayasinghe, C., Michel-Schmidt, R., et al. 2007. Reduced expression of TLR4 is associated with the metastatic status of human colorectal cancer. *International Journal of Molecular Medicine,* 20**,** 21-29.

Smith, P. D., Ochsenbauer-Jambor, C. & Smythies, L. E. 2005. Intestinal macrophages: unique effector cells of the innate immune system. *Immunological Reviews,* 206**,** 149-159.

Smythies, L. E., Sellers, M., Clements, R. H., Mosteller-Barnum, M., Meng, G., et al. 2005. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *The Journal of clinical investigation,* 115**,** 66-75.

Sorek, R. & Ast, G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome research,* 13**,** 1631-1637.

Stanislawski, M. A., Lange, L. A., Raffield, L. M., Zakai, N. A., Meyer, M., et al. 2021. Soluble CD14 Levels in the Jackson Heart Study: Associations With Cardiovascular Disease Risk and Genetic Variants. *Arteriosclerosis, thrombosis, and vascular biology,* 41**,** e369-e378.

Stehle, J. R., Jr, Leng, X., Kitzman, D. W., Nicklas, B. J., Kritchevsky, S. B., et al. 2012. Lipopolysaccharide-Binding Protein, a Surrogate Marker of Microbial Translocation, Is Associated With Physical Function in Healthy Older Adults. *The Journals of Gerontology: Series A,* 67**,** 1212-1218.

Stepanikova, I., Bateman, L. B. & Oates, G. R. 2017. Systemic Inflammation in Midlife: Race, Socioeconomic Status, and Perceived Discrimination. *American Journal of Preventive Medicine,* 52**,** S63-S76.

Strutz, F., Heller, G., Krasemann, K., Krone, B. & Muller, G. A. 1999. Relationship of antibodies to endotoxin core to mortality in medical patients with sepsis syndrome. *Intensive Care Med,* 25**,** 435-444.

Su, G. L., Dorko, K., Strom, S. C., Nüssler, A. K. & Wang, S. C. 1999. CD14 expression and production by human hepatocytes. *Journal of Hepatology,* 31**,** 435-442.

Swingler, S., Mann, A., Jacqué, J. M., Brichacek, B., Sasseville, V. G., et al. 1999. HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nature Medicine,* 5**,** 997-1003.

Sylwester, A. W., Mitchell, B. L., Edgar, J. B., Taormina, C., Pelte, C., et al. 2005. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *The Journal of experimental medicine,* 202**,** 673-685.

Takeuchi, F., Yanai, K., Inomata, H., Kuzuya, N., Kajio, H., et al. 2007. Search of type 2 diabetes susceptibility gene on chromosome 20q. *Biochemical and biophysical research communications,* 357**,** 1100-1106.

Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. & Hammer, S. M. 2008. The challenge of HIV-1 subtype diversity. *The New England journal of medicine,* 358**,** 1590-1602.

Thevaranjan, N., Puchta, A., Schulz, C., Naidoo, A., Szamosi, J. C., et al. 2017. Age-Associated Microbial Dysbiosis Promotes Intestinal Permeability, Systemic Inflammation, and Macrophage Dysfunction. *Cell host & microbe,* 21**,** 455-466.e454.

Thompson, P. A., Tobias, P. S., Viriyakosol, S., Kirkland, T. N. & Kitchens, R. L. 2003. Lipopolysaccharide (LPS)-binding protein inhibits responses to cell-bound LPS. *The Journal of biological chemistry,* 278**,** 28367-28371.

Tilves, C. M., Zmuda, J. M., Kuipers, A. L., Nestlerode, C. S., Evans, R. W., et al. 2015. Association of Lipopolysaccharide-Binding Protein With Aging-Related Adiposity Change and Prediabetes Among African Ancestry Men. *Diabetes Care,* 39**,** 385-391.

Tobias, P. S., Soldau, K. & Ulevitch, R. J. 1989. Identification of a lipid A binding site in the acute phase reactant lipopolysaccharide binding protein. *The Journal of biological chemistry,* 264**,** 10867-10871.

Trøseid, M., Lind, A., Nowak, P., Barqasho, B., Heger, B., et al. 2013. Circulating levels of HMGB1 are correlated strongly with MD2 in HIV-infection: possible implication for TLR4-signalling and chronic immune activation. *Innate immunity,* 19**,** 290-297.

Tucci, S. & Akey, J. M. 2019. The long walk to African genomics. *Genome Biology,* 20**,** 130.

Ulevitch, R. J. & Tobias, P. S. 1995. Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. *Annual review of immunology,* 13**,** 437-457.

UNAIDS. (2022) Global HIV & AIDS statistics- Factsheet. Available at: https://www.unaids.org/en/resources/fact-sheet (Accessed: 10 February 2023).

Unkelbach, K., Gardemann, A., Kostrzewa, M., Philipp, M., Tillmanns, H., et al. 1999. A new promoter polymorphism in the gene of lipopolysaccharide receptor CD14 is associated with expired myocardial infarction in patients with low atherosclerotic risk profile. *Arteriosclerosis, thrombosis, and vascular biology,* 19**,** 932-938.

Valdez, H., Connick, E., Smith, K. Y., Lederman, M. M., Bosch, R. J., et al. 2002. Limited immune restoration after 3 years' suppression of HIV-1 replication in patients with moderately advanced disease. *AIDS* 16**,** 1859-1866.

Valencia, P., Dias, A. P. & Reed, R. 2008. Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 3386-3391.

Vasl, J., Prohinar, P., Gioannini, T. L., Weiss, J. P. & Jerala, R. 2008. Functional activity of MD-2 polymorphic variant is significantly different in soluble and TLR4-bound forms: decreased endotoxin binding by G56R MD-2 and its rescue by TLR4 ectodomain. *Journal of immunology,* 180**,** 6107-6115.

Vaure, C. & Liu, Y. 2014. A Comparative Review of Toll-Like Receptor 4 Expression and Functionality in Different Animal Species. *Frontiers in Immunology,* 5.

Vesterbacka, J., Rivera, J., Noyan, K., Parera, M., Neogi, U., et al. 2017. Richer gut microbiota with distinct metabolic profile in HIV infected Elite Controllers. *Scientific Reports,* 7**,** 6269.

Vesy, C. J., Kitchens, R. L., Wolfbauer, G., Albers, J. J. & Munford, R. S. 2000. Lipopolysaccharide-binding protein and phospholipid transfer protein release lipopolysaccharides from gram-negative bacterial membranes. *Infection and immunity,* 68**,** 2410-2417.

Vidyant, S., Chatterjee, A. & Dhole, T. N. 2019. A single-nucleotide polymorphism in TLR4 is linked with the risk of HIV-1 infection. *British journal of biomedical science,* 76**,** 59-63.

Vihinen, M. 2022. When a Synonymous Variant Is Nonsynonymous. *Genes* 13.

Viladés, C., Escoté, X., López-Dupla, M., Martinez, E., Domingo, P., et al. 2014. Involvement of the LPS-LPB-CD14-MD2-TLR4 inflammation pathway in HIV-1/HAART-associated lipodystrophy syndrome (HALS). *The Journal of antimicrobial chemotherapy,* 69**,** 1653-1659.

Vinhaes, C. L., Araujo-Pereira, M., Tibúrcio, R., Cubillos-Angulo, J. M., Demitto, F. O., et al. 2021. Systemic Inflammation Associated with Immune Reconstitution Inflammatory Syndrome in Persons Living with HIV. *Life,* 11**,** 65.

Viriyakosol, S. & Kirkland, T. N. 1995. A region of human CD14 required for lipopolysaccharide binding. *The Journal of biological chemistry,* 270**,** 361-368.

Visintin, A., Latz, E., Monks, B. G., Espevik, T. & Golenbock, D. T. 2003. Lysines 128 and 132 enable lipopolysaccharide binding to MD-2, leading to Toll-like receptor-4 aggregation and signal transduction. *The Journal of biological chemistry,* 278**,** 48313-48320.

Visintin, A., Mazzoni, A., Spitzer, J. A. & Segal, D. M. 2001. Secreted MD-2 is a large polymeric protein that efficiently confers lipopolysaccharide sensitivity to Toll-like receptor 4. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 12156-12161.

Visseaux, B., Damond, F., Matheron, S., Descamps, D. & Charpentier, C. 2016. Hiv-2 molecular epidemiology. *Infection, Genetics and Evolution,* 46**,** 233-240.

Wang, H., Wei, Y., Zeng, Y., Qin, Y., Xiong, B., et al. 2014. The association of polymorphisms of TLR4 and CD14 genes with susceptibility to sepsis in a Chinese population. *BMC medical genetics,* 15**,** 123-123.

Wang, L., Quan, Y., Zhu, Y., Xie, X., Wang, Z., et al. 2022. The regenerating protein 3A: a crucial molecular with dual roles in cancer. *Molecular Biology Reports,* 49**,** 1491-1500.

Weber, J. 2001. The pathogenesis of HIV-1 infection. *British Medical Bulletin,* 58**,** 61-72.

Weiss, J. 2003. Bactericidal/permeability-increasing protein (BPI) and lipopolysaccharide-binding protein (LBP): structure, function and regulation in host defence against Gram-negative bacteria. *Biochemical Society transactions,* 31**,** 785-790.

WHO (2022) HIV. Available at: https://www.who.int/data/gho/data/themes/hiv-aids (Accessed: 10 February 2023).

Widera, D., Mart, #237, nez Aguilar, R., #237, et al. 2019. Toll-like receptor 4 and protease-activated receptor 2 in physiology and pathophysiology of the nervous system: more than just receptor cooperation? *Neural Regeneration Research,* 14**,** 1196-1201.

Willie, B., Hall, N. B., Stein, C. M., Jurevic, R. J., Weinberg, A., et al. 2014. Association of Toll-like receptor polymorphisms with HIV status in North Americans. *Genes and immunity,* 15**,** 569-577.

Wilms, E., Troost, F. J., Elizalde, M., Winkens, B., de Vos, P., et al. 2020. Intestinal barrier function is maintained with aging – a comprehensive study in healthy subjects and irritable bowel syndrome patients. *Scientific Reports,* 10**,** 475.

Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. 1990. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science,* 249**,** 1431-1433.

Wu, Q., Xu, X., Ren, J., Liu, S., Liao, X., et al. 2017. Association between the -159C/T polymorphism in the promoter region of the CD14 gene and sepsis: a meta-analysis. *BMC anesthesiology,* 17**,** 11.

Xu, H. & Ghishan, F. K. 2018. Chapter 10 - Molecular Physiology of Gastrointestinal Function During Development. *In:* SAID, H. M. (ed.) *Physiology of the Gastrointestinal Tract (Sixth Edition).* Academic Press.

Xu, J.-J., Liu, K.-Q., Ying, Z.-M., Zhu, X.-W., Xu, X.-J., et al. 2019. Effect of CD14 polymorphisms on the risk of cardiovascular disease: evidence from a meta-analysis. *Lipids in Health and Disease,* 18**,** 74.

Xue, Y., Zhao, Z. Q., Hong, D., Zhao, M. Y., Zhang, Y. X., et al. 2010. Lack of association between MD-2 promoter gene variants and tuberculosis. *Genetics and molecular research,* 9**,** 1584-1590.

Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., et al. 2015. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Scientific reports,* 5**,** 15145-15145.

Yang, O. O., Cumberland, W. G., Escobar, R., Liao, D. & Chew, K. W. 2017. Demographics and natural history of HIV-1-infected spontaneous controllers of viremia. *AIDS,* 31**,** 1091-1098.

Yong, Y. K., Shankar, E. M., Solomon, A., Spelman, T., Fairley, C. K., et al. 2016. Polymorphisms in the CD14 and TLR4 genes independently predict CD4+ T-cell recovery in HIV-infected individuals on antiretroviral therapy. *AIDS,* 30**,** 2159-2168.

Yongtang, W., Jianxin, J., Qing, L., Zhaoxia, D., Wei, G., et al. 2009. TLR4 −2242 T→C variant increases transcriptional activity of its promoter. *Journal of Medical Colleges of PLA,* 24**,** 69-75.

Younas, M., Psomas, C., Reynes, C., Cezar, R., Kundura, L., et al. 2019. Microbial Translocation Is Linked to a Specific Immune Activation Profile in HIV-1-Infected Adults With Suppressed Viremia. *Frontiers in immunology,* 10**,** 2185.

Younas, M., Psomas, C., Reynes, J. & Corbeau, P. 2016. Immune activation in the course of HIV-1 infection: Causes, phenotypes and persistence under therapy. *HIV Medicine,* 17**,** 89-105.

Yu, L., Wang, L. & Chen, S. 2010. Endogenous toll-like receptor ligands and their biological significance. *Journal of cellular and molecular medicine,* 14**,** 2592-2603.

Zeng, L., Gu, W., Zhang, A. Q., Zhang, M., Zhang, L. Y., et al. 2012a. A functional variant of lipopolysaccharide binding protein predisposes to sepsis and organ dysfunction in patients with major trauma. *Annals of surgery,* 255**,** 147-157.

Zeng, L., Zhang, A.-q., Gu, W., Zhou, J., Zhang, L.-y., et al. 2012b. Identification of Haplotype Tag SNPs Within the Whole Myeloid Differentiation 2 Gene and Their Clinical Relevance in Patients With Major Trauma. *Shock,* 37**,** 366-372.

Zeng, Z., Surewaard, B. G., Wong, C. H., Geoghegan, J. A., Jenne, C. N., et al. 2016. CRIg Functions as a Macrophage Pattern Recognition Receptor to Directly Bind and Capture Blood-Borne Gram-Positive Bacteria. *Cell Host Microbe,* 20**,** 99-106.

Zhang, P., Nguyen, J., Abdulla, F., Nelson, A. T., Beckman, J. D., et al. 2021. Soluble MD-2 and Heme in Sickle Cell Disease Plasma Promote Pro-Inflammatory Signaling in Endothelial Cells. *Frontiers in Immunology,* 12.

Zheng, S. L., Augustsson-Bälter, K., Chang, B., Hedelin, M., Li, L., et al. 2004. Sequence variants of toll-like receptor 4 are associated with prostate cancer risk: results from the CAncer Prostate in Sweden Study. *Cancer research,* 64**,** 2918-2922.

Zheng, Y., Valdez, P. A., Danilenko, D. M., Hu, Y., Sa, S. M., et al. 2008. Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Medicine,* 14**,** 282-289.

Zhou, L., Wei, B., Xing, C., Xie, H., Yu, X., et al. 2011. Polymorphism in 3'-untranslated region of toll-like receptor 4 gene is associated with protection from hepatitis B virus recurrence after liver transplantation. *Transplant infectious disease,* 13**,** 250-258.

Zweigner, J., Gramm, H.-J., Singer, O. C., Wegscheider, K. & Schumann, R. R. 2001. High concentrations of lipopolysaccharide-binding protein in serum of patients with severe

sepsis or septic shock inhibit the lipopolysaccharide response in human monocytes. *Blood*, 98**,** 3800-3808.

APPENDIX 1

Supplementary Tables for Chapter 4

Supplementary Table 1: Minor allele frequencies of *LBP* genetic variants (introns) in the South African population compared to select populations from the 1000 Genomes Project.

Location on gene	Position (chr:20)	Accession number	Base change (wt/mut)	n(minnor allele frequency, %)					
				SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
	38346681	rs142875263	G/T	1 (0.6)	0	0	0	7 (0.7)	0
	38347074	rs5741813	A/G	81 (46.6)	96 (44.4)	64 (32.3)	135 (13.4)	74 (7.3)	237 (24.2)
	38347301	rs866771816	T/C	1 (0.6)	0	0	0	0	0
	38347341	rs11536975	G/A	2 (1.1)	3 (1.4)	8 (4)	0	0	0
	38347371	rs112363124	T/C	1 (0.6)	0	0	39 (3.9)	5 (0.5)	88 (9)
	38347473	rs146512052	G/C	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38347624	rs6099228	A/C	68 (39.1)	55 (25.5)	42 (21.2)	16 (1.6)	0	32 (3.3)
	38347700	rs7267449	A/G	67 (38.5)	55 (25.5)	42 (21.2)	17 (1.7)	0	32 (3.3)
	38347924	rs7267600	A/G	67 (38.5)	55 (25.5)	42 (21.2)	16 (1.6)	0	32 (3.3)
	38348258	rs2076807646	C/A	1 (0.6)	0	0	0	0	0
	38348300	rs8120088	A/G	68 (39.1)	55 (25.5)	42 (21.2)	16 (1.6)	0	32 (3.3)
Intron 1	38348317	rs766381612	**A/AT**	1 (0.6)	-	-	-	-	-
	38348353	rs1031122290	G/A	1 (0.6)	-	-	-	-	-
	38348505	rs6123628	C/T	19 (10.9)	9 (4.2)	21 (10.6)	2 (0.2)	138 (13.7)	4 (0.4)
	38348544	rs148840294	C/T	1 (0.6)	6 (2.8)	8 (4.0)	0	0	0
	38348545	rs899221984	G/T	1 (0.6)	-	-	-	-	-
	38348550	rs6025117	G/C	3 (1.7)	1 (0.5)	0	0	0	0
	38348776	rs1491192102	**TAG/T**	1 (0.6)	-	-	-	-	-
	38348824	rs11536978	C/T	9 (5.2)	15 (6.9)	7 (3.5)	0	0	0
	38349121	rs11536940	G/A	29 (16.7)	24 (11.1)	20 (10.1)	14 (1.4)	0	32 (3.3)
	38349144	rs140501730	G/A	43 (24.7)	47 (21.8)	35 (17.7)	1 (0.1)	0	0
	38349233	rs6092380	G/A	29 (16.7)	31 (14.4)	35 (17.7)	11 (1.1)	1 (0.1)	24 (2.5)
	38349344	rs1001474580	C/T	2 (1.1)	0	0	0	0	0
	38349684	rs370012899	C/T	1 (0.6)	0	0	0	0	0
	38349779	rs533481732	C/T	12 (6.9)	-	-	-	-	-
	38350107	rs6025196	C/T	21 (12.1)	30 (13.9)	8 (4.0)	0	1 (0.1)	0
	38350192	rs192827551	G/A	7 (4.0)	2 (0.9)	2 (1.0)	0	0	0
Intron 2	38350280	rs146955220	G/A	3 (1.7)	8 (3.7)	11 (5.6)	0	0	0
	38350359	rs2076840912	**CA/C**	4 (2.3)	-	-	-	-	-
	38350397	rs370151435	A/G	1 (0.6)	1 (0.5)	0	0	0	0
	38350608	rs6099236	G/C	80(46.0)	88(40.7)	67(33.8)	845(16.0)	240(23.8)	347(35.5)
	38350646	rs5741814	C/T	53 (30.5)	56 (25.9)	46 (23.2)	95 (9.4)	68 (6.7)	150 (15.3)
	38350730	rs1780627	T/C	80(46.0)	120(55.6)	99(50.0)	476(52.7)	845(83.8)	745(76.2)
	38350967	rs2232584	G/A	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38351111	rs5741815	T/C	76 (43.7)	87 (40.3)	65 (32.8)	132 (13.1)	52 (5.2)	230 (23.5)
	38351189	rs112288993	C/T	6 (3.4)	-	-	-	-	-
	38351261	rs6025082	T/C	1 (0.6)	2 (0.9)	8 (4.0)	0	0	0
	38351384	rs149349880	**TA/T**	40 (23.0)	38 (17.6)	34 (17.2)	1 (0.1)	0	0
	38351393	rs376769663	C/T	7 (4.0)	0	0	0	0	0
	38351418	rs1057402282	A/G	2 (1.1)	-	-	-	-	-
	38351619	rs6025119	G/A	3 (1.7)	4 (1.9)	0	0	0	0
	38351693	rs1259960883	G/T	13 (7.5)	-	-	-	-	-
	38352038	rs1375812198	**T/TA**	1 (0.6)	-	-	-	-	-
	38352050	rs144933085	C/G	2 (1.1)	2 (0.9)	1 (0.5)	0	0	0
	38352311	rs1046104659	T/A	1 (0.6)	0	0	0	0	0
	38352580	rs6099254	T/C	51 (29.3)	-	-	-	-	-
	38352668	rs6099309	G/T	39 (22.4)	64 (29.6)	54 (27.3)	84 (8.3)	66 (6.5)	194 (19.8)
Intron 3	38352669	rs145718383	C/T	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38352742	rs141862127	A/C	3 (1.7)	2 (0.9)	2 (1)	0	0	0
	38352768	rs11536948	G/T	5 (2.9)	2 (0.9)	5 (2.5)	0	0	0
	38352830	rs566402241	T/C	4 (2.3)	0	0	0	0	1 (0.1)
	38353035	rs6014804	G/C	74 (42.5)	86 (39.8)	73 (36.9)	121 (12.0)	51 (5.1)	203 (20.8)
	38353055	rs898782437	C/T	1 (0.6)	-	-	-	-	-
	38353056	rs6014805	A/T	74 (42.5)	84 (38.9)	71 (35.9)	41 (4.1)	5 (0.5)	87 (8.9)
	38353359	rs1018470	G/T	13 (7.5)	18 (8.3)	28 (14.1)	467 (46.4)	75 (7.4)	241 (24.6)
	38353484	rs76585941	C/T	51 (29.3)	30 (13.9)	38 (19.2)	11 (1.1)	1 (0.1)	24 (2.5)
	38353504	rs1568834425	C/T	3 (1.7)	-	-	-	-	-
	38353549	rs137996210	**T/TATAA**	31 (17.8)	36 (16.7)	37 (18.7)	121 (12.0)	50 (5.0)	205 (21.0)
	38353609	*LBP* NI-1	A/G	1 (0.6)	-	-	-	-	-
	38353618	rs111067091	C/T	4 (2.3)	27 (12.5)	12 (6.1)	40 (4.0)	5 (0.5)	87 (8.9)
	38353941	rs6025106	A/G	1 (0.6)	2 (0.9)	8 (4.0)	0	0	0
	38354160	rs866351464	T/C	1 (0.6)	-	-	-	-	-
	38354451	rs1385004224	T/G	1 (0.6)	-	-	-	-	-
	38354467	rs6025191	T/C	29 (16.7)	63 (29.2)	25 (12.6)	43 (4.3)	6 (0.6)	90 (9.2)
	38354488	rs2232588	C/T	1 (0.6)	0	6 (3.0)	2 (0.2)	135 (13.4)	4 (0.4)
	38354501	rs953713220	G/A	1 (0.6)	-	-	-	-	-
	38354582	rs756001841	T/C	9 (5.2)	-	-	-	-	-
	38354713	rs999577398	G/A	1 (0.6)	-	-	-	-	-
	38354795	rs2076832874	A/G	1 (0.6)	-	-	-	-	-
Intron 4	38354851	rs1032769678	G/C	3 (1.7)	-	-	-	-	-
	38354927	rs2122620263	T/C	3 (1.7)	-	-	-	-	-
	38355000	rs5741818	G/A	21 (12.1)	30 (13.9)	8 (4.0)	0	1 (0.1)	0
	38355090	rs11906594	A/G	31 (17.8)	37 (17.1)	37 (18.7)	121 (12.0)	52 (5.2)	205 (21.0)
	38355173	rs142240566	**GCACACACCA/G**	13 (7.5)	8 (3.7)	9 (4.5)	68 (6.8)	46 (4.6)	117 (12.0)
	38355230	rs60588211	T/C	33 (19.0)	41 (19.0)	31 (15.7)	15 (1.5)	0	32 (3.3)
	38355270	rs6025083	C/T	34 (19.5)	62 (28.7)	75 (37.9)	571 (56.8)	497 (49.3)	409 (41.8)
	38355274	rs5744207	C/T	5 (2.9)	2 (0.9)	5 (2.5)	0	0	0

	38355449	rs201030283	**GT/G**	119 (68.4)	-	-	-	-	-
	38355518	*LBP* NI-2	T/C	1 (0.6)	-	-	-	-	-
	38355575	rs75640389	T/C	3 (1.7)	7 (3.2)	1 (0.5)	0	0	0
	38355579	rs2232592	G/A	12 (6.9)	7 (3.2)	16 (8.1)	2 (0.2)	136 (13.5)	4 (0.4)
	38355607	rs5744206	G/A	80 (46)	74 (34.3)	54 (27.3)	6 (0.6)	1 (0.1)	22 (2.2)
	38355608	rs2232609	A/G	21 (12.1)	30 (13.9)	8 (4.0)	0	1 (0.1)	0
	38355614	rs6127842	C/T	5 (2.9)	6 (2.8)	2 (1.0)	0	0	0
	38355756	rs145588209	C/T	3 (1.7)	2 (0.9)	1 (0.5)	0	6 (0.6)	0
	38355815	rs918313285	G/A	1 (0.6)	-	-	-	-	-
	38356058	rs76833431	A/C	29 (16.7)	41 (19.0)	37 (18.7)	110 (10.9)	189 (18.8)	209 (21.4)
	38356102	rs142567294	**CT/C**	7 (4)	-	-	-	-	-
	38356130	rs200069451	T/C	1 (0.6)	-	-	-	-	-
	38356133	rs1410967743	T/C	1 (0.6)	-	-	-	-	-
	38356221	rs11536951	G/A	43 (24.7)	47 (21.8)	35 (17.7)	1 (0.1)	0	1 (0.1)
	38356227	rs149255417	G/A	3 (1.7)	7 (3.2)	6 (3.0)	0	0	0
	38356347	rs6123638	A/G	15 (8.6)	7 (3.2)	20 (10.1)	2 (0.2)	151 (15.0)	6 (0.6)
	38356388	rs1388740773	A/C	1 (0.6)	-	-	-	-	-
	38356398	rs146018769	**AC/A**	1 (0.6)	4 (1.9)	0	0	0	0
	38356441	rs16987263	T/C	44 (25.3)	47 (21.8)	35 (17.7)	1 (0.1)	0	0
	38356511	rs11536949	G/T	13 (7.5)	17 (7.9)	29 (14.6)	77 (7.7)	1 (0.1)	51 (5.2)
	38356556	rs144732000	G/A	1 (0.6)	2 (0.9)	8 (4.0)	0	0	0
	38356567	rs6127855	G/A	17 (9.8)	7 (3.2)	20 (10.1)	2 (0.2)	131 (13.0)	4 (0.4)
	38356587	rs148541728	A/C	1 (0.6)	2 (0.9)	8 (4.0)	0	0	0
	38356647	rs16987261	A/G	52 (29.9)	53 (24.5)	37 (18.7)	1 (0.1)	0	0
	38356704	rs116743201	C/T	52 (29.9)	66 (30.6)	38 (19.2)	2 (0.2)	0	0
	38356825	rs141831997	C/T	45 (25.9)	60 (27.8)	36 (18.2)	2 (0.2)	0	0
	38356860	rs1739655	G/A	52(29.9)	48(22.2)	41(20.7)	965(4.1)	166(16.5)	141(14.4)
	38356929	rs6127831	G/A	48(27.6)	81(37.5)	80(40.4)	135(86.6)	819(81.3)	742(75.9)
	38356950	rs557025533	G/A	1 (0.6)	-	-	-	-	-
	38356981	rs139715955	G/A	10 (5.7)	3 (1.4)	5 (2.5)	0	0	0
	38356986	rs114631382	T/G	54 (31)	75 (34.7)	46 (23.2)	2 (0.2)	1 (0.1)	0
	38357097	rs147651643	T/C	58 (33.3)	75 (34.7)	43 (21.7)	2 (0.2)	1 (0.1)	0
	38357098	rs2232616	C/T	44 (25.3)	72 (33.3)	59 (29.8)	85 (8.4)	59 (5.9)	191 (19.5)
	38357100	rs116538384	T/C	58 (33.3)	75 (34.7)	43 (21.7)	2 (0.2)	1 (0.1)	0
	38357107	rs149725981	A/C	11 (6.3)	5 (2.3)	9 (4.5)	0	0	1 (0.1)
	38357124	rs145503460	A/G	58 (33.3)	75 (34.7)	43 (21.7)	2 (0.2)	1 (0.1)	0
	38357162	rs73285089	T/C	7 (4.0)	30 (13.9)	18 (9.1)	40 (4.0)	5 (0.5)	87 (8.9)
	38357231	rs2232617	C/T	15 (8.6)	41 (19.0)	24 (12.1)	73 (7.3)	57 (5.7)	160 (16.4)
	38357381	rs2232593	C/T	12 (6.9)	6 (2.8)	15 (7.6)	2 (0.2)	136 (13.5)	4 (0.4)
Intron 5	38357413	rs113030679	T/C	5 (2.9)	5 (2.3)	5 (2.5)	0	1 (0.1)	0
	38357493	rs148853836	G/A	1 (0.6)	1 (0.5)	3 (1.5)	0	1 (0.1)	0
	38357650	rs191771004	A/G	1 (0.6)	1 (0.5)	0	0	0	0
	38357856	rs115635753	A/G	8 (4.6)	6 (2.8)	2 (1.0)	0	0	0
	38358273	rs2232598	G/A	5 (2.9)	7 (3.2)	3 (1.5)	0	0	0
	38358344	rs146425008	C/T	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38358376	rs78713715	A/G	8 (4.6)	20 (9.3)	2 (1)	1 (0.1)	0	0
	38358471	rs76658360	C/T	8 (4.6)	17 (7.9)	2 (1)	1 (0.1)	0	0
	38358673	rs2076863761	C/T	2 (1.1)	-	-	-	-	-
	38358698	rs911476950	G/A	2 (1.1)	-	-	-	-	-
	38358867	rs11086581	**T/TA**	44 (25.3)	72 (33.3)	59 (29.8)	84 (8.3)	59 (5.9)	191 (19.5)
	38359006	rs12624843	G/A	17 (9.8)	31 (14.4)	36 (18.2)	388 (38.6)	696 (69)	414 (42.3)
	38359164	rs6069875	T/C	7 (4.0)	30 (13.9)	18 (9.1)	40 (4.0)	5 (0.5)	88 (9.0)
	38359231	rs540710354	**GA/G**	17 (9.8)	19 (8.8)	18 (9.1)	30 (3.0)	161 (16)	31 (3.2)
	38359276	rs6025169	G/A	22 (12.6)	21 (9.7)	17 (8.6)	36 (3.6)	5 (0.5)	90 (9.2)
	38359319	rs11536954	G/T	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38359359	rs11536955	C/T	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38359424	rs11536956	T/A	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38359601	rs147999452	T/C	2 (1.1)	2 (0.9)	1 (0.5)	0	0	0
	38359602	rs2232612	T/C	22 (12.6)	30 (13.9)	8 (4)	0	1 (0.1)	0
	38359609	rs543189741	A/C	8 (4.6)	6 (2.8)	2 (1.0)	0	1 (0.1)	0
	38359659	rs11536957	A/G	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38359700	rs150498373	G/A	2 (1.1)	2 (0.9)	1 (0.5)	0	0	0
	38359715	rs1401500360	**G/GA**	1 (0.6)	-	-	-	-	-
	38359717	rs1392211278	C/T	1 (0.6)	-	-	-	-	-
	38359721	rs1333551204	G/C	1 (0.6)	-	-	-	-	-
	38359722	rs1374925378	T/A	1 (0.6)	-	-	-	-	-
	38359837	rs11536961	G/A	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38359989	rs11536962	C/T	8 (4.6)	20 (9.3)	2 (1.0)	1 (0.1)	0	0
	38360004	rs11536963	C/A	10 (5.7)	3 (1.4)	2 (1.0)	0	0	0
	38360215	rs559170776	**A/AG**	2 (1.1)	2 (0.9)	2 (1.0)	0	0	0
	38360273	rs1339224831	**AAC/A**	8 (4.6)	-	-	-	-	-
	38360427	rs73909011	C/T	1 (0.6)	8 (3.7)	8 (4.0)	1 (0.1)	0	0
	38360442	rs6099277	A/C	2 (1.1)	1 (0.5)	5 (2.5)	51 (5.1)	18 (1.8)	21 (2.1)
	38360492	rs2076855946	G/T	1 (0.6)	-	-	-	-	-
	38360561	rs1600726016	C/T	1 (0.6)	-	-	-	-	-
	38360615	rs182432670	G/T	16 (9.2)	7 (3.2)	19 (9.6)	2 (0.2)	146 (14.5)	6 (0.6)
	38360648	rs1780624	G/C	69 (39.7)	78 (36.1)	71 (35.9)	503 (50.0)	121 (12.0)	357 (36.5)
	38360680	rs2232594	T/G	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	0	0
	38360681	rs1739638	G/A	17(9.8)	27(12.5)	34(17.2)	603(40.1)	707(70.1)	408(41.7)

	38360869	rs8183330	G/A	5 (2.9)	2 (0.9)	5 (2.5)	2 (0.2)	146 (14.5)	6 (0.6)
	38360889	rs574484429	G/A	1 (0.6)	0	0	1 (0.1)	0	0
	38361036	rs187398369	C/T	1 (0.6)	3 (1.4)	1 (0.5)	0	0	0
	38361049	rs2076858423	C/G	1 (0.6)	-	-	-	-	-
	38361068	rs143197158	A/G	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	0	0
	38361078	rs146873840	A/C	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	1 (0.1)	0
	38361168	rs11536959	T/G	2 (1.1)	23 (10.6)	9 (4.5)	1 (0.1)	0	1 (0.1)
	38361358	rs76328747	T/C	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	0	0
	38361377	rs146541773	C/T	5 (2.9)	21 (9.7)	6 (3.0)	48 (4.8)	80 (7.9)	147 (15.0)
	38361388	rs79741204	T/C	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	0	0
	38361527	rs149011164	C/G	6 (3.4)	24 (11.1)	5 (2.5)	1 (0.1)	0	0
	38361768	rs143748743	T/C	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	38361813	rs11696479	A/G	23 (13.2)	40 (18.5)	19 (9.6)	121 (12.0)	141 (14.0)	303 (31.0)
	38362074	rs1187483579	**TG/T**	1 (0.6)	-	-	-	-	-
	38362077	rs1165222643	G/T	1 (0.6)	-	-	-	-	-
	38362190	rs564425592	C/T	1 (0.6)	3 (1.4)	1 (0.5)	0	0	0
	38362215	rs190254523	G/A	5 (2.9)	21 (9.7)	6 (3.0)	48 (4.8)	80 (7.9)	147 (15.0)
	38362216	rs139169382	C/T	5 (2.9)	21 (9.7)	6 (3.0)	47 (4.7)	80 (7.9)	146 (14.9)
	38362254	rs1631643	T/A	17(9.8)	26(12.0)	34(17.2)	604(40.0)	709(70.3)	413(42.2)
	38362293	rs62201547	T/C	5 (2.9)	21 (9.7)	6 (3.0)	47 (4.7)	80 (7.9)	147 (15.0)
	38362333	rs117832305	G/A	6 (3.4)	24 (11.1)	7 (3.5)	47 (4.7)	80 (7.9)	146 (14.9)
Intron 6	38362336	rs566746432	G/C	1 (0.6)	3 (1.4)	1 (0.5)	0	0	0
	38362346	rs193188296	C/T	2 (1.1)	3 (1.4)	7 (3.5)	0	0	0
	38362356	rs111633395	T/C	10 (5.7)	0	0	0	0	0
	38362367	rs11536965	G/T	5 (2.9)	21 (9.7)	6 (3.0)	48 (4.8)	79 (7.8)	146 (14.9)
	38362392	rs1444787557	T/C	1 (0.6)	-	-	-	-	-
	38362422	rs13043870	A/G	78 (44.8)	108 (50)	85 (42.9)	565 (56.2)	153 (15.2)	450 (46)
	38362554	rs11536964	G/A	6 (3.4)	24 (11.1)	7 (3.5)	48 (4.8)	80 (7.9)	146 (14.9)
	38362570	rs11536968	G/A	5 (2.9)	21 (9.7)	6 (3.0)	48 (4.8)	79 (7.8)	145 (14.8)
	38362618	rs546284088	**AG/A**	1 (0.6)	3 (1.4)	0	0	0	0
	38362766	rs11536967	A/G	6 (3.4)	24 (11.1)	7 (3.5)	48 (4.8)	79 (7.8)	146 (14.9)
	38362768	rs181255547	C/T	1 (0.6)	3 (1.4)	2 (1.0)	0	0	0
	38362777	rs185844320	A/G	1 (0.6)	3 (1.4)	1 (0.5)	0	0	0
	38362897	rs2076876352	**T/TAA**	5 (2.9)	-	-	-	-	-
	38362923	rs1780628	A/T	17(9.8)	25(11.6)	36(18.2)	558(44.5)	793(78.7)	551(56.3)
	38362953	rs1204006801	G/T	1 (0.6)	-	-	-	-	-
	38362997	rs11536970	T/C	1 (0.6)	4 (1.9)	2 (1.0)	0	0	0
	38363237	rs11536972	T/C	6 (3.4)	21 (9.7)	7 (3.5)	48 (4.8)	79 (7.8)	146 (14.9)
	38363470	rs11536974	C/T	6 (3.4)	21 (9.7)	7 (3.5)	48 (4.8)	79 (7.8)	145 (14.8)
	38363566	rs6127841	C/T	12 (6.9)	5 (2.3)	15 (7.6)	2 (0.2)	138 (13.7)	4 (0.4)
	38363737	rs1739640	T/C	80(46.0)	120(55.6)	99(50.0)	472(53.1)	852(84.5)	743(76.0)
	38363818	rs2232599	G/C	10 (5.7)	29 (13.4)	11 (5.6)	48 (4.8)	79 (7.8)	145 (14.8)
	38363896	rs2232600	G/A	1 (0.6)	3 (1.4)	3 (1.5)	0	0	0
	38364073	rs1444691492	C/T	2 (1.1)	-	-	-	-	-
	38364156	rs5741816	A/G	82(47.1)	129(59.7)	125(63.1)	13(98.7)	1007(99.9)	954(97.5)
	38364195	rs11536977	G/A	1 (0.6)	4 (1.9)	0	0	0	0
Intron 7	38364460	rs1600727734	G/A	1 (0.6)	-	-	-	-	-
	38364501	rs113464344	G/A	12 (6.9)	0	0	0	1 (0.1)	0
	38364525	rs2232605	G/C	1 (0.6)	0	3 (1.5)	47 (4.7)	76 (7.5)	145 (14.8)
	38364552	rs2232606	A/G	7 (4.0)	19 (8.8)	16 (8.1)	42 (4.2)	6 (0.6)	90 (9.2)
	38364938	rs11536992	T/C	3 (1.7)	4 (1.9)	0	0	0	0
	38365005	rs1780626	T/G	82(47.1)	123(56.9)	99(50.0)	509(49.4)	790(78.4)	667(68.2)
	38365051	rs6025095	C/T	51 (29.3)	50 (23.1)	37 (18.7)	0	0	0
	38365233	rs149506270	C/T	5 (2.9)	7 (3.2)	6 (3)	0	0	0
	38365276	rs1010323636	T/A	1 (0.6)	-	-	-	-	-
	38365289	rs185085448	G/A	1 (0.6)	1 (0.5)	0	0	0	0
	38365340	rs998520711	G/A	5 (2.9)	-	-	-	-	-
	38365354	rs113966205	T/C	12 (6.9)	1 (0.5)	0	0	0	0
	38365471	rs1780625	C/T	70 (40.2)	84 (38.9)	78 (39.4)	499 (49.6)	67 (6.6)	305 (31.2)
	38365500	rs563241248	C/T	7 (4.0)	2 (0.9)	2 (1)	0	0	0
	38365508	rs74802071	**TTTTG/T**	3 (1.7)	-	-	-	-	-
	38365620	rs548140872	C/G	1 (0.6)	0	2 (1)	0	0	0
	38365654	rs6025214	C/A	16 (9.2)	23 (10.6)	5 (2.5)	0	1 (0.1)	0
Intron 8	38365731	rs2232581	T/G	83 (47.7)	95 (44.0)	65 (32.8)	135 (13.4)	74 (7.3)	238 (24.3)
	38365789	rs139872326	A/G	2 (1.1)	3 (1.4)	2 (1)	0	0	0
	38365863	rs181695299	T/C	7 (4.0)	2 (0.9)	2 (1)	0	0	0
	38365890	rs1739641	T/C	83(47.7)	122(56.5)	101(51.0)	473(53.0)	852(84.5)	740(75.7)
	38365999	rs111874373	T/A	12 (6.9)	0	0	0	0	0
	38366025	rs6014890	A/C	22 (12.6)	37 (17.1)	20 (10.1)	88 (8.7)	57 (5.7)	168 (17.2)
	38366199	rs5741817	C/T	85(48.9)	113(52.3)	105(53.0)	552(45.1)	930(92.3)	582(59.5)
	38366261	rs11536936	C/T	9 (5.2)	3 (1.4)	5 (2.5)	0	0	0
	38366336	rs113365137	G/A	1 (0.6)	-	-	-	-	-
	38366365	rs11536979	C/G	1 (0.6)	4 (1.9)	0	0	0	0
	38366472	rs2232602	G/A	9 (5.2)	28 (13.0)	10 (5.1)	48 (4.8)	76 (7.5)	145 (14.8)
	38366583	rs11536939	C/T	23 (13.2)	23 (10.6)	20 (10.1)	14 (1.4)	0	32 (3.3)
	38366696	rs1739639	T/C	70 (40.2)	84 (38.9)	77 (38.9)	505 (50.2)	67 (6.6)	305 (31.2)
	38366752	rs6099305	T/C	23 (13.2)	24 (11.1)	31 (15.7)	11 (1.1)	1 (0.1)	24 (2.5)

	38366854	rs2232611	A/T	1 (0.6)	4 (1.9)	0	0	0	0
	38366873	rs113489543	C/T	12 (6.9)	-	-	-	-	-
	38366973	rs76186816	A/C	25 (14.4)	41 (19)	39 (19.7)	110 (10.9)	190 (18.8)	209 (21.4)
	38367020	rs3819023	A/G	1 (0.6)	0	2 (1.0)	47 (4.7)	76 (7.5)	147 (15.0)
	38367046	rs6025064	T/C	29 (16.7)	50 (23.1)	29 (14.6)	120 (11.9)	73 (7.2)	206 (21.1)
	38367274	rs11536994	C/T	15 (8.6)	41 (19.0)	24 (12.1)	73 (7.3)	57 (5.7)	167 (17.1)
	38367446	rs140231834	A/T	3 (1.7)	2 (0.9)	2 (1.0)	0	0	0
	38367448	rs34553900	C/T	30 (17.2)	37 (17.1)	36 (18.2)	121 (12.0)	51 (5.1)	205 (21.0)
	38367550	rs186406309	A/T	1 (0.6)	4 (1.9)	0	0	0	0
	38367585	rs41332344	C/T	30 (17.2)	44 (20.4)	45 (22.7)	122 (12.1)	187 (18.6)	211 (21.6)
	38367646	rs113579225	A/C	12 (6.9)	0	0	0	0	0
	38367808	rs112318889	C/T	12 (6.9)	0	0	0	0	0
	38367921	rs6025195	C/T	9 (5.2)	30 (13.9)	8 (4)	0	1 (0.1)	0
Intron 9	38367980	rs139111516	C/T	1 (0.6)	2 (0.9)	0	0	0	0
	38368034	rs6069931	A/G	15 (8.6)	56 (25.9)	25 (12.6)	73 (7.3)	58 (5.8)	167 (17.1)
	38368104	rs35133348	C/T	18(10.3)	21(9.7)	39(19.7)	608(39.6)	291(28.9)	289(29.6)
	38368117	rs188701226	T/C	7 (4)	2 (0.9)	2 (1.0)	0	0	0
	38368127	rs113031135	C/G	12 (6.9)	-	-	-	-	-
	38368195	rs562936892	A/G	1 (0.6)	2 (0.9)	2 (1.0)	0	0	0
	38368328	rs11536991	G/A	5 (2.9)	2 (0.9)	5 (2.5)	0	0	0
	38368335	rs111848745	A/T	12 (6.9)	0	0	0	0	0
	38368341	rs113734860	C/G	12 (6.9)	0	0	0	0	0
	38368705	rs6025208	T/C	41(23.6)	64(29.6)	45(22.7)	557(44.6)	784(77.8)	553(56.5)
	38368724	rs187402189	C/T	2 (1.1)	0	1 (0.5)	0	0	0
	38368747	rs113203674	C/T	12 (6.9)	-	-	-	-	-
	38368889	rs112685620	T/A	12 (6.9)	-	-	-	-	-
	38368988	rs6025198	G/A	9 (5.2)	15 (6.9)	9 (4.5)	0	0	0
	38369193	rs112098655	A/C	12 (6.9)	-	-	-	-	-
	38369336	rs11536983	C/A	1 (0.6)	14 (6.5)	1 (0.5)	1 (0.1)	0	0
	38369496	rs113366471	T/C	12 (6.9)	-	-	-	-	-
	38369505	rs11481047	**T/TA**	58 (33.3)	87 (40.3)	60 (30.3)	85 (8.4)	60 (6)	191 (19.5)
	38369598	rs1714281893	**A/ATTTGCTCTGTTC**	12 (6.9)	-	-	-	-	-
	38369641	rs6014861	A/G	58 (33.3)	72 (33.3)	59 (29.8)	85 (8.4)	59 (5.9)	191 (19.5)
	38369645	rs2232589	G/A	30 (17.2)	44 (20.4)	45 (22.7)	122 (12.1)	187 (18.6)	210 (21.5)
Intron 10	38369667	rs6025241	A/G	45 (25.9)	72 (33.3)	59 (29.8)	85 (8.4)	59 (5.9)	191 (19.5)
	38369851	rs11698149	T/C	7 (4.0)	15 (6.9)	18 (9.1)	73 (7.3)	58 (5.8)	167 (17.1)
	38369907	rs554483061	C/A	3 (1.7)	1 (0.5)	1 (0.5)	0	0	0
	38370291	rs140999190	G/A	10 (5.7)	3 (1.4)	2 (1.0)	0	0	1 (0.1)
	38370296	rs6127837	G/A	48(27.6)	81(37.5)	80(40.4)	123(87.8)	817(81.1)	746(76.3)
	38370399	rs113369278	**TA/T**	17 (9.8)	4 (1.9)	4 (2.0)	0	0	0
	38370686	rs112052603	C/T	12 (6.9)	0	0	0	0	0
	38370814	rs201651468	G/T	7 (4)	2 (0.9)	2 (1.0)	0	0	0
	38370887	rs112924468	A/G	12 (6.9)	0	0	0	0	0
	38371029	rs5741819	A/C	62 (35.6)	88 (40.7)	62 (31.3)	85 (8.4)	60 (6.0)	191 (19.5)
Intron 11	38371167	rs182922719	C/T	1 (0.6)	0	1 (0.5)	0	0	0
	38371231	rs6014862	T/A	44 (25.3)	72 (33.3)	59 (29.8)	85 (8.4)	59 (5.9)	191 (19.5)
	38371235	rs188388337	C/T	7 (4)	2 (0.9)	2 (1.0)	0	0	0
	38371253	rs1780622	A/C	48 (27.6)	83 (38.4)	85 (42.9)	845 (84)	767 (76.1)	633 (64.7)
	38371501	rs541743179	A/G	1 (0.6)	0	1 (0.5)	0	0	0
	38371549	*LBP* NI-4	A/T	1 (0.6)	-	-	-	-	-
	38371639	rs936053972	A/G	1 (0.6)	-	-	-	-	-
	38371808	rs6025102	T/C	48(27.6)	86(39.8)	90(45.5)	133(86.8)	955(94.7)	748(76.5)
	38371818	rs113207369	**AG/A**	12 (6.9)	0	0	0	0	0
	38372041	rs145688223	**G/GCA**	12 (6.9)	-	-	-	-	-
	38372132	rs944704809	C/T	1 (0.6)	-	-	-	-	-
Intron 12	38372285	rs112314157	T/A	12 (6.9)	0	0	0	0	0
	38372600	rs1013317527	G/C	1 (0.6)	-	-	-	-	-
	38372691	rs9753666	G/A	5 (2.9)	7 (3.2)	3 (1.5)	0	0	0
	38372750	rs149832282	T/C	2 (1.1)	3 (1.4)	2 (1.0)	0	0	0
	38372765	rs6127849	C/T	17 (9.8)	7 (3.2)	20 (10.1)	2 (0.2)	138 (13.7)	4 (0.4)
	38372814	rs1609800	A/T	20 (11.5)	43 (19.9)	44 (22.2)	472 (46.9)	75 (7.4)	241 (24.6)
	38372909	rs1780623	C/T	20(11.5)	30(13.9)	37(18.7)	616(38.8)	696(69)	410(41.9)
	38372987	rs11696597	T/G	16 (9.2)	32 (14.8)	21 (10.6)	108 (10.7)	52 (5.2)	204 (20.9)
	38373290	rs6014840	T/C	9 (5.2)	30 (13.9)	8 (4.0)	0	1 (0.1)	0
	38373313	rs2232590	C/G	30 (17.2)	44 (20.4)	45 (22.7)	122 (12.1)	187 (18.6)	211 (21.6)
	38373329	rs6123629	C/G	16 (9.2)	7 (3.2)	19 (9.6)	2 (0.2)	138 (13.7)	4 (0.4)
	38373379	rs6069876	T/G	16 (9.2)	50 (23.1)	21 (10.6)	41 (4.1)	6 (0.6)	88 (9.0)
Intron 13	38373450	rs6025098	C/T	31 (17.8)	37 (17.1)	37 (18.7)	121 (12.0)	52 (5.2)	205 (21)
	38373612	rs113915430	C/A	12 (6.9)	0	0	0	0	0
	38373736	rs113525611	G/A	12 (6.9)	0	0	0	0	0
	38373741	rs11536995	**AGTTACTTG/A**	11 (6.3)	33 (15.3)	20 (10.1)	73 (7.3)	57 (5.7)	167 (17.1)
	38373883	rs374356852	C/T	1 (0.6)	0	1 (0.5)	0	0	0

	38374083	rs112652150	G/A	12 (6.9)	0	0	0	0	1 (0.1)
	38374274	rs140486354	G/A	1 (0.6)	6 (2.8)	8 (4)	0	0	0
	38374336	rs6025188	C/T	9 (5.2)	15 (6.9)	7 (3.5)	0	0	1 (0.1)
	38374337	rs556659001	G/A	1 (0.6)	0	0	0	0	0
	38374426	rs2232595	C/T	16 (9.2)	7 (3.2)	19 (9.6)	2 (0.2)	141 (14)	6 (0.6)
	38374553	rs114618390	T/G	21 (12.1)	26 (12.0)	31 (15.7)	11 (1.1)	1 (0.1)	24 (2.5)
	38374594	rs562385988	**CA/C**	1 (0.6)	6 (2.8)	10 (5.1)	32 (3.2)	15 (1.5)	14 (1.4)
	38374659	rs138755345	C/T	5 (2.9)	12 (5.6)	13 (6.6)	89 (8.8)	57 (5.7)	168 (17.2)
	38374802	rs11536993	G/A	15 (8.6)	41 (19)	24 (12.1)	73 (7.3)	57 (5.7)	167 (17.1)
	38374846	rs112038028	A/G	12 (6.9)	0	0	0	0	0
	38374847	rs6014889	G/A	10 (5.7)	49 (22.7)	19 (9.6)	73 (7.3)	58 (5.8)	167 (17.1)
	38374947	rs6099300	A/G	50 (28.7)	79 (36.6)	53 (26.8)	54 (5.4)	7 (0.7)	115 (11.8)
	38374969	rs1028648588	G/T	11 (6.3)	-	-	-	-	-
	38374971	rs888806866	A/T	7 (4.0)	-	-	-	-	-
Intron 14	38375082	rs2232587	T/C	26 (14.9)	21 (9.7)	27 (13.6)	11 (1.1)	1 (0.1)	24 (2.5)
	38375177	rs549719574	G/C	1 (0.6)	-	-	-	-	-
	38375195	rs6025049	G/A	11(6.3)	33(15.3)	25(12.6)	861(14.4)	240(23.8)	315(32.2)
	38375394	rs6069847	G/A	29 (16.7)	42 (19.4)	38 (19.2)	120 (11.9)	51 (5.1)	207 (21.2)
	38375462	rs6064469	T/C	6 (3.4)	28 (13.0)	14 (7.1)	90 (8.9)	59 (5.9)	168 (17.2)
	38375473	rs6127861	G/A	16 (9.2)	7 (3.2)	20 (10.1)	2 (0.2)	146 (14.5)	6 (0.6)
	38375656	*LBP* NI-5	T/G	1 (0.6)	-	-	-	-	-
	38375671	rs2083954674	A/G	1 (0.6)	-	-	-	-	-
	38375766	rs73909020	G/A	1 (0.6)	23 (10.6)	6 (3.0)	0	1 (0.1)	0
	38375815	rs73909021	C/T	1 (0.6)	23 (10.6)	6 (3.0)	0	1 (0.1)	0
	38375817	rs73909022	A/T	1 (0.6)	23 (10.6)	6 (3.0)	0	1 (0.1)	0
	38375982	rs140069602	G/C	21 (12.1)	4 (1.9)	3 (1.5)	0	0	0
	38376206	rs114181807	C/T	1 (0.6)	8 (3.7)	5 (2.5)	0	0	0
	38376256	rs11086565	A/G	21 (12.1)	4 (1.9)	5 (2.5)	67 (6.7)	47 (4.7)	116 (11.9)
	38376323	rs572665635	A/C	21 (12.1)	8 (3.7)	5 (2.5)	67 (6.7)	48 (4.8)	113 (11.6)
	38376389	rs113826326	G/C	8 (4.6)	-	-	-	-	-

Wt/mut: wildtype/mutant referring to major and minor alleles, respectively; **SA:** South Africa; **YRI:** Yoruba in Ibadan, Nigeria; **LWK:** Luhya from Webuye, Kenya; **EUR:** European population; **EAS:** East Asians population; **SAS:** South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); **-:** No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic). NI: *LBP* newly identified SNP.

Supplementary Table 2: Minor allele frequencies of *LY96* genetic variants (introns) in the South African population compared to select populations from the 1000 Genomes Project.

				n(minnor allele frequency, %)					
Location on gene	Position (Chr:8)	Accession number	Base change (wt/mut)	SA (N=87)	YRI (N=108)	LWK (N=99)	EUR (N=503)	EAS (N=504)	SAS (N=489)
	73991901	rs16938755	T/C	66 (37.9)	82 (38.0)	89 (44.9)	151 (15.0)	117 (11.6)	160 (16.4)
	73992134	rs7829369	A/T	24 (13.8)	28 (13.0)	19 (9.6)	289 (28.7)	193 (19.1)	286 (29.2)
	73992737	rs969055264	T/C	2 (1.1)	-	-	-	-	-
	73992768	rs149392518	C/T	6 (3.4)	4 (1.9)	12 (6.1)	1 (0.1)	2 (0.2)	0
	73992781	rs187100096	G/C	1 (0.6)	0	0	0	5 (0.5)	0
	73992915	rs10086976	C/T	53 (30.5)	70 (32.4)	70 (35.4)	151 (15.0)	117 (11.6)	160 (16.4)
	73992917	rs192492675	C/T	3 (1.7)	0	1 (0.5)	0	1 (0.1)	0
	73992920	rs10102014	T/G	79 (45.4)	82 (38.0)	91 (46.0)	170 (16.9)	146 (14.5)	210 (21.5)
	73993021	rs1816042505	**A/AT**	3 (1.7)	-	-	-	-	-
	73993222	rs1586643110	T/A	2 (1.1)	-	-	-	-	-
	73993528	rs4738410	C/T	5 (2.9)	0	4 (2.0)	19 (1.9)	29 (2.9)	50 (5.1)
Intron 1	73993529	rs186659780	G/A	1 (0.6)	1 (0.5)	1 (0.5)	0	0	0
	73993535	rs4738411	C/T	71 (40.8)	106 (49.1)	88 (44.4)	548 (54.5)	669 (66.4)	482 (49.3)
	73993752	rs139208711	A/T	6 (3.4)	5 (2.3)	7 (3.5)	0	0	0
	73993767	rs149923973	T/C	2 (1.1)	17 (7.9)	3 (1.5)	0	0	0
	73993793	rs144988354	G/A	3 (1.7)	5 (2.3)	2 (1.0)	0	0	0
	73993796	rs7009353	G/C	71 (40.8)	106 (49.1)	88 (44.4)	548 (54.5)	669 (66.4)	482 (49.3)
	73993822	rs55722404	**T/TC**	71 (40.8)	106 (49.1)	88 (44.4)	548 (54.5)	669 (66.4)	482 (49.3)
	73993988	rs534972696	**G/GT**	4 (2.3)	-	-	-	-	-
	73994075	rs147059682	C/T	5 (2.9)	0	3 (1.5)	19 (1.9)	20 (2.0)	31 (3.2)
	73994279	rs10808798	C/T	71 (40.8)	105 (48.6)	87 (43.9)	548 (54.5)	669 (66.4)	482 (49.3)
	73994494	rs10957691	A/G	64 (36.8)	83 (38.4)	88 (44.4)	151 (15.0)	118 (11.7)	160 (16.4)
	73994599	rs546722838	G/A	9 (5.2)	-	-	-	-	-
	73994648	rs141586860	G/A	4 (2.3)	2 (0.9)	5 (2.5)	0	0	0
	73995014	rs79142015	A/G	5 (2.9)	0	4 (2.0)	19 (1.9)	30 (3.0)	49 (5.0)
	73995052	rs75274643	C/G	8 (4.6)	24 (11.1)	16 (8.1)	1 (0.1)	2 (0.2)	0
	73995163	rs537366563	T/C	4 (2.3)	-	-	-	-	-
	73995323	rs7821865	C/T	24 (13.8)	27 (12.5)	19 (9.6)	287 (28.5)	194 (19.2)	286 (29.2)
	73995387	rs75724780	C/A	2 (1.1)	18 (8.3)	3 (1.5)	0	0	0
	73995555	rs7822054	G/A	60 (34.5)	76 (35.2)	68 (34.3)	548 (54.5)	666 (66.1)	482 (49.3)
	73995578	rs7822407	G/A	53 (30.5)	70 (32.4)	71 (35.9)	159 (15.8)	116 (11.5)	162 (16.6)
	73995728	rs11409980	**A/AT**	65 (37.4)	82 (38.0)	85 (42.9)	150 (14.9)	118 (11.7)	160 (16.4)
	73995782	rs7822709	A/G	61 (35.1)	79 (36.6)	69 (34.8)	548 (54.5)	666 (66.1)	482 (49.3)
	73995912	rs58916587	G/A	1 (0.6)	3 (1.4)	3 (1.5)	81 (8.1)	1 (0.1)	13 (1.3)
	73995948	rs139870678	C/T	1 (0.6)	0	1 (0.5)	0	1 (0.1)	0
	73996229	rs112512079	**CT/C**	1 (0.6)	1 (0.5)	3 (1.5)	0	2 (0.2)	0
	73996266	rs28435175	G/A	65 (37.4)	82 (38.0)	86 (43.4)	151 (15.0)	116 (11.5)	161 (16.5)
	73996280	rs28718132	T/C	65 (37.4)	82 (38.0)	86 (43.4)	151 (15.0)	116 (11.5)	161 (16.5)
	73996294	rs537792961	C/G	1 (0.6)	0	0	0	0	1 (0.1)
	73996345	rs895555776	C/A	2 (1.1)	-	-	-	-	-
	73996349	rs949793915	C/A	2 (1.1)	-	-	-	-	-
	73996355	rs1491254890	**TCC/T**	2 (1.1)	-	-	-	-	-
	73996358	rs1563707666	**TTCCTTCCTTCATTC/T**	2 (1.1)	-	-	-	-	-
	73996463	*LY96* NI-1	G/C	1 (0.6)	-	-	-	-	-
	73996547	rs7387160	A/G	40 (23.0)	45 (20.8)	26 (13.1)	307 (30.5)	224 (22.2)	335 (34.3)
	73996708	rs7388563	T/G	42 (24.1)	45 (20.8)	26 (13.1)	306 (30.4)	224 (22.2)	335 (34.3)
	73996711	rs367942020	G/A	2 (1.1)					
	73996758	rs142448050	T/A	8 (4.6)	6 (2.8)	11 (5.6)	0	0	0
	73996964	rs148940132	C/T	16 (9.2)	18 (8.3)	7 (3.5)	19 (1.9)	31 (3.1)	49 (5.0)
	73997053	rs138082733	**TTTAG/T**	2 (1.1)	18 (8.3)	3 (1.5)	0	0	0
Intron 2	73997124	rs115006184	C/T	6 (3.4)	5 (2.3)	7 (3.5)	0	0	1 (0.1)
	73997312	rs115869250	C/T	6 (3.4)	4 (1.9)	12 (6.1)	1 (0.1)	2 (0.2)	0
	73997501	rs77952155	T/C	12 (6.9)	12 (5.6)	14 (7.1)	0	0	0
	73997562	rs115855837	A/T	3 (1.7)	4 (1.9)	2 (1.0)	1 (0.1)	0	0
	73998047	rs1207937804	G/A	1 (0.6)	-	-	-	-	-
	73998141	rs76061819	A/C	11 (6.3)	18 (8.3)	3 (1.5)	0	0	0
	73998330	rs12546552	G/A	42 (24.1)	45 (20.8)	26 (13.1)	306 (30.4)	224 (22.2)	335 (34.3)
	73998347	rs12546519	C/T	42 (24.1)	45 (20.8)	26 (13.1)	306 (30.4)	224 (22.2)	335 (34.3)
	73998481	rs141192619	G/A	4 (2.3)	6 (2.8)	4 (2.0)	0	0	0
	73998655	rs116892026	C/G	5 (2.9)	0	3 (1.5)	19 (1.9)	21 (2.1)	31 (3.2)
	73998691	rs60367712	**C/CA**	58 (33.3)	79 (36.6)	87 (43.9)	153 (15.2)	119 (11.8)	161 (16.5)
	73999108	rs11781740	C/T	25 (14.4)	27 (12.5)	19 (9.6)	288 (28.6)	194 (19.2)	286 (29.2)
	73999490	rs184492149	A/T	2 (1.1)	4 (1.9)	2 (1.0)	0	0	0
	73999799	rs7841862	A/G	73 (42.0)	92 (42.6)	85 (42.9)	548 (54.5)	666 (66.1)	482 (49.3)
	74000006	rs60515500	A/C	73 (42.0)	92 (42.6)	85 (42.9)	548 (54.5)	666 (66.1)	482 (49.3)
	74000060	rs76447356	G/A	2 (1.1)	18 (8.3)	3 (1.5)	0	0	0
	74000274	rs112390569	G/C	2 (1.1)	2 (0.9)	0	0	0	0
	74000278	rs79664747	**G/GTGTAGATCCTTGCTGAT**	23 (13.2)	25 (11.6)	18 (9.1)	276 (27.4)	187 (18.6)	275 (28.1)
	74000293	rs202196303	C/G	1 (0.6)	-	-	-	-	-
	74000294	rs200299427	T/A	1 (0.6)	-	-	-	-	-
	74000296	*LY96* NI-2	**A/AGTGT**	1 (0.6)	-	-	-	-	-
	74000299	rs1325156550	T/A	1 (0.6)	-	-	-	-	-
	74000301	*LY96* NI-3	**C/CTCGGTGGTCGCCGTAT**	1 (0.6)	-	-	-	-	-
	74000305	*LY96* NI-4	**TTTTC/T**	1 (0.6)	-	-	-	-	-
	74000309	*LY96* NI-5	**C/CAAA**	1 (0.6)	-	-	-	-	-
	74000311	rs1274874437	C/A	3 (1.7)	-	-	-	-	-
	74000484	rs72661869	C/T	24 (13.8)	27 (12.5)	19 (9.6)	287 (28.5)	194 (19.2)	286 (29.2)

	74000758	rs147206553	C/A	1 (0.6)	1 (0.5)	5 (2.5)	0	0	0
	74001285	rs62509331	G/A	24 (13.8)	27 (12.5)	19 (9.6)	287 (28.5)	194 (19.2)	286 (29.2)
	74001412	rs148462447	G/A	6 (3.4)	8 (3.7)	13 (6.6)	1 (0.1)	2 (0.2)	0
	74001652	rs62509332	G/A	42 (24.1)	45 (20.8)	26 (13.1)	307 (30.5)	224 (22.2)	335 (34.3)
	74001688	rs548591196	**C/CA**	5 (2.9)	-	-	-	-	-
	74001802	rs562243530	C/T	2 (1.1)	-	-	-	-	-
	74001817	rs115700825	G/A	2 (1.1)	18 (8.3)	3 (1.5)	0	0	0
	74001850	rs1008726652	A/C	1 (0.6)	-	-	-	-	-
	74001886	rs369815557	**GTATTTCCAT/G**	5 (2.9)	6 (2.8)	2 (1.0)	1 (0.1)	0	0
	74001959	rs1406963683	**CTT/C**	9 (5.2)	-	-	-	-	-
	74001962	rs1161163048	**CCTTCCTTCCT/C**	9 (5.2)	-	-	-	-	-
	74001975	rs190959982	C/A	13 (7.5)	19 (8.8)	11 (5.6)	1 (0.1)	0	0
	74001980	rs145178284	C/T	24 (13.8)	-	-	-	-	-
	74001984	rs796270377	C/T	7 (4.0)	-	-	-	-	-
	74001995	rs1312134372	**CTTT/C**	18 (10.3)	-	-	-	-	-
	74002133	rs911522615	T/C	1 (0.6)	-	-	-	-	-
	74002458	rs118149143	C/T	5 (2.9)	0	3 (1.5)	19 (1.9)	21 (2.1)	31 (3.2)
	74002527	rs1482388882	**AG/A**	3 (1.7)	-	-	-	-	-
	74002556	rs116821531	G/T	2 (1.1)	18 (8.3)	3 (1.5)	1 (0.1)	0	0
	74002569	rs143496381	A/C	1 (0.6)	7 (3.2)	1 (0.5)	3 (0.3)	3 (0.3)	2 (0.2)
	74002657	rs111747962	C/T	15 (8.6)	17 (7.9)	10 (5.1)	27 (2.7)	0	4 (0.4)
	74002746	rs34767907	**A/AT**	54 (31.0)	57 (26.4)	40 (20.2)	307 (30.5)	224 (22.2)	335 (34.3)
	74002799	rs189437001	C/T	6 (3.4)	4 (1.9)	12 (6.1)	1 (0.1)	2 (0.2)	0
	74002816	rs11775465	T/C	51 (29.3)	55 (25.5)	40 (20.2)	307 (30.5)	224 (22.2)	335 (34.3)
	74002825	rs181954162	C/T	4 (2.3)	-	-	-	-	-
	74002980	*LY96* NI-6	T/C	1 (0.6)	-	-	-	-	-
	74003202	rs138140428	**CAGG/C**	42 (24.1)	45 (20.8)	26 (13.1)	307 (30.5)	224 (22.2)	335 (34.3)
	74003316	rs114596749	T/C	5 (2.9)	14 (6.5)	12 (6.1)	0	0	0
	74003434	rs4738412	T/C	24 (13.8)	27 (12.5)	19 (9.6)	288 (28.6)	194 (19.2)	286 (29.2)
	74003534	rs114188248	C/T	6 (3.4)	4 (1.9)	12 (6.1)	1 (0.1)	2 (0.2)	0
	74003535	rs115211571	G/A	12 (6.9)	12 (5.6)	14 (7.1)	0	0	0
	74003582	rs116585932	T/C	4 (2.3)	2 (0.9)	5 (2.5)	0	0	0
	74003699	rs4738413	A/G	54 (31.0)	57 (26.4)	40 (20.2)	307 (30.5)	224 (22.2)	335 (34.3)
	74003779	rs56343979	C/T	42 (24.1)	45 (20.8)	26 (13.1)	307 (30.5)	223 (22.1)	336 (34.4)
	74003813	rs139146287	C/T	2 (1.1)	0	7 (3.5)	0	0	0
	74004109	rs11786591	C/T	54 (31)	57 (26.4)	40 (20.2)	307 (30.5)	223 (22.1)	335 (34.3)
	74004111	rs78144508	G/T	1 (0.6)	12 (5.6)	0	0	0	0
	74004415	rs549574553	T/C	1 (0.6)	0	0	0	0	0
	74004482	rs7001511	A/T	24 (13.8)	27 (12.5)	19 (9.6)	288 (28.6)	193 (19.1)	286 (29.2)
	74004517	rs113037498	T/C	9 (5.2)	-	-	-	-	-
Intron 2	74004655	rs148538964	G/A	7 (4.0)	1 (0.5)	9 (4.5)	0	0	0
	74004917	rs11466000	T/G	12 (6.9)	12 (5.6)	14 (7.1)	0	0	0
	74004993	rs11466001	A/G	24 (13.8)	27 (12.5)	19 (9.6)	288 (28.6)	195 (19.3)	286 (29.2)
	74005397	rs75878116	C/G	13 (7.5)	6 (2.8)	3 (1.5)	1 (0.1)	0	0
	74005936	rs11783456	G/A	54 (31)	57 (26.4)	40 (20.2)	307 (30.5)	223 (22.1)	336 (34.4)
	74005960	rs116507106	G/A	12 (6.9)	12 (5.6)	14 (7.1)	0	0	0
	74006254	rs1209244217	G/T	1 (0.6)	-	-	-	-	-
	74006487	rs137964758	C/T	11 (6.3)	12 (5.6)	14 (7.1)	0	0	0
	74006636	rs1426060	T/C	71 (40.8)	75 (34.7)	52 (26.3)	319 (31.7)	233 (23.1)	336 (34.4)
	74006822	rs149064938	C/T	8 (4.6)	6 (2.8)	11 (5.6)	0	0	0
	74007062	rs114898192	C/A	4 (2.3)	8 (3.7)	15 (7.6)	3 (0.3)	4 (0.4)	0
	74007094	rs1991262	C/G	37(21.3)	56(25.9)	59(29.8)	771(76.6)	866(85.9)	744(76.1)
	74007210	rs550479990	G/A	2 (1.1)	-	-	-	-	-
	74007245	rs73335828	G/A	33 (19.0)	26 (12.0)	24 (12.1)	48 (4.8)	14 (1.4)	43 (4.4)
	74007478	rs10504554	C/T	78 (44.8)	86 (39.8)	105 (53)	811 (80.6)	896 (88.9)	840 (85.9)
	74007514	rs16938758	T/A	84 (48.3)	101 (46.8)	110 (55.6)	817 (81.2)	895 (88.8)	840 (85.9)
	74007528	rs16938759	G/A	84 (48.3)	100 (46.3)	111 (56.1)	815 (81)	896 (88.9)	840 (85.9)
	74007592	rs189685235	T/A	2 (1.1)	0	2 (1.0)	0	0	0
	74007770	rs113548678	C/T	35 (20.1)	60 (27.8)	35 (17.7)	1 (0.1)	0	0
	74007806	rs146998002	C/A	6 (3.4)	4 (1.9)	11 (5.6)	3 (0.3)	3 (0.3)	0
	74007828	rs113459347	C/T	35 (20.1)	60 (27.8)	35 (17.7)	1 (0.1)	0	0
	74007831	rs141717659	C/T	6 (3.4)	4 (1.9)	11 (5.6)	3 (0.3)	3 (0.3)	0
	74007904	rs112793867	C/T	23 (13.2)	23 (10.6)	12 (6.1)	1 (0.1)	0	0
	74007975	rs28579835	G/C	2 (1.1)	13 (6.0)	9 (4.5)	0	0	0
	74008039	rs142840016	C/T	4 (2.3)	10 (4.6)	3 (1.5)	0	0	0
	74008103	rs16938760	G/A	42(24.1)	61(28.2)	71(35.9)	826(82.1)	864(85.7)	827(84.6)
	74008144	rs756135618	G/A	1 (0.6)	-	-	-	-	-
	74008339	rs73335834	C/T	31 (17.8)	43 (19.9)	27 (13.6)	1 (0.1)	0	0
	74008363	rs16938761	G/A	30(17.2)	54(25.0)	54(27.3)	818(81.3)	855(84.8)	814(83.2)
	74008410	rs535774727	G/T	1 (0.6)	1 (0.5)	0	0	0	0
	74008572	rs16938762	G/A	31 (17.8)	43 (19.9)	27 (13.6)	1 (0.1)	0	0
	74008619	rs145843246	G/C	6 (3.4)	4 (1.9)	11 (5.6)	3 (0.3)	3 (0.3)	0
	74008620	rs200051782	**AAT/A**	6 (3.4)	4 (1.9)	11 (5.6)	3 (0.3)	3 (0.3)	0
	74008621	rs879519398	**ATTCT/A**	6 (3.4)	-	-	-	-	-
	74008905	rs111614245	T/C	2 (1.1)	0	1 (0.5)	0	0	0
	74009186	rs142908478	G/A	3 (1.7)	4 (1.9)	11 (5.6)	3 (0.3)	3 (0.3)	0
	74009351	rs186800931	C/T	1 (0.6)	0	0	0	0	1 (0.1)
	74009397	rs1563714175	**A/AG**	8 (4.6)	-	-	-	-	-
	74009408	rs141329231	G/A	1 (0.6)	0	0	0	0	0
	74009413	rs148081046	G/A	1 (0.6)	0	0	0	0	0
	74009418	rs373694698	G/A	2 (1.1)	0	0	0	0	0
	74009547	rs1448213899	A/T	4 (2.3)	-	-	-	-	-

	74019877	rs147547464	A/G	5 (2.9)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74019902	rs6472813	G/C	36 (20.7)	69 (31.9)	64 (32.3)	822 (81.7)	867 (86.0)	826 (84.5)
	74019915	rs142925908	G/A	35 (20.1)	49 (22.7)	30 (15.2)	1 (0.1)	0	0
	74020084	rs7819708	A/G	36 (20.7)	69 (31.9)	64 (32.3)	822 (81.7)	867 (86.0)	826 (84.5)
	74020100	rs146658813	T/C	12 (6.9)	8 (3.7)	3 (1.5)	1 (0.1)	0	0
	74020206	rs146530931	A/G	23 (13.2)	23 (10.6)	12 (6.1)	1 (0.1)	0	0
	74020243	rs143140587	G/A	5 (2.9)	20 (9.3)	15 (7.6)	0	0	0
	74020579	rs560070947	C/T	2 (1.1)	0	0	1 (0.1)	0	0
	74021017	rs6993513	C/T	104 (59.8)	97 (44.9)	101 (51.0)	183 (18.2)	141 (14.0)	151 (15.4)
	74021169	rs541016211	C/T	3 (1.7)					
	74021344	rs113066143	A/G	73 (42.0)	71 (32.9)	76 (38.4)	147 (14.6)	109 (10.8)	97 (9.9)
	74021523	rs529510601	A/G	3 (1.7)	1 (0.5)	1 (0.5)	0	0	0
	74021723	rs191612377	A/G	1 (0.6)	1 (0.5)	5 (2.5)	0	0	0
	74021806	rs180826310	C/A	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74022021	rs183365413	G/A	6 (3.4)	3 (1.4)	0	0	0	0
	74022051	rs143055276	C/A	1 (0.6)	0	0	4 (0.4)	0	1 (0.1)
	74022102	rs55917243	G/A	73 (42.0)	71 (32.9)	77 (38.9)	147 (14.6)	109 (10.8)	97 (9.9)
	74022145	rs11998059	A/C	29 (16.7)	50 (23.1)	33 (16.7)	1 (0.1)	0	0
	74022301	rs77754189	C/G	73 (42.0)	71 (32.9)	78 (39.4)	147 (14.6)	110 (10.9)	97 (9.9)
	74022303	rs115331465	G/A	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74022422	rs71733528	**C/CA**	29 (16.7)	-	-	-	-	-
	74022597	rs74940078	G/T	19 (10.9)	-	-	-	-	-
	74022599	rs1190589339	G/T	2 (1.1)	-	-	-	-	-
	74022639	rs1402990689	C/T	3 (1.7)	-	-	-	-	-
	74022648	rs28513078	C/A	31(17.8)	62(28.7)	57(28.8)	813(80.8)	859(85.2)	809(82.7)
	74022649	rs187932458	G/A	27 (15.5)	16 (7.4)	20 (10.1)	36 (3.6)	22 (2.2)	37 (3.8)
	74022689	rs552270680	C/A	3 (1.7)	1 (0.5)	1 (0.5)	0	0	0
	74022727	rs28667879	T/C	29 (16.7)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
Intron 3	74022924	rs557953887	A/G	2 (1.1)	0	0	0	0	0
	74023074	rs28496706	G/C	29 (16.7)	50 (23.1)	33 (16.7)	1 (0.1)	0	0
	74023173	rs75276516	G/T	6 (3.4)	20 (9.3)	15 (7.6)	0	0	0
	74023329	rs528880274	C/T	2 (1.1)	0	0	0	0	0
	74023376	rs7838114	G/A	36 (20.7)	69 (31.9)	64 (32.3)	822 (81.7)	867 (86.0)	826 (84.5)
	74023438	rs75913616	T/G	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74023464	rs7838017	C/T	106 (60.9)	97 (44.9)	101 (51.0)	183 (18.2)	141 (14.0)	152 (15.5)
	74023573	rs75084868	C/T	4 (2.3)	10 (4.6)	4 (2.0)	4 (0.4)	10 (1.0)	22 (2.2)
	74023854	rs114128949	T/C	1 (0.6)	4 (1.9)	2 (1.0)	0	0	0
	74024292	rs1037493201	T/C	3 (1.7)	-	-	-	-	-
	74024457	rs112541809	G/A	23 (13.2)	23 (10.6)	13 (6.6)	1 (0.1)	0	0
	74024593	rs80343839	C/T	26 (14.9)	16 (7.4)	20 (10.1)	36 (3.6)	22 (2.2)	38 (3.9)
	74024594	rs56230757	G/A	73 (42.0)	71 (32.9)	77 (38.9)	147 (14.6)	109 (10.8)	98 (10.0)
	74024640	rs73335860	T/A	27 (15.5)	16 (7.4)	20 (10.1)	36 (3.6)	22 (2.2)	37 (3.8)
	74024705	rs115865211	A/C	14 (8.0)	14 (6.5)	9 (4.5)	0	0	0
	74024733	rs559630662	T/G	3 (1.7)	-	-	-	-	-
	74024829	rs73335866	A/G	23 (13.2)	23 (10.6)	12 (6.1)	1 (0.1)	0	0
	74025203	rs73335869	G/C	23 (13.2)	18 (8.3)	11 (5.6)	1 (0.1)	0	0
	74025388	rs111233473	T/C	29 (16.7)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
	74025733	rs147904637	C/T	1 (0.6)	3 (1.4)	1 (0.5)	0	0	0
	74025782	rs189419975	C/A	2 (1.1)	3 (1.4)	2 (1.0)	0	0	0
	74025964	rs141504244	C/T	5 (2.9)	5 (2.3)	2 (1.0)	0	0	0
	74025977	rs10106026	C/T	29 (16.7)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
	74026100	rs148158912	A/T	3 (1.7)	0	10 (5.1)	0	0	0
	74026145	rs112264636	**C/CTATT**	109 (62.6)	147 (68.1)	134 (67.7)	184 (18.3)	141 (14)	152 (15.5)
	74026148	rs200345119	**TA/T**	29 (16.7)	-	-	-	-	-
	74026152	rs879272932	**A/ATATT**	29 (16.7)	-	-	-	-	-
	74026466	rs201476283	**C/CA**	6 (3.4)	-	-	-	-	-
	74026468	rs199501702	**AG/A**	6 (3.4)	17 (7.9)	15 (7.6)	0	0	0
	74026471	rs180939394	T/C	6 (3.4)	19 (8.8)	15 (7.6)	0	0	0
	74026940	rs11379988	**C/CT**	7 (4.0)	-	-	-	-	-
	74027018	rs573220510	C/T	3 (1.7)	0	0	0	0	0
	74027024	rs183250931	G/C	1 (0.6)	2 (0.9)	1 (0.5)	0	0	0
	74027146	rs1179055086	C/T	1 (0.6)	-	-	-	-	-
	74027439	rs138887132	G/A	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)
	74027829	rs60745123	A/C	34 (19.5)	50 (23.1)	32 (16.2)	1 (0.1)	0	0
	74027865	rs191067849	G/A	1 (0.6)	2 (0.9)	0	0	0	0
	74027953	rs116539549	T/C	4 (2.3)	3 (1.4)	1 (0.5)	0	0	0
	74027996	rs565907234	G/A	1 (0.6)	-	-	-	-	-
Intron 4	74028276	rs7820086	T/C	73 (42)	71 (32.9)	77 (38.9)	147 (14.6)	109 (10.8)	97 (9.9)
	74028312	rs182486475	G/A	3 (1.7)	0	0	0	0	0
	74028442	rs544252525	A/G	2 (1.1)	-	-	-	-	-
	74028472	rs75218174	T/C	2 (1.1)	10 (4.6)	4 (2.0)	0	0	0
	74028613	rs146198500	G/T	3 (1.7)	0	4 (2.0)	0	0	0
	74028651	rs2114169	T/A	36 (20.7)	69 (31.9)	64 (32.3)	822 (81.7)	866 (85.9)	826 (84.5)
	74028658	rs193263455	C/A	1 (0.6)	1 (0.5)	0	0	0	0
	74028771	rs2891354	C/T	37 (21.3)	69 (31.9)	64 (32.3)	822 (81.7)	867 (86)	826 (84.5)
	74028793	*LY96* NI-10	G/A	1 (0.6)	-	-	-	-	-
	74028948	rs11466003	C/T	4 (2.3)	10 (4.6)	4 (2.0)	0	10 (1.0)	18 (1.8)

Wt/mut: wildtype/mutant referring to major and minor alleles, respectively; **SA:** South Africa; **YRI:** Yoruba in Ibadan, Nigeria; **LWK:** Luhya from Webuye, Kenya; **EUR:** European population; **EAS:** East Asians population; **SAS:** South Asian population; Grey shaded boxes represent significant comparisons to the South African population (SA); **-:** No non reference calls reported for this variant in 1000 genomes (assumption that they are monomorphic). NI: Newly identified variant. Indels are indicated with bold base changes.

APPENDIX 2

Ethical clearance

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Professor CT Tiemessen, et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M190995

NAME: (Principal Investigator)	Professor CT Tiemessen, et al
DEPARTMENT:	National Institute for Communicable Diseases Centre for HIV and Sexually-Transmitted Infections Sandringham
PROJECT TITLE:	HIV-1 positive South African Elite and Long-term Controllers: viral and host targets for functional cure strategies
DATE CONSIDERED:	Ad hoc
DECISION:	Aproved unconditionally
CONDITIONS:	Renewal of M140926
SUPERVISOR:	Not applicable
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2019/10/04

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the 3rd Floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to submit details to the Committee. I **agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **September** and will therefore reports and re-certification will be due early in the month of **September** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

5 October 2019
Date

PLEASE QUOTE THE CLEARANCE CERTIFICATE NUMBER IN ALL ENQUIRIES