

https://dspace.org/

# The role of CCL4/Mip-1b in HIV infection

**Bharuthram, Avani**

**2015**
http://hdl.handle.net/10539/19971

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

# The Role of CCL4/Mip-1β in HIV infection

**Avani Bharuthram**

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree Master of Science.

Johannesburg 2015

# DECLARATION

I, Avani Bharuthram, hereby declare that:

a) The research reported in this thesis, except where otherwise indicated is my original work.

b) This thesis has not been submitted for any degree or examination to any other University

c) The research reported in this dissertation, except where otherwise indicated is my original work.

d) This dissertation has not been submitted for any degree or examination to any other University

e) This dissertation does not contain other people's data, pictures, graphs or information, unless specifically stated as being of sourced from other people.

f) This dissertation does not contain another person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   I. Their words have been re-written but the general information attributed to them has been referenced;

   II. Where their exact words have been used, their writing has been placed inside quotation marks and referenced.

g) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was written by myself alone and have fully referenced such publications.

h) This dissertation does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being listed in the dissertation and in the References section.

_____ _____

Avani Bharuthram Date

# ABSTRACT

The controversy surrounding the findings that copy number variation (CNV), of the CCL3 encoding genes, influences HIV-1 infection and disease progression has been in part attributed to the variable results obtained from methods used for copy number evaluation. Like CCL3, the genes encoding the CC chemokine CCL4, also a natural ligand of the CCR5 receptor, are found to occur in population-specific multiple copy number and have been shown to play a protective role against HIV-1. In this study, we evaluated the standard method of quantitative Real-Time PCR (qPCR) and Droplet Digital PCR (ddPCR) for *CCL4L* gene copy number determination. The CCL4 encoding genes are *CCL4*, occurring in two copies per diploid genome (pdg), and the non-allelic *CCL4L* genes, comprised of *CCL4L1* and *CCL4L2*, which are both found in multiple copies pdg. Copy number of *CCL4L*, *CCL4L1* and *CCL4L2* was determined in a cohort of HIV-1-uninfected individuals from the South African Black (n=23) and Caucasian (n=32) population groups using qPCR and ddPCR, with the addition of another 30 black individuals to the ddPCR cohort. A stronger correlation between the number of *CCL4L* copies and the sum of *CCL4L1* and *CCL4L2* copies generated by ddPCR ($r=0.99$, $p<0.0001$) compared to qPCR ($r=0.87$, $p<0.0001$) was observed. Real-Time qPCR exhibited greater inaccuracy at higher copy numbers which is particularly relevant to our cohort of Black individuals who have a higher range of *CCL4L* copies (3-6) compared to Caucasians (0-4) and a higher population median (4 and 2, respectively). Medians and ranges of *CCL4L1* (Black: 2, 0-4, Caucasian: 0, 0-2) and *CCL4L2* (Black: 2, 1-5, Caucasian: 2, 0-3) were also higher in the Black population. Droplet Digital PCR was shown to be a far superior method to qPCR for assessment of *CCL4* gene copy number variation, the accuracy of which is essential for studies of the contribution of variable gene copy number to phenotypic outcomes of host infection and disease course. We further used the *CCL4L* copy number data to examine variation of these genes with

reference to measured stimulated CCL4 production by the same individuals. Although significant differences in the copy number range medians and patterns of distribution of the genes *CCL4L1*, *CCL4L2* and combined *CCL4L* were observed between the two populations, on a whole, the two populations do not differ significantly with respect to CCL4 production. Caucasian females however had a higher level of protein production per copy of *CCL4L* than Black females. When stratifying production based on population specific copy number median, Black individuals showed a decreased level of protein production at a *CCL4L* copy number below the median, although this was not maintained when the *CCL4L1* and *CCL4L2* genes were analysed individually. CCL4 copy number and production data was compared to data generated for CCL3 in the same cohort. CCL4 chemokine levels were significantly higher in both the Black and Caucasian populations and Black individuals had a higher number of gene copies of the *CCL3L* genes giving rise to functional CCL3 protein than the *CCL4L* genes producing functional CCL4 protein. These results highlight genetic differences between divergent populations, differences in distribution of CCL4 and CCL3 encoding genes and protein production, and suggest an intricate regulation of the *CCL4L* encoding genes. In order to investigate the role of CCL4 in HIV-1 control, we next assessed variation in the numbers of *CCL4L* copies in relation to CCL4 production in a cohort of 14 long-term nonprogressors (LTNPs). While no associations between copy number and CCL4 production were observed, the LTNP cohort had significantly lower levels of CCL4 production than the HIV-1-uninfected cohort, and this was maintained when Black and Caucasian individuals were examined individually. When the LTNP cohort individuals were divided based on viral loads, individuals with viral loads <400 RNA copies/ml had significantly lower CCL4 production than those with viral loads >400 RNA copies/ml. This finding suggests a role for the amount of CCL4 produced in the reduced pace of HIV-1 progression observed in LTNP individuals. Since genetic variation other than copy number can also influence CCL4 protein production, we then proceeded with a thorough genetic characterization of the *CCL4* gene of uninfected individuals and investigated any possible relationships with select variants and CCL4 production. Sequencing the complete gene

and flanking regions of 23 Black and 32 Caucasian revealed several intra as well as extragenic SNPs, with one newly identified SNP at position -1063. The Black population exhibited a higher degree of variability compared to the Caucasian population. We described four haplotypes in the Caucasian population, three haplotypes in the Black population, and a single haplotype shared between both population groups. Of the four indels that were identified, a three bp deletion (rs3216921) was the only indel present in the Caucasian population and was not identified in the Black population. Of all the indels, this indel had the highest allelic frequency (14%). Comparisons of haplotypes and prevalent SNPs with protein production in both population groups did not show any significant differences. Caucasian individuals harbouring the 3bp deletion however, had significantly higher levels of CCL4 production ($p=0.024$). These results form a good base from which to further investigate the impact of select genetic variants on CCL4 production and possibly HIV-1 control. This study has succeeded in optimising a ddPCR assay for the copy number determination of the CCL4 genes and has interrogated the relationship between *CCL4L* copy number and CCL4 production in both HIV-1-uninfected individuals as well as a subset of LTNPs. The results suggest a complex, intricate regulation of CCL4 that appears to play a role in HIV-1 control. In conclusion, this study forms the basis for future work to build on and to further explore the role of this CCR5 ligand in HIV-1 disease.

# ACKNOWLEDGEMENTS

This study was made possible by funding from the Poliomyelitus Research Foundation, The National Health Laboratory Service Research Trust and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa. I would like to acknowledge the true heroes of this study, the volunteers who selflessly put themselves through a bit of discomfort to help answer the questions that we had.

To my supervisors, Prof Caroline Tiemessen and Dr Maria Paximadis: Thank you for leading me through this process. It is obvious that you are both strong and immensely intelligent women, but what is most striking is the loving and caring way in which you deal with your students. Caroline, I admire the way that you step in when students need rescuing. Maria, you have been like the North Star, always steering me back on course, although it has sometimes taken me a long time to look through the clouds. Thank you to Dr Anabela Picton for teaching me so many techniques and data analysis methods, for your positive attitude and for making me feel like a regular sized human being in a world filled with short people. Humour aside, I am sincerely grateful to all of you for the many things that I have learnt and for the time that you have invested in helping me through this degree.

To the soon to be Doctors, Ria, Nikki and Heather, my office mates and a lovely bunch of girls, thank you for the laughs, the interesting debates (Sharks vs. Spiders), the empathetic 'I feel your pain', and for being happy when I have achieved something. I am going to miss you all like R7 life-sustaining toasted cheese sandwiches. Thank you to Shayne and Gemma, my vegan buddies, I think we've done a good job in convincing the rest of the lab that vegetarian food is actually quite nice. Thank you to

the super sportswomen, Di and Sharon and to Leonard, the man in the corner who turned out to be quite chatty. A friendly smile from Thuli, who is likely to win a jackpot and buy the NHLS someday, was a great way to start a day. And an acknowledgement to the enigma that is Mihloti, for providing many thought provoking moments. These individuals and the many others that have passed through our lab, even if only for a short while have taught me a variety of things. When the majority of your day is spent in one environment, it is comforting to have people like the ones mentioned above surrounding you.

Thank you to my friends and family members, old and new, my 'soul buddies', Bhaskaran, Heena, my musical family, my Durban connections (many of whom are scattered around the world; and in Ermelo), and a special thank you to my uncle Pravesh.

To my favourite parents, Ramesh and Sharita, I greatly appreciate and value your motivating talks, support, and the strong protests every time I announced, 'I'm a loser,' after many tough days. I am also grateful to my little sister Nirvana, who thinks that I'm the coolest (jokes), for having to take on the role of a big sister at times and for trying to wake me up at a reasonable time every morning. Without the support from my loving family, I would not be submitting this dissertation.

Thank you to my late grandparents, my Nani and my Nana. The vivid memories of their love and soothing words, still give me comfort today. During tough times, these memories of them consoled me more than anybody on this earth could have.

The grace that has brought me to this point has not gone unnoticed.

# TABLE OF CONTENTS

Declaration ..... i

Abstract ..... ii

Acknowledgements ..... v

Table of Contents ..... vii

List of Tables ..... xi

List of Figures ..... xii

Abbreviations ..... xv

Preface ..... xviii

Chapter 1 Introduction ..... 1

1.1. HIV treatment: successes and challenges ..... 5

1.2. Host immune factors and genetic variability ..... 6

1.3. HIV-1 Viral Entry and the C-C chemokine receptor-5 (CCR5) ..... 10

1.4. CC Chemokines ..... 11

1.5. Genetic structure of the CCL3-CCL4 region of chromosome 17 ..... 14

1.6. Contribution of CCL3 and CCL4 gene copy number and genetic variants to HIV-1 disease .. 16

1.7. Rationale for current study ..... 19

1.8. Study objectives ..... 21

Chapter 2 Materials and Methods ..... 22

2.1. Cohort Description ..... 23

2.1.1. HIV-1-uninfected individuals ..... 23

2.1.2. Additional samples for CCL4L copy number determination ..... 23

2.1.3. Long-term nonprogressor (LTNP) cohort ..... 23

2.2. Peripheral Blood Mononuclear Cell (PBMC) Isolation and Culture ..... 25

2.3. DNA Extraction ..... 26

2.4. Gene Nomenclature ..... 27
2.4.1. CCL4L nomenclature ..... 27
2.4.2. CCL3L nomenclature ..... 28
2.5. Copy Number Determination Assays ..... 29
2.5.1. TaqMan primers and probes ..... 29
2.5.2. Evaluation of gene copy numbers by quantitative Real-Time PCR ..... 29
2.5.3. Evaluation of gene copy number using the Droplet Digital PCR system ..... 30
2.5.4. PCR amplification and sequencing of CCL4L of select individuals ..... 32
2.6. CCL4 Production Analysis ..... 33
2.7. Comparison of CCL4 and CCL3 production and Copy Number Variation ..... 33
2.8. Genetic Characterisation of the *CCL4* Gene ..... 34
2.8.1. PCR amplification of CCL4 ..... 34
2.8.2. Sequencing ..... 35
2.8.3. Sequencing analysis ..... 36
2.8.4. Assigning the wild type (WT) allele ..... 36
2.8.5. Haplotype prediction/analysis ..... 36
2.8.6. Characterization of putative insertions/deletions ..... 37
2.8.7. Hardy-Weinberg equilibrium ..... 37
2.8.8. Bioinformatic tools used ..... 37
2.9. Statistical Analysis ..... 37
Chapter 3 Results ..... 39
3.1. Comparison of qPCR and ddPCR for *CCL4L* Copy Number Determination ..... 40
3.1.1. Assay validation ..... 40
3.1.2. Reproducibility of results ..... 44
3.1.3. Population specific medians ..... 46
3.2. *CCL4L* Copy Number Analysis ..... 47
3.2.1. Gender and copy number distribution ..... 47

3.2.2. CCL4L copy number variation between the Black and Caucasian population groups ..... 48
3.3. CCL4 Protein Production and correlation to copy number ..... 50
3.3.1. Population and gender differences in CCL4 production ..... 50
3.3.2. Correlation of CCL4 production with CCL4L copy number ..... 52
3.3.3. CCL4 production per CCL4L copy number ..... 55
3.3.4. Comparison of CCL3L and CCL4L gene copy number and CCL3 and CCL4 production ..... 57
i. CCL3L and CCL4L copy number distribution ..... 57
ii. CCL3 and CCL4 protein production ..... 59
3.4. CCL4 and HIV-1-infected long-term nonprogressors (LTNPs) ..... 61
3.4.1. CCL4 copy number and production comparisons between uninfected Black individuals and Black LTNPs ..... 61
3.4.2. CCL4 production in LTNPS differs from HIV-1-uninfected individuals ..... 62
3.4.3. CCL4 protein production is not higher than CCL3 protein production in LTNPs ..... 63
3.4.4. CCL4 and CCL3 production and HIV-1 viral load (VL) ..... 65
3.5. Genetic Characterization of the *CCL4* Gene ..... 67
3.5.1. Single nucleotide polymorphisms ..... 67
3.5.2. Influence of exonic SNPs on protein translation ..... 70
3.5.3. Insertions/Deletions ..... 72
3.5.4. Putative haplotypic structure of CCL4 and flanking regions ..... 74
3.5.5. Hardy-Weinberg Equilibrium ..... 76
3.5.6. The influence of Genetic Variation on CCL4 production ..... 80
Chapter 4 Discussion and Conclusions ..... 82
References ..... 99
Appendix A: Supplementary Data for Chapter 2 ..... 107
Appendix B: Supplementary Data for Chapter 3 ..... 108
Appendix C: Amino Acid Codes ..... 109
Appendix D: Ethical Clearance ..... 110

Appendix E: Turn It in Report.....111

# LIST OF TABLES

Table 2.1.1 Characteristics of HIV-1-infected long-term nonprogressor study participants ..... 24

Table 2.8.1 Primer pairs used for the amplification of the *CCL4* gene using the Expand High Fidelity PCR System ..... 35

Table 3.1.1 Population specific medians and ranges generated by qPCR and ddPCR for the South African Black and Caucasian populations ..... 42

Table 3.5.1 SNPs and indels identified in the South African Black and Caucasian populations ..... 69

# LIST OF FIGURES

Figure 1.1.1 Infographic of HIV statistics (A) Reduction in new HIV infections in adults and children since 2001. (B) 2012 HIV statistics ..... 4

Figure 1.2.1 Schematic diagram of factors to be considered when assessing vaccine trial data (Tomaras and Haynes 2014). ..... 7

Figure 1.3.1 Viral entry of HIV-1 into CD4+ cells. HIV-1 first attaches to cells (a) then binds to the CD4 receptor via the trimeric gp120 subunit of the Env protein (b).. ..... 10

Figure 1.4.1 Schematic description of the structure of the different families of chemokines and the receptors and ligands which belong to the family..... 12

Figure 1.5.1 Schematic representation of the CCL4 encoding genes and transcripts. ..... 16

Figure 2.4.1 Schematic representation of the relative position, order and nomenclature of the CCL3 and CCL4 encoding genes on chromosome 17q12. ..... 28

Figure 3.1.1 Concordance between *CCL4L* copy number and the sum of the copy number of *CCL4L1* and *CCL4L2* generated by two assays ..... 41

Figure 3.1.2 Clustering of unrounded average copy number values for all three *CCL4L* genes generated by quantitative Real-Time PCR and Droplet Digital PCR around whole integer copy numbers. ..... 44

Figure 3.1.3 Copy number frequencies of the sum of genes *CCL4L1* and *CCL4L2* ..... 45

Figure 3.2.1 Gender differences in copy number distribution between Black South African females and males for the (A) *CCL4L* genes, (B) *CCL4L1* and (C) *CCL4L2* individually. ..... 47

Figure 3.2.2 Copy number variation of the *CCL4L* genes in the South African Black and Caucasian populations ..... 49

Figure 3.2.3 Differences in distribution and frequency of *CCL4L* copy number and *CCL4L1* and *CCL4L2* copy number combinations in HIV-1-uninfected (A) Black (n=53) and (B) Caucasian (N=32) South African cohorts..... 49

Figure 3.2.4 Inverse significant correlation between *CCL4L1* and *CCL4L2* copy numbers in (A) Black but not in (B) Caucasian individuals ..... 50

Figure 3.3.1 No Significant differences in CCL4 protein production in two South African populations but Black females have a higher PHA-stimulated CCL4 production than Caucasian females.. ..... 51

Figure 3.3.2 Correlations of PHA-stimulated CCL4 production to (i) *CCL4L,* (ii) *CCL4L1* and (ii) *CCL4L2* copy number variants in (A) Black and (B) Caucasian South African cohorts..... 53

Figure 3.3.3 PHA-stimulated CCL4 production relative to population specific median copy numbers in two South African populations ..... 54

Figure 3.3.4 PHA-stimulated CCL4 production per gene copy number of the *CCL4L* encoding genes in the Black and Caucasian South African population ..... 56

Figure 3.3.5 Comparison of PHA-stimulated CCL4 production per *CCL4L* gene copies in Black and Caucasian individuals having equal number of copies of *CCL4L* and comparable patterns of distribution of *CCL4L1* and *CCL4L2*. ..... 56

Figure 3.3.6 Correlation of *CCL3L* and *CCL4L* copy numbers generated in individuals of the same cohorts of South African (A) Black and (B) Caucasian individuals. ..... 58

Figure 3.3.7 Comparison of the copy number distribution of the *CCL4L2* and *CCL3La* genes (functional isoforms of the CCL4 and CCL3 genes, respectively) in a South African (A) Black and (B) Caucasian cohort. ..... 58

Figure 3.3.8 Correlations of CCL4 production to CCL3 production in a cohort of (A) 55 HIV-1-uninfected individuals comprising of (B) Black and (C) Caucasian cohorts. ..... 59

Figure 3.3.9 CCL3 and CCL4 protein production measured from (A) unstimulated PBMC supernatants and (B) PHA-stimulated supernatants, were compared in the cohort of (i) 55 HIV-1-uninfeced South Africans, comprising of (i) 23 Black and (iii) 32 Caucasian individuals. ..... 60

Figure 3.4.1 Neither (A) *CCL4L*, (B) *CCL4L1* nor (C) *CCL4L2* copy number distribution differed significantly when a cohort of HIV-1-uninfected Black individuals was compared to a cohort of LTNPs. ..... 61

Figure 3.4.2 *CCL4L* gene copy numbers do not correlate with CCL4 production in a cohort of 9 Black South African LTNPs. ..... 62

Figure 3.4.3 CCL4 production from (A) unstimulated and (B) PHA-stimulated peripheral PBMCs isolated from HIV-1-uninfected individuals and LTNPs ..... 63

Figure 3.4.4 No significant differences in (A) unstimulated and (B) PHA-stimulated CCL3 and CCL4 production from peripheral PBMCs. P values and number of individuals are indicated. ..... 64

Figure 3.4.5 Correlation of PHA-stimulated CCL4 and CCL3 production in a group of HIV-1-infected LTNPs ..... 65

Figure 3.4.6 Comparisons of (i) Unstimulated and (ii) PHA-stimulated (A) CCL3 and (B) CCL4 production between HIV-uninfected Black South Africans and HIV-infected Black LTNPs ..... 66

Figure 3.5.1 (A) Schematic representation of the CCL4 gene ..... 69

Figure 3.5.2 Influence of CCL4 exonic SNPs on CCL4 amino acid sequence. ..... 71

Figure 3.5.3 Schematic representation of putative CCL4 haplotypes identified in the South African Black and Caucasian populations. ..... 73

Figure 3.5.4 Pairwise LD among SNPs (>5% prevalence) across the *CCL4* gene in South African Black individuals. ..... 77

Figure 3.5.5 Pairwise LD among SNPs (>5% prevalence) across the *CCL4* gene in South African Black individuals ..... 78

Figure 3.5.6 Pairwise LD among SNPs (>5% prevalence) across the predicted *CCL4* haplotypes in South African Black individuals ..... 79

Figure 3.5.7 Comparison of CCL4 production in Caucasian individuals with and without the rs3216921 deletion in the CCL4 gene ..... 81

# ABBREVIATIONS

| | |
|---|---|
| *A* | *Adenine* |
| *Ab* | *Antibody* |
| *ADCC* | *Antibody-dependent cellular cytotoxicity* |
| *AIDS* | *Acquired Immunodeficiency Syndrome* |
| *BGB* | *Human beta globin* |
| *bp* | *base pair* |
| *°C* | *Degrees Celsius* |
| *C* | *Cytosine* |
| *Cd* | *Cluster of differentiation* |
| *CCL3* | *Chemokine (C-C motif) ligand 3 or MIP-1α* |
| *CCL3L* | *C-C motif chemokine ligand 3-like* |
| *CCL4* | *Chemokine (C-C motif) ligand 4 or MIP-1β* |
| *CCL4L* | *C-C motif chemokine ligand 4-like* |
| *CCL5* | *Chemokine (C-C motif) ligand 5 or RANTES* |
| *CCL8* | *Chemokine (C-C motif) ligand 8 or MCP-2* |
| *CCR* | *C-C chemokine receptor* |
| *CCR5* | *C-C chemokine receptor 5* |
| *CCR5Δ32* | *CCR5 delta 32 (32 bp deletion)* |
| *CD* | *Cluster of differentiation* |
| *CT* | *Cycle threshold* |
| *CXCR4* | *CXC-chemokine receptor 4* |
| *ddPCR* | *Droplet Digital PCR* |
| *DNA* | *Deoxyribonucleic acid* |
| *EC* | *Elite controller* |
| *EDTA* | *Ethylenediaminetetraacetic acid* |
| *eg.* | *Example* |
| *Env* | *HIV viral envelope glycoprotein* |
| *et al.* | *And others* |
| *g* | *gram* |
| *G* | *Guanine* |

| | |
|---|---|
| *h* | *Hours* |
| *HAART* | *Highly active antiretroviral therapy* |
| *HESN* | *Highly exposed seronegative* |
| *HIV-1* | *Human Immunodeficiency Virus type 1* |
| *HLA* | *Human Leukocyte Antigens* |
| *kb* | *Kilo bases* |
| *kDa* | *Kilo Daltons* |
| *LTNP* | *Long-term nonprogressor* |
| *M* | *Molar* |
| *MDG* | *Millennium Development Goals* |
| *MCP-2* | *Monocytic chemotactic protein-2 or CCL8* |
| *MGB* | *Minor groove binder* |
| *Min* | *Minutes* |
| *MHC* | *Major histocompatibility complex* |
| *MIP-1α* | *Macrophage inflammatory protein-1α or CCL3* |
| *MIP-1β* | *Macrophage inflammatory protein-1β or CCL4* |
| *ml* | *Millilitre* |
| *mRNA* | *Messenger RNA* |
| *MTCT* | *Mother-to-child transmission* |
| *n* | *Number of study participants* |
| *NI* | *Newly Identified* |
| *NK* | *Natural killer* |
| *nM* | *nanomolar* |
| *PBMC* | *Peripheral blood mononuclear cells* |
| *PBS* | *Phosphate Buffered Saline* |
| *PCR* | *Polymerase Chain Reaction* |
| *pdg* | *Per diploid genome* |
| *pg* | *Picogram* |
| *PHA* | *Phytohaemagglutinin* |
| *PRT* | *Paralogue ratio test* |
| *qPCR* | *Quantitative real-time polymerase chain reaction* |
| *RANTES* | *Regulated upon Activation, Normal T-cell Expressed, and Secreted or CCL5* |
| *RNA* | *Ribonucleic acid* |
| *RPMI* | *Roswell Park Memorial Institute medium* |

| | |
|---|---|
| *SIV* | *Simian Immunodeficiency Virus* |
| *SNP* | *Single nucleotide polymorphism* |
| *T* | *Thymine* |
| *T* | *Temperature* |
| *UTR* | *Untranslated region* |
| *vs* | *Versus* |
| *WT* | *Wild type* |

# PREFACE

Part of this work has been published:

Bharuthram A, Paximadis M, Picton AC, Tiemessen CT. 2014. Comparison of a quantitative Real-Time PCR assay and droplet digital PCR for copy number analysis of the CCL4L genes. Infection Genetics and Evolution (**25**).28-35.

# CHAPTER 1 INTRODUCTION

The first reported cases of Acquired Immune Deficiency Syndrome (AIDS), then an unidentified disease, were in 1981 (Gottlieb 1981; Gottlieb et al. 1981; Siegal et al. 1981). The infectious agent of the disease was identified as a retrovirus (Barre-Sinoussi et al. 1983; Gallo et al. 1984; Popovic et al. 1984) which was later named Human Immunodeficiency Virus (HIV) (Coffin et al. 1986). Since then, the disease has become a global health pandemic and as early as 1987, the World Health Organization had set up a programme to probe the spread of the disease and possible treatments and strategies to manage it (Mann 1987). By pledging to work towards the Millennium Development Goals (MDG), compiled in 2000, the United Nations member states and several other organizations committed to working towards eight goals of international significance, one of them being to combat diseases such as HIV, and in 2001, the expected outcomes that were to be met with regards to improving the global scenario of HIV were clearly defined (UNAIDS 2013c).

As the target date (2015) to reach the MDGs approaches, advances in spheres such as decreasing the number of new infections and increasing the number of individuals with access to treatment have been positive (Figure 1.1.1A and B). However, there are still an estimated 35.5 million people living with HIV worldwide, and 2.3 million new infections in adults and 260 thousand new infections in infants were reported in 2012 (Figure 1.1.1B) . The largest disease burden is borne by countries with low socio-economic climates, with Sub-Saharan Africa being the most affected (Figure 1.1.1C). South Africa in particular, has the highest infection rate and HIV prevalence in the world (Figure 1.1.1D and E).

In 2009, for example, it was reported that while the rate of mother to child transmission of HIV was as low as 1% in developed countries, in some resource-poor African countries, the rate was as high as 50% (Paintsil and Andiman 2009). While implementation of administration of treatment to HIV-infected mothers can prevent mother-to-child transmission (MTCT), study of these mother and infant

pairs before the current recommended drug regimen have helped to elucidate possible host immune factors that contribute to the exposed uninfected phenotype which may be applied to studies investigating protection mechanisms seen in sero-discordant couples.

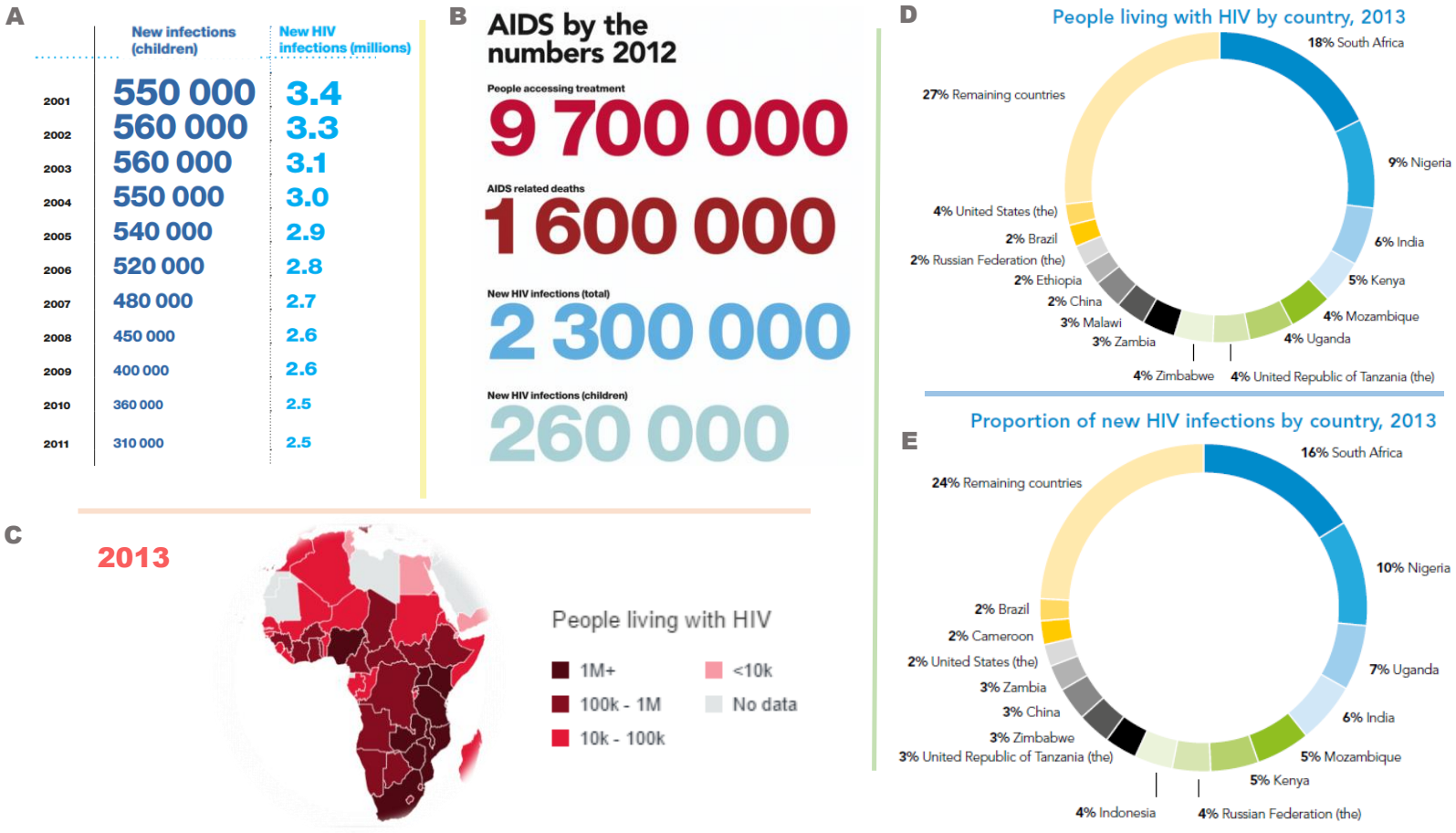



**Figure 1.1.1 Infographic of HIV statistics** (A) Reduction in new HIV infections in adults and children since 2001. (B) 2012 HIV statistics. (C) Heat map of people living with HIV in sub-Saharan Africa in 2013. (D) Percentage of individuals living with HIV and (E) proportion of HIV infections showing the contribution of individual countries to the total in 2013. (UNAIDS 2013a; UNAIDS 2013b; UNAIDS 2014).

## 1.1. **HIV TREATMENT: SUCCESSES AND CHALLENGES**

Highly Active Antiretroviral Therapy (HAART), a combination therapy of protease inhibitors, entry inhibitors, nucleoside and non-nucleoside reverse transcriptase inhibitors and integrase inhibitors (AIDSinfo 2014) has been a medical triumph as it has significantly increased the expected life-span of HIV-infected individuals as well as decreased the number of AIDS related deaths (Palella et al. 1998; Walensky et al. 2006). Even though there were some challenges in initiation of the roll out of HAART in South Africa, the therapy has been a success in both adults and children who are provided with the therapy, despite their low resource surroundings (Coetzee et al. 2004; Reddi et al. 2007).

HAART has been shown to not only effectively prevent free virus transmission to cells, but cell to cell transmission as well (Agosto et al. 2014). Although one MDG was to provide treatment for 15 million HIV-infected individuals in 2015, by 2012, this quota was only met by 65% (Figure 1.1.1B). HAART, however is not a cure for HIV infection and has to be adhered to for a patient's lifetime, as periods of non-adherence can result in increases in viral loads as viral reservoirs and latent viruses are not cleared from cells. Furthermore, the treatment has driven the evolution of drug resistant viral strains, is beyond the reach of many individuals in poor countries, and can lead to mild toxicities in patients, which might be a factor contributing to the increased occurrence of non-communicable diseases such as diabetes and liver disease, observed in HIV positive individuals (Richman et al. 2009).

## 1.2. HOST IMMUNE FACTORS AND GENETIC VARIABILITY

The development of prophylactic therapies such as vaccines, have taken several different approaches. The difficulties faced in inducing neutralizing antibodies to HIV-1 and the variation observed within HIV envelope proteins motivated the change to a vaccine design that would induce a cellular response (Sekaly 2008). The STEP trial, which followed this design, however, failed to protect against HIV-1 infection and in some cases exacerbated infection rates (Gray et al. 2010). The outcomes of the trial were in part attributed to previous immunity of the study participant to the adenovirus vector that was used (Gray et al. 2010). Furthermore, the macaque study, which prompted the testing of the STEP regimen in humans, revealed that protection in those animals may have been associated with specific Human Leukocyte Antigen (HLA) genotypes (Sekaly 2008). The most successful HIV vaccine trial to date, the RV144 Thai trial, used a combination of two vaccines, a canarypox vector-based vaccine harbouring genetically engineered versions of three HIV-1 genes (*env* (envelope glycoprotein) , *gag* (Group Specific Antigen) and *pol* (polymerase)) and a second vaccine composed of genetically engineered gp120, which is an HIV-1 envelope protein subunit. This trial was estimated to offer protection against HIV-1 infection to 31% of study participants compared those who received the placebo; however the vaccine had no effect on viral set point in the vaccinated individuals that became infected (Donald and McNeil 2009). An interesting outcome of the trial however was that correlates of protection that were identified were not neutralising antibodies or cellular immunity, but likely to have included antibody-dependent cellular cytotoxicity (ADCC) antibodies, that is, antibodies that classically form part of an innate immune response (Donald and McNeil 2009). These examples all highlight that, to date, the factors which constitute a protective immune response to HIV-1 are still unknown, and that a better understanding of the immunology of HIV-1 infection is necessary for the design of more effective alternate therapies as well as immunogens for vaccine development.

It has been suggested that the extraction of the most knowledge from vaccine efficacy trials would need to include the examination of host immune responses, viral sequences (which includes identifying viral strains, in South Africa for example, HIV-1 subtype C is the prevalent circulating strain) as well as host genetics and how these factors overlap and interact with/influence each other over time (Tomaras and Haynes 2014) (Figure 1.2.1).

Studying the genetic contribution to varying host immune factors in HIV-1 infection traditionally involves the determination of genetic variability, in the form of Single Nucleotide Polymorphisms (SNP) and insertions and deletions (indels) and the ways in which these naturally occurring variants may influence dynamics such as protein production, and how varying levels of production, in turn can influence the disease at several levels, ranging from acquisition to progression and disease attenuation. During the past few years however, gene copy number variation (CNV) has emerged as a factor that is responsible for a considerable component of the genetic variability that influences many diseases (Freeman et al. 2006; Wain et al. 2009).

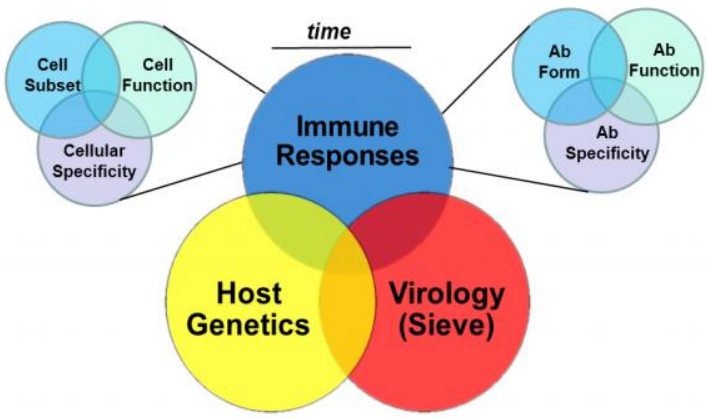



**Figure 1.2.1 Schematic diagram of factors to be considered when assessing vaccine trial data (Tomaras and Haynes 2014).**

A small group of individuals who become HIV-infected but manage to remain healthy, without the use of HAART for long periods of time, offer a unique opportunity to investigate the correlates of HIV-1 protective immunity. Differences in host responses to HIV-1 infection, host genetic variation as well as differences in viral strains have been noted across the progression spectrum of HIV-1-infected individuals, ranging from rapid on the one side of the spectrum, to slow progression on the opposite end of the spectrum (Poropatich and Sullivan 2011). Broadly termed long-term nonprogressors (LTNPs) in this study, these individuals generally maintain normal $CD4^+$ T cell counts, control viral replication and exhibit no clinical symptoms of disease, in the absence of antiretroviral therapy, for extended periods of time (up to 20 years) (Mikhail et al. 2003). The stringent criteria for the catergorisation of LTNPs are however not clearly defined and may differ from study to study. One observation was that viral copy numbers as well as viral mRNA expression in LTNP individuals are comparable to progressors just after seroconversion, suggesting that LTNPs differ from progressors in that they are able to maintain contol over viral replication (Vesanen et al. 1996). After a few years, however, LTNPs do eventually start showing symptoms of progression (Lefrere et al. 1997; Rodes et al. 2004).

Teasing out the host factors contributing to HIV-1 protection, requires a back to basics scientific approach which starts much earlier than the vaccine development stage to first establish baseline data in specific populations, i.e. first understanding these factors in an HIV-1 free backdrop, then elucidating the impact of possible changes/responses of these factors in the context of HIV-1 infection. Furthermore, these studies can provide insights into what possible, additional host responses should be looked at as potential contributers to vaccine trial outcomes.

Reported findings that African populations display the greatest genetic heterogeneity (Tishkoff et al. 2009) compared to other populations, can be explained by the widely accepted theory that modern humans originated in Africa from where they migrated to other parts of the world approximately 150 000 years ago (Seielstad et al. 1999). The intra-ethnic diversity as well as the ethnic diversity displayed in South Africa, combined with the fact that sub-Saharan African countries, more specifically South Africa, carry one of the highest HIV-1 disease burdens worldwide (UNAIDS 2014), emphasises the importance of conducting studies relating to HIV-1 protective immunity in these populations.

Genetic variability in a number of genes, classified as AIDS restriction genes, was shown to influence cytokine regulation, viral entry into cells as well as modulation of innate and acquired immunity. These variants can result in full or partial resistance to HIV-1 infection or they may have an effect on disease progression in individuals who are already infected (O'Brien and Nelson 2004). Some of the genes that have been classified as AIDS restricting genes are those involved in viral entry.

## 1.3. HIV-1 VIRAL ENTRY AND THE C-C CHEMOKINE RECEPTOR-5 (CCR5)

The entry of HIV-1 into a variety of CD4+ host cells, including dendritic cells, T-helper cells, macrophages and monocytes, is not limited to the blood or sites of infection but also several organs, such as the brain and heart (Levy 2006). In brief, the entry process begins with the attachment of the virus to the cell membrane, facilitated by various virus and immune factors (Figure 1.3.1), followed by binding to the CD4 receptor (Dalgleish et al. 1984; Maddon et al. 1986). This involves a high affinity interaction of the trimeric gp120 peptide of the viral envelope glycoproteins (Env) to the CD4 receptor (McDougal et al. 1986) and is followed by a conformational change in the gp120 subunit, exposing its chemokine receptor binding region which binds to either the CXCR4 or CCR5 co-receptor (Deng et al. 1996; Dragic et al. 1996) Thereafter, the virus and cell membrane fuse, forming a hairpin structure and the transmembrane gp41 is inserted into the host cell and the virion released.

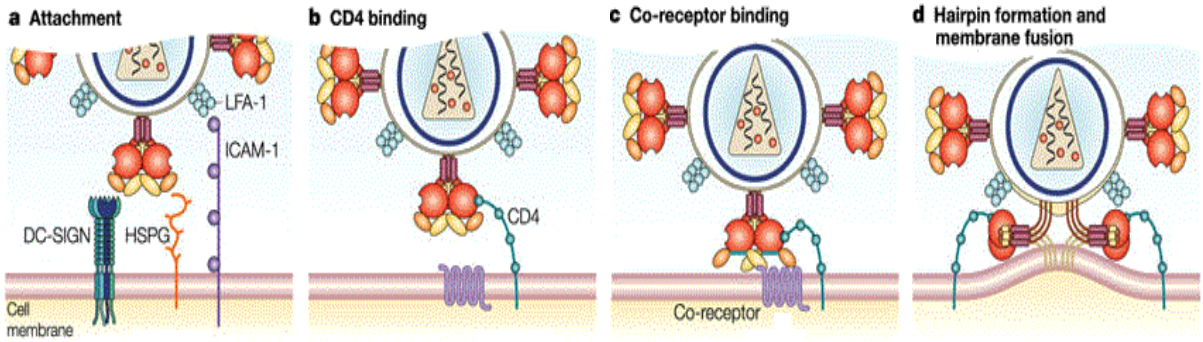



**Figure 1.3.1 Viral entry of HIV-1 into CD4+ cells.** HIV-1 first attaches to cells (a) then binds to the CD4 receptor via the trimeric gp120 subunit of the Env protein (b). Followed by a conformational change in gp120 which facilitates binding to the co-receptor (c) either CXCR4 or CCR5. A hairpin loop is then formed and viral and host membrane fuse (d) (Shattock and Moore 2003).

The CXCR4 receptor is utilised by T-cell-tropic or X4 strains of the virus while the receptor utilised by macrophage-tropic or R5 strains is CCR5. Macrophage tropic HIV-1 is the major virus transmitted from person to person and is the predominant virus in the early stages of infection (Connor et al. 1997).

The CCR5 receptor has 7 transmembrane domains and is a member of the G-protein-coupled receptor family (Samson et al. 1996a). Due to its role in viral entry, CCR5 has been a point of interest in the design of HIV-1 therapies. A 32 base pair (bp) deletion, *CCR5Δ32*, in the *CCR5* gene, leading to a truncated protein that does not display receptor functionality has been shown to offer protection against HIV-1 in individuals who are homozygous or heterozygous for the mutation (Dean et al. 1996; Eugen-Olsen et al. 1997). This deletion only occurs at a frequency of 0.1% in Black South African individuals (Williamson et al. 2000). However, many other SNPs have been reported in the *CCR5* gene which influence HIV-1 disease in a variety of ways, and occur at different frequencies in different populations (Petersen et al. 2001), for example, an association between a SNP (G245A) and a longer time for virologic control by HIV-1 treatment was linked to African ancestry (Cheruvu et al. 2014). Many of the CCR5 SNPs have been described as belonging to specific haplotypes and some of these haplotypes have been associated with HIV-1 disease progression (Martin et al. 1998; Gonzalez et al. 1999; Mummidi et al. 2000). Certain haplotypes, that were identified in multiple populations were found to have opposing effects depending on the population group in which they were present (Catano et al. 2011).

## 1.4. CC CHEMOKINES

Chemokines are small immunoregulatory molecules which belong to different families (C, CC, CXC and CX3C chemokine families), based on the position of conserved cysteine residues in their amino-terminus (Zlotnik et al. 2006) A variety of receptors and their corresponding ligands belong to each family (Figure 1.4.1).

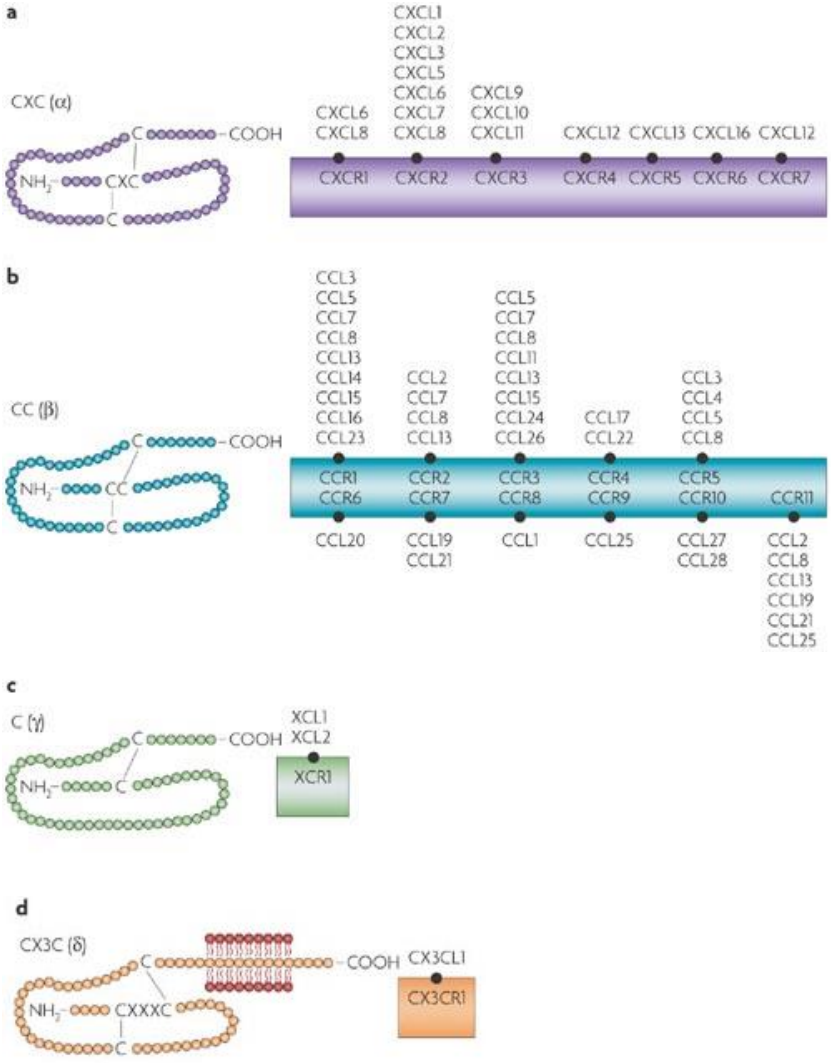



**Figure 1.4.1 Schematic description of the structure of the different families of chemokines and the receptors and ligands which belong to the family (Rostene et al. 2007).**

Non-human primates have been a good research tool since the advent of the chimeric lentivirus (Lu et al., 1999). *In vivo* and *ex vivo* experiments showed that measured levels of chemokines, correlated with protection elicited by anti-Simian Immunodeficieny Virus (SIV) vaccines in macaques (Ahmed et al. 2005). The role for innate immunity in protective immunity and outcome of infection was demonstrated using these models (Lehner et al. 1996; Ahmed et al. 1999; Ahmed et al. 2001). Complete or partial protection against SIV was seen in both naïve and vaccinated macaques who had high levels of production of CC Chemokines as well as CD8+ T-cell antiviral factors. The production of

these chemokines is highly variable in naïve macaques and those exhibiting the highest levels of chemokine production were found to be included in the group that resisted a pathogenic SIVsm challenge after vaccination (Ahmed et al. 1999; Ahmed et al. 2001), suggesting that genetic or environmental factors may affect CC chemokine production. Innate immunity may assist in generating HIV/SIV specific responses upon primary encounter with HIV/SIV.

Cocchi et al. (1995) first demonstrated that the human CC chemokines, C-C motif chemokine ligand-3 (CCL3) (macrophage inflammatory protein, MIP-1α), CCL4 (MIP-1β) and CCL5 (RANTES, regulation upon activation normal T expressed and secreted) were the major HIV-suppressive factors produced by CD8+ T cells. CCL3, CCL4 and CCL5 were initially demonstrated to be the natural ligands of the CCR5 receptor (Combadiere et al. 1996; Raport et al. 1996; Samson et al. 1996b) and later CCL8 (MCP-2, Monocytic chemotactic protein-2) was also identified as a CCR5 ligand (Blanpain et al. 1999). The protection offered by the CC chemokines has therefore been attributed to some or possibly a combination of all of the following proposed methods of hindering viral entry: competitive binding between these ligands and R5 strains of HIV-1, to the CCR5 receptor; downregulation of the CCR5 receptor; or, the induction of receptor dimerisation (Cocchi et al. 1995; Oravecz et al. 1996; Zagury et al. 1998; Vila-Coro et al. 2000). Interestingly, although the chemokine MCP-3 binds to the CCR5 receptor with a high affinity, it does not offer much protection against HIV-1, possibly because, as an antagonist of the receptor, downstream pathways after receptor binding would not be activated and CCR5 down-modulation would not occur (Blanpain et al. 1999).

## 1.5. GENETIC STRUCTURE OF THE CCL3-CCL4 REGION OF CHROMOSOME 17

The genes encoding both CCL3 and CCL4 are located on chromosome 17q21 (Irving et al. 1990; Hirashima et al. 1992). The multiple copy number, allelic isoforms of *CCL3* and *CCL4* (which each occur in two copies per diploid genome (pdg)), are believed to have resulted from a segmental duplication in this region of the chromosome (Menten et al. 2002; Modi 2004). Different population groups exhibit varying copy number ranges and mean gene copy numbers (Colobran et al. 2008) and these genes fall within a group of genes that have been identified as being most stratified by copy number in different populations (Sudmant et al. 2010). While the contribution of CCL3 to HIV-1 disease progression as well as to mother-to-child HIV-1 transmission has been extensively investigated, the same cannot be said about the role of CCL4 in HIV-1 disease (Shostakovich-Koretskaya et al. 2009). The understated possible significance of CCL4 has been suggested to be as a result of failure of studies to separately study, for example, CNV in the *C-C motif chemokine ligand 4-like 1 (CCL4L1)* and *CCL4L2* genes (Shostakovich-Koretskaya et al. 2009). Collectively termed *CCL4L,* these non-allelic isoforms share a 95% sequence similarity with *CCL4* and can be distinguished from each other by a SNP within an acceptor splice site in intron two, giving rise to aberrantly spliced products of *CCL4L2* that do not seem to display classical CCL4 functionality (Colobran et al. 2005) (Figure 1.5.1). The mature, functional protein product produced by *CCL4L1* however, differs from its isoform CCL4, by just one amino acid (Howard et al. 2004; Colobran et al. 2005). These two proteins are therefore not easily differentiated from each other *in vitro* when using monoclonal antibodies currently available. Functional CCL4 is secreted by a number of cell types, such as monocytes, T and B lymphocytes, NK cells, dendritic cells as well as neutrophils, and participates in the immune response through its chemoattractant properties, directing various cell types to sites of inflammation (Menten et al. 2002).

Although the *CCL3L* and *CCL4L* genes are located on the same segmental duplication, there are several break points in this region, which means that the genes will not necessarily occur in the same copy number (Cardone et al. 2008). While CNV of these two genes showed a positive correlation in a Ukrainian population, *CCL3L* occurred at higher copy numbers than *CCL4L* (Shostakovich-Koretskaya et al. 2009). Using the paralogue ratio test (PRT) to asses *CCL3L* copy number, Walker et al. (2009) showed that the combined copy number of *CCL3L1 and CCL3L3* (i.e. the *CCL3*-like genes that produce a functional CCL3 protein and excluding *CCL3L2*, the truncated CCL3 gene) was equal to the copy number of *CCL4L* in a Caucasian population group, however this direct relationship between *CCL3L* and *CCL4L* has not been unequivocally proven for populations with very high copy numbers, like African population groups. Interestingly, a study conducted by Aklillu et al. (2013), using both the PRT method and fluorescence *in situ* hybridization (FISH) with probes to validate copy number calls at higher copy numbers, showed that sub-Saharan Africans display "heterogeneity in *CCL3L1* repeat structure not yet observed in Europeans" and points out that more research into the structural variation at this locus needs to be undertaken.

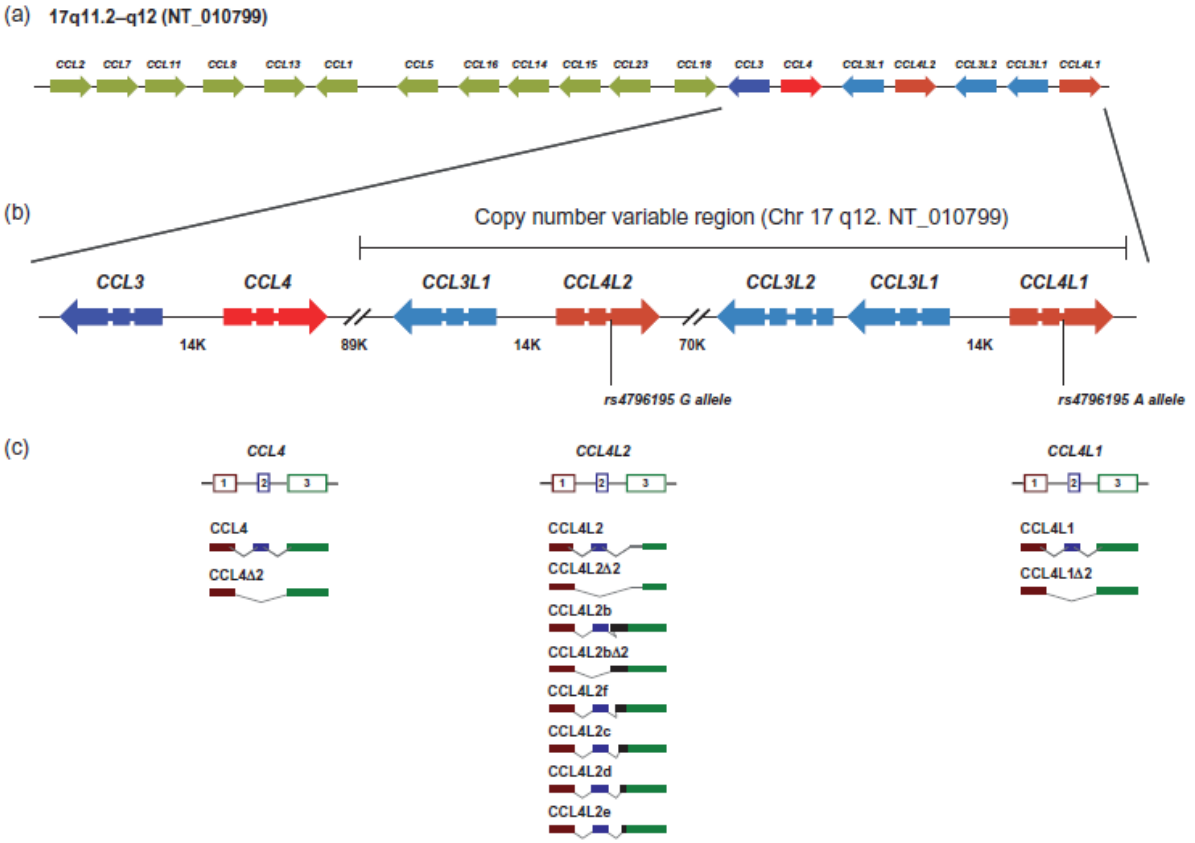



**Figure 1.5.1 Schematic representation of the CCL4 encoding genes and transcripts.** (a) The genes located within the 17q11.2-q12 and amplification of the region containing the CCL3 and CCL4 encoding genes (b). The *CCL4* and *CCL4L1* gene give rise to transcripts with the same functionality while the *CCL4L2* gene which harbours a mutation ('G' allele) in the second intron gives rise to aberrantly spliced products (c) (Colobran et al. 2010).

## 1.6. CONTRIBUTION OF CCL3 AND CCL4 GENE COPY NUMBER AND GENETIC VARIANTS TO HIV-1 DISEASE

While several studies have shown a protective anti-HIV-1 effect of a higher copy number of the genes *CCL3L1* (Townson et al. 2002; Meddows-Taylor et al. 2006; Kuhn et al. 2007; Shalekoff et al. 2008; Shostakovich-Koretskaya et al. 2009) and *CCL4L1* (Shostakovich-Koretskaya et al. 2009), other studies have failed to replicate these results (Bhattacharya et al. 2009; Urban et al. 2009), leading to much controversy in the field (Field et al. 2009; He et al. 2009; Shrestha et al. 2009). Possessing copies of

*CCL4L1* only, (that is, no copies of *CCL4L2*), has been associated with higher levels of CCL4 mRNA production compared to individuals who possessed copies of *CCL4L2* only and no copies of *CCL4L1* (Colobran et al. 2005). A different study suggested that possessing copies of *CCL4L2* only but no copies of *CCL4L1*, led to faster progression to AIDS of HIV-1 infected Ukrainian children (Shostakovich-Koretskaya et al. 2009). These two results however, have not been replicated by others (Shao et al. 2007). Furthermore, the combination of an individual's *CCL3L* and *CCL4L* copy number has also been suggested to be functionally relevant (Shostakovich-Koretskaya et al. 2009). Experiments conducted in lipopolysaccharide stimulated monocytes have found a correlation between copy number of *CCL3L* and CCL3 production suggesting that it is likely that *CCL3L* copy number may have an effect on CCL3 production *in vivo* (Townson et al., 2002). In a MTCT study, it was demonstrated that lower levels of foetal CCL3 associated with an increased risk of intrapartum-acquired HIV-1 infection (Meddows-Taylor et al. 2006). However, further genetic characterisation of these genes, in the hope of developing a genotypic assay to test for increased risk of infection or a more rapid disease progression, has to be carried out because copy number could only partially explain the observed phenotype, suggesting that those infants that did become infected have non or less functional copies of the gene (Gonzalez et al. 2005; Meddows-Taylor et al. 2006). This may be a result of mutations in some of the copies of *CCL3L* or even in the two copy *CCL3* gene, which would explain why there may not always be a direct association between actual copy number and protection, but rather an association between protection and specific population median (Gonzalez et al., 2005). A SNP in CCL5 (28G), for example, was associated with delayed disease progression, most likely through increased expression of CCL5 (Liu et al. 2004). Serum levels of CCL5 however remained the same in the presence or absence of the mutation. Several SNPs identified across the CCL3 and CCL4-encoding genes in both European populations and those of African ancestry were associated with a faster progression to AIDS (Modi et al. 2006). As mentioned above, *CCL3L* did not completely account for the increased expression of CCL3 in infants in an MTCT study (Meddows-Taylor et al. 2006) however, a later study also conducted

on the same African infants, suggested that copy numbers of *CCL3L1* above the population specific median offered protection against acquiring HIV-1 infection in infants but this association was not maintained when mothers took a dose of nevirapine or if their viral loads were low (Kuhn et al. 2007).

The role of increased CCL3 and CCL4 production in LTNP individuals as well as highly-exposed seronegative individuals (HESN) is to date poorly understood. A study following commercial sex workers in Benin found that individuals who were uninfected had higher serum levels of CCL3 and CCL4 than those that became infected over time (Lajoie et al. 2010), but at the mucosa this was not mimicked, suggesting that perhaps recruitment of CD4+ cells to the mucosa might facilitate infection but having them high in the blood is good to prevent infection. Interestingly, it has been suggested that chemokine production may be influenced by hormones, which means that production levels may differ between males and females (Barrenäs et al. 2008).

In a recent study conducted in our laboratory, the contribution of *CCL3L* copy number to CCL3 production in two ethnically divergent (Black and Caucasian) South African populations was investigated (Picton et al. 2013). The study revealed that despite the significant differences in *CCL3L* copy number, no difference in CCL3 production was noted between the two population groups upon *in vitro* stimulation of isolated peripheral blood mononuclear cells (PBMCs), and in addition, no correlation between CCL3 production and *CCL3L* copy number was noted in either the Black or Caucasian cohort (Picton et al. 2013).

The contradictory findings of association studies between copy number variation and HIV-1 protection have been, in part, attributed to the variability of assays used to determine gene copy numbers, such as the quantitative Real-Time PCR (qPCR) TaqMan assay, which displays low accuracy at higher copy

numbers. Thus, an accurate and easily implementable copy number determination assay for populations of African origin is needed.

The recently developed technique of droplet digital PCR (ddPCR) involves emulsification of the amplification reaction into thousands of small droplets, allowing PCR amplification to take place in each droplet. The droplets are subsequently assessed for fluorescence (TaqMan chemistry), indicative of a positive end-point PCR. This technique has proven to generate accurate results in a number of applications (Henrich et al. 2012; Heredia et al. 2013; Strain et al. 2013), including copy number determination of the CCL3-encoding genes (Hindson et al. 2011).

## 1.7. RATIONALE FOR CURRENT STUDY

Taking all the above into consideration and with the aim of trying to understand the relationship between variable copy number-encoded chemokines and chemokine production, ddPCR for copy number determination will be assessed in the *CCL4L*-encoding genes, *CCL4L1* and *CCL4L2*, to compare the same TaqMan assay in two different platforms, qPCR and ddPCR, using the BioRad QX100 system. The *CCL4L*-Taqman assay (Shostakovich-Koretskaya et al., 2009) that will be implemented in this study has an intrinsic validation in that *CCL4L* copy number = *CCL4L1* copy number + *CCL4L2* copy number, and thus the accuracy of the two methods is easily evaluated.

Folllowing copy number determination, the data will be used to investigate the relationship between *CCL4L* copy number and CCL4 protein production in two genetically divergent (Black and Caucasian) South African populations free of HIV-1 infection, since HIV-1 infection in itself impacts significantly on CCL4 production as well as on the production of a number of other chemokines. Since the same two populations cohorts were previously studied by Picton et al (2013), with respect to CCL3

production and *CCL3L* copy number, we will be able to compare and jointly analyse the two datasets. In addition, having access to a small cohort of HIV-1-infected LTNPs, will allow a preliminary comparison of *CCL4* gene copy number variation and protein production between these interesting individuals and the HIV-1-uninfected cohort individuals. Lastly, genetic variation, in terms of SNPs and indels in the two copy pdg, *CCL4* gene in HIV-1 uninfected South African Black and Caucasian individuals will be investigated and these variations and putative haplotypes will be assessed for their contribution to CCL4 protein production.

## 1.8. STUDY OBJECTIVES

The specific objectives of this study are thus to:

i. Evaluate a copy number determination assay of the *CCL4L* genes on the platforms of qPCR and ddPCR

ii. Determine differences in copy number distribution between two HIV-1-uninfected, ethnically divergent South African populations of Black and Caucasian individuals

iii. Assess CCL4 protein production and the influence of *CCL4L* copy number on production in the two populations

iv. Compare *CCL4L* copy number distribution, CCL4 production levels as well as the influence of copy number variation on production in the context of HIV-1-infection in a cohort of long-term nonprogressors

v. Describe genetic variability within the *CCL4* gene in the HIV-1-uninfected cohort and assess the influence of these variations on CCL4 production

# CHAPTER 2 MATERIALS AND METHODS

## 2.1. COHORT DESCRIPTION

### *2.1.1. HIV-1-uninfected individuals*

Copy number variation of the genes *CCL4L*, *CCL4L1* and *CCL4L2,* CCL4 protein production as well as genetic variants in *CCL4* were investigated in a previously well studied cohort of 55 HIV-1-uninfected volunteers (Picton et al. 2012a; Picton et al. 2012b; Lassauniere et al. 2013; Picton et al. 2013; Shalekoff et al. 2013), consisting of 23 South African Black and 32 South African Caucasian individuals. The Black cohort comprised 9 males (39.1%) and 14 females (60.9%) with a mean age of 36.7 years (range: 23-62 years). The Caucasian cohort comprised 11 males (34.4%) and 21 females (65.6%) with an average age of 40.5 years (range: 25-67 years). There were no significant statistical differences in the ages and male:female ratio between the Black and Caucasian cohorts (Mann-Whitney P=0.2674 and Fishers' exact test P=0.7808, respectively).

### *2.1.2. Additional samples for CCL4L copy number determination*

An additional cohort of 30 HIV-1-uninfected Black females, recruited as controls in a previous study that investigated mother-to-child transmission of HIV-1, was used to increase the number of study participants in the Black cohort for the determination copy number variation of *CCL4L* by Droplet Digital PCR (ddPCR).

### *2.1.3. Long-term nonprogressor (LTNP) cohort*

CCL4 production and copy number variation was also investigated in a cohort of 14 LTNPs. Requirements for the inclusion into this cohort were maintenance of stable $CD4^+$ T cell counts >350 cells/µl and a viral load <20 000 RNA copies/ml for more than 5 years in the absence of antiretroviral treatment, at the time of the study. These individuals were prospectively recruited and disease progression will be continually followed to observe whether they maintain these criteria, as some

individuals have not been HIV-1-infected >10 years, the time span for which some studies state an infected individual should remain asymptomatic for in order to be classified as an LTNP.

The LTNP cohort consisted of ten Black and four Caucasian individuals. Measurements of viral load and $CD4^+$ T cell count for the LTNP individuals are shown in Table 2.1.1. The median age of the Black LTNPs was 40.5 years (range: 33-54 years) and comprised four (40%) males and six (60%) females. Among LTNPs, three individuals (LTNP6, LTNP7 and LTNP13) met the criteria of elite controllers, i.e. patients with plasma HIV RNA levels of <50 copies/ml (Pereyra et al. 2008).

**Table 2.1.1 Characteristics of HIV-1-infected long-term nonprogressor study participants**

| Patient ID | Lab ID | Age (years) | Sex | Viral load (copies/ml) | $Log_{10}$ (viral load)[1] | $CD4^+$ T cell count (cells/µl) | Time since diagnosis (years) |
|---|---|---|---|---|---|---|---|
| **Black South Africans (n=10)** | | | | | | | |
| LTNP1 | TG1 | 38 | M | 6 070 | 3.78 | 334 | 9 |
| LTNP2 | TG2 | 47 | F | 5 780 | 3.76 | 400 | 6 |
| LTNP3 | TG4 | 35 | M | 183 | 2.26 | 910 | 9 |
| LTNP4 | TG7 | 54 | M | 23 300 | 4.37 | 968 | 4 |
| LTNP5 | TG9 | 46 | F | <400 | 2.60 | 327 | 9 |
| LTNP6 | TG11 | 32 | F | <40 | 1.59 | 693 | 7 |
| LTNP7 | Pru1 | 54 | F | <40 | 1.59 | >2000 | 14 |
| LTNP8 | Pru2 | 43 | M | 1 155 | 3.06 | 637 | 14 |
| LTNP9 | Pru3 | 36 | F | 1 410 | 3.15 | 775 | 11 |
| LTNP10 | Pru4 | 38 | F | 124 | 2.09 | 379 | 13 |
| **Caucasian South Africans (n=4)** | | | | | | | |
| LTNP11 | TG3 | 53 | M | <400 | 2.60 | 877 | 5 |
| LTNP12 | TG5 | 33 | F | 5 420 | 3.73 | 396 | 5 |
| LTNP13 | TG8 | 67 | M | <40 | 1.59 | 811 | 6 |
| LTNP14 | TG10 | 40 | M | <400 | 2.60 | 327 | 18 |

[1]Where viral load was determined to be <40 or <400 RNA copies/ml, $Log_{10}$(viral load) was determined for values of 39 and 399, respectively.

Pink shaded LTNPs are considered to be Elite Controllers.

For each of the above cohorts, written informed consent was obtained from the participants at the time of enrolment and ethical approval was provided by the Human Research on Human Subjects Committee of the University of the Witwatersrand (Appendix D: Ethical Clearance Certificate).

## 2.2. PERIPHERAL BLOOD MONONUCLEAR CELL (PBMC) ISOLATION AND CULTURE

Fresh whole blood was collected from the cohort of 55 HIV-uninfected individuals as well as the 14 LTNP individuals in tubes coated with ethylenediaminetetraacetic acid (EDTA) as an anticoagulant. PBMCs were isolated by centrifugation of the blood on Ficoll-Plaque Plus (GE Healthcare, Sweden) using standard methods. Plasma was removed and stored at -80°C for future studies. Harvested cells were counted using a haemocytometer and resuspended in RPMI (Roswell Park Memorial Institute) medium (Gibco, Invitrogen, Paisley, UK) supplemented with 10% foetal bovine serum (Gibco, Invitrogen, Paisley, UK) at concentrations of 2 x $10^6$ cells/ml. Cells were aliquoted into 24-well tissue culture plates (Corning, NY, USA) (1.4 x $10^6$ cells/well) and one half of the wells (variable depending on amount of PBMCs isolated) were left unstimulated while the rest of the cells were stimulated by the addition of the plant mitogen, phytohaemagglutinin (PHA) at a final concentration of 12.5 µg/ml. Plates were incubated for 20 hours overnight at 37 °C in 5% $CO_2$. Following incubation, cells were pelleted by centrifugation and supernatants (harbouring the secreted CCL4) were collected and stored at -80°C until further analysis. Cell pellets were also stored at -80°C for genomic DNA extraction.

## 2.3. **DNA EXTRACTION**

Total genomic DNA for use in *CCL4L* copy number determination by quantitative Real-Time PCR (qPCR) and for *CCL4* genetic characterisation, was extracted from fresh EDTA-anticoagulated whole blood samples using the QIAmp DNA Blood Mini Kit (QIAGEN, Dusseldorf, Germany), and stored at -20°C for periods >1 year prior to use.

For genotyping using ddPCR, DNA was extracted using the same kit, from previously isolated and frozen PBMC pellets (isolation described above, 2.2.). The qPCR assay was the standard method for copy number determination in our laboratory, and the *CCL4L* qPCR assays were carried out prior to the acquisition of the Bio-Rad QX100 ddPCR system. It was therefore not possible to use the same DNA for both assays as the DNA used for qPCR had also been used for a number of other genotypic assays and experiments (as already mentioned, this is an extensively studied cohort) and it was therefore necessary to re-extract DNA from stored PBMCs processed at the same time as the initial DNA extractions from blood. To exclude the possibility of DNA quality interfering with the PCR in the qPCR assays, we further assessed DNA, extracted and stored in the same way as used in qPCR testing, from the additional cohort of 30 Black individuals, by ddPCR.

## 2.4. GENE NOMENCLATURE

### *2.4.1. CCL4L nomenclature*

The term *CCL4L* refers collectively to the genes *CCL4L1* and *CCL4L2*, which are non-allelic isoforms of *CCL4* (Gene ID: 6351). Various sources (Colobran et al. 2005; Shostakovich-Koretskaya et al. 2009) concur that on chromosome 17, the sequence of the CCL4 encoding genes begins with *CCL4,* followed by the aberrantly spliced *CCL4L* gene and then by the functional *CCL4L* gene (Figure 2.4.1). The *CCL4L* genes, however, are differently named in different studies, with some sources referring to the upstream gene as *CCL4L1* and the downstream copy as *CCL4L2*, while others refer to the variable copy number gene immediately downstream from *CCL4* as *CCL4L2* and the gene downstream from *CCL4L2* as *CCL4L1*. In the current study, the naming of the genes will follow the latter, in concordance with the annotation updated in GenBank (Benson et al. 2006) in June 2014 and in the study by Colobran *et al.* (2005). The aberrantly spliced *CCL4L* gene which is immediately downstream from *CCL4* will therefore be referred to as *CCL4L2* (Gene ID: 9560) while the fully functional *CCL4L* gene, downstream from *CCL4L2* will be referred to as *CCL4L1* (Gene ID: 388372). These genes are schematically depicted in Figure 2.4.1. The study by Shostakovich-Koretskaya *et al*. (2009), from which the assay used to determine *CCL4L* copy number was adapted, uses a nomenclature in which '1' and '2' are replaced by 'a' and 'b' respectively, based on the sequence of genes on the chromosome as well as gene functionality (Figure 2.4.1). The alphabet naming system will be maintained for comparisons involving *CCL3L* gene copy number.

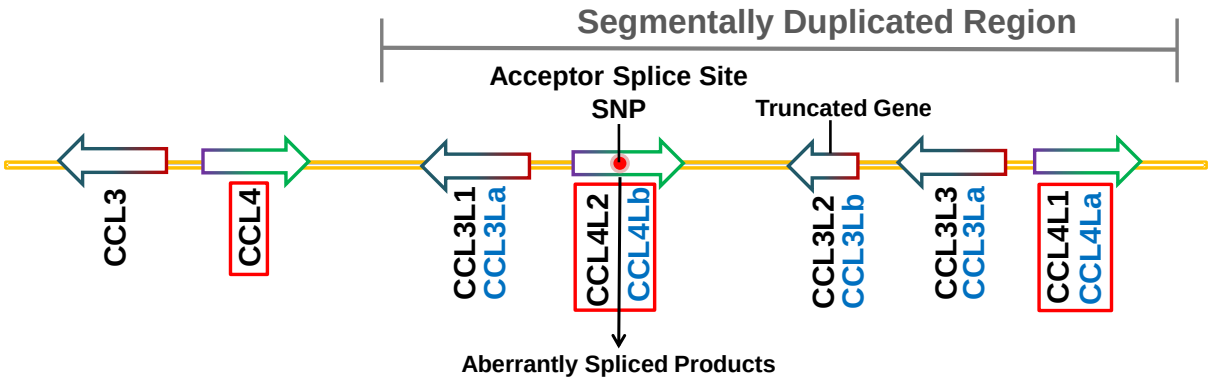



**Figure 2.4.1 Schematic representation of the relative position, order and different nomenclature of the CCL3 and CCL4 encoding genes on chromosome 17q12.**

### *2.4.2.CCL3L nomenclature*

The qPCR assay employed by Picton et al. (2013) was adapted from that described by Shostakovich-Koretskaya et al. (2009). The *CCL3L* genes consist of *CCL3L1* (Gene ID: 6349), *CCL3L2* (Gene ID: 390788) and *CCL3L3* (Gene ID: 414062). Since both *CCL3L1* and *CCL3L3* encode functional proteins (CCL3L), the two isoforms cannot be separately detected or quantified (or distinguished from CCL3 produced by the two copy pdg gene, namely *CCL3* (Gene ID: 6348)) by standard immunological techniques, and are collectively referred to as *CCL3La* (Figure 2.4.1). *CCL3Lb* is the name that has been assigned to *CCL3L2*, which has been referred to as a 'pseudogene' that does not produce the standard functional protein (Shostakovich-Koretskaya et al. 2009).

## 2.5. COPY NUMBER DETERMINATION ASSAYS

In addition to determining copy number variation of the *CCL4L* genes in our cohorts, the methods of qPCR and ddPCR were compared to identify the method that was most efficient and more appropriate for our populations.

### *2.5.1. TaqMan primers and probes*

Primers and probes used to detect the *CCL4L* genes by qPCR and ddPCR were as published by Shostakovich-Koretskaya *et al.* (2009) and are capable of distinguishing the two genes, i.e. *CCL4L1* and *CCL4L2*. Four minor groove binder (MGB) probes were used for targeting the genes *CCL4L* (fluorescent label: 6-carboxyfluorescein (FAM)), *CCL4L1* (fluorescent label: FAM), *CCL4L2* (fluorescent label: VIC) and the *betaglobin* gene (*BGB*, fluorescent label: VIC), which served as an endogenous control, were used for qPCR.

In addition to the probes mentioned above, a FAM-labelled *BGB* probe with the same sequence as the VIC-labelled *BGB* probe was used for ddPCR, as each PCR reaction contained both a gene-specific and an endogenous control primer/probe set (i.e. duplex reactions). The PCR reaction to assess copy numbers of the *CCL4L2* gene using a VIC-labelled probe, therefore, also included a FAM-labelled *BGB* probe, while FAM-labelled *CCL4L* and *CCL4L1* were assessed together with a VIC-labelled *BGB* probe.

### *2.5.2.Evaluation of gene copy numbers by quantitative Real-Time PCR*

A qPCR assay as described by Shostakovich-Koretskaya *et al* (2009) was used to assess copy numbers of the genes *CCL4L*, *CCL4L1* and *CCL4L2* in the cohort of 55 HIV-uninfected individuals. In this method, standard curves of $C_T$ (threshold cycle) values against log [DNA, ng] were generated for each of the four probes (*CCL4L*, *CCL4L1*, *CCL4L2* and *BGB)* from duplicate reactions for six doubling dilutions of

genomic DNA (25ng to 0.78ng) extracted from K562 and A431 cell lines (ATCC, Manassas, VA, USA), having two copies of *CCL4L1,* and two copies of *CCL4L* and *CCL4L2,* respectively. *BGB* was used as an internal control as it is known to occur in two copies pdg. Duplicate wells of each sample were run for each of the four probes. The TaqMan Universal Master Mix (Applied Biosystems, Foster City, CA, USA) was added in all reactions, together with 5 ng of sample DNA.

The quantity of initial template present for each sample was determined by comparing generated $C_T$ values to the standard curves of the respective genes. Copy number is the ratio of the template quantity for the gene of interest to the template quantity for the endogenous control (*BGB*), multiplied by two.

Copy number values were determined for each test replicate and the mean of these values determined. If the standard deviation between replicates was >10% relative to the replicate mean, copy number quantitation was repeated in duplicate, irrespective of whether the sum of *CCL4L1* and *CCL4L2* did not equal *CCL4L*. The mean copy number value was rounded to the nearest integer. If the result was ambiguous, that is, if gene copy number happened to fall on a decimal of around 0.5, the qPCR was repeated for that sample. PCR was conducted in an Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems, Foster City, CA, USA).

### *2.5.3. Evaluation of gene copy number using the Droplet Digital PCR system*

For ddPCR, a double digest was performed on DNA with the restriction enzymes *Bam*HI and *Sac*I (Promega, Madison, Wisconsin, USA) for two hours at 37°C, followed by heat-inactivation of the restriction enzymes for 20 minutes at 65°C. To select which restriction enzymes to use in the digests, reference gene sequences of *CCL4L1* and *CCL4L2* were downloaded from GenBank (Benson et al. 2006). Sequencher version 4.5 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI

USA http://www.genecodes.com) was used to align primer and probe sequences to the genes. Restriction enzymes with cutting sites flanking the amplified and probed gene region, and with similar reaction conditions, were selected. This step is essential to ensure proper target fragmentation for the tandem, syntenic copies of the *CCL4L* genes.

Reaction mixtures (20 µl) containing 75 ng of digested sample DNA, ddPCR Supermix for probes (Bio-Rad, Hercules, CA, USA), 1000 nM of each primer and 200 nM of each probe were loaded into the QX100 Droplet Generator together with 70 µl droplet oil as per the manufacturer's instructions. Post droplet generation, the oil/reagent emulsion was transferred to a 96-well plate (Eppendorf AG, Hamburg, Germany) and the samples were amplified on the conventional Bio-Rad T100 Thermal Cycler (95°C for 10 minutes, followed by 40 cycles of 94°C for 30 seconds and 60°C for 60 seconds, with a final elongation step of 98°C for 10 minutes). The plate, containing the droplet amplicons, was subsequently loaded into the QX100 Droplet Reader (Bio-Rad, Hercules, CA, USA). Standard reagents and consumables supplied by Bio-Rad were used, including, cartridges and gaskets, droplet generation oil and droplet reader oil.

Reactions for each sample were run in triplicate for each gene, *CCL4L*, *CCL4L1* and *CCL4L2*. Each reaction well contained a gene-specific target as well as an endogenous control target PCR (*BGB*), in a duplex reaction.

In the QX100 Droplet Reader, individual droplets are counted as positive or negative based on the presence or absence of fluorescence of the DNA molecules that are enclosed within the droplets. These readings are then compared to the count obtained for *BGB,* which is known to occur as two copies pdg. The number of gene copies present is determined by the QuantaSoft Software version 1.2.10.0 (Bio-Rad, Hercules, CA, USA) by applying Poisson statistics to the fraction of end-point

positive reactions (Hindson et al. 2011). The average number of droplets generated in each reaction was 11 946. This method was used to generate CCL4 copy number data for the cohort of 55 HIV-1-uninfected individuals and a further 30 Black individuals as well as the cohort of 14 LTNPs.

For ddPCR, since the standard deviation between replicates was <10% for all but one sample, we decided to apply more stringent criteria and repeated samples (in triplicate) on the rare occasion (3/55) that the sum of the copy number of *CCL4L1* and *CCL4L2* was not equal to the copy number of *CCL4L* (see below).

### *2.5.4.PCR amplification and sequencing of CCL4L of select individuals*

Since samples that presented with unusual copy number patterns in the ddPCR assay were in the low copy number range and were maintained upon repetition, we decided to investigate further. To validate the presence of the *CCL4L* genes in these samples, primers were designed (Same method of primer design and amplification used as described in 2.8.1.) (fwd: 5' GCAGTCTATCAGTCAATGGT 3', rev: 5' CACATCACCTCTAACAGC 3') to amplify the *CCL4L* genes collectively by conventional PCR (Expand High Fidelity PCR System, Roche, Mannheim, Germany).

Positive amplifications were indicative of the presence of *CCL4L*, however to determine whether *CCL4L1, CCL4L2* or both genes were present, amplicons were sequenced (sequencing methodology and analysis explained below in 2.8.2. and 2.8.3.). Sequencing data were analysed at the single SNP position (rs4796195), that distinguishes the two genes (Colobran et al. 2005) and possible SNPs in the binding sites of primers and probes used for the TaqMan assays, which may have contributed to the unexpected copy numbers detected by ddPCR, were investigated.

## 2.6. CCL4 PRODUCTION ANALYSIS

Quantification of CCL4 in unstimulated and PHA-stimulated PBMC supernatants from the HIV-1-uninfected cohort as well as the LTNP cohort, was performed by sandwich enzyme-linked immunosorbent assay (ELISA; DuoSet ELISA Development Systems: R&D systems, Minneapolis, MN, USA). Reactions were carried out in triplicate following methods prescribed by the manufacturer. The minimal detectable levels were <10 pg/ml. A standard curve was generated using a duplicate doubling serial dilution (1000-15.625 pg/ml) of the CCL4 standard and this was used to calculate the CCL4 concentration of samples based on mean optical density, which was measured using the Oragnon Teknika Microwell System Reader. Plasma samples were assessed undiluted, while the unstimulated supernatants and the stimulated supernatants were diluted by factors of 10 and 100, respectively.

## 2.7. COMPARISON OF CCL4 AND CCL3 PRODUCTION AND COPY NUMBER VARIATION

Data generated for CCL4 and its encoding genes by the methods described above were compared with CCL3 raw data generated by and analysed in detail by Picton *et al.* (2013) in the same cohort of 55 HIV-uninfected individuals. CCL3 and CCL4 production data were both determined by ELISA as described above (2.6), under the same conditions at the same time point.

For the quantitation of CCL3, supernatants from PHA-stimulated cells were diluted by a factor of 1000 (Picton et al. 2013). Copy number variation of the *CCL3L* genes, however, was assessed using quantitative Real-Time (qPCR) unlike *CCL4L* copy number variation which was measured using ddPCR.

Raw CCL3 copy number and production data generated and analysed by Picton et al (unpublished) in the same cohort of 14 LTNPs was used to investigate differences between CCL3 and CCL4 production and the influence of copy number variation on chemokine production in these individuals.

## 2.8. GENETIC CHARACTERISATION OF THE *CCL4* GENE

### *2.8.1. PCR amplification of CCL4*

An alignment of the *CCL4*, *CCL4L1* and *CCL4L2* reference sequences, obtained from GenBank, was generated using DNAMAN phylogenetic software (Version 4.0, Lynnon Biosoft, Pointe-Claire, Quebec, Canada) and was used to design PCR primers to amplify the *CCL4* gene. Designing primers based on this alignment was necessitated by the fact that the *CCL4* and *CCL4L* genes share such a high sequence similarity, and it was therefore necessary to ensure that the primers designed selectively amplified the *CCL4* gene alone and not the *CCL4L* genes. Primer design was further complicated by the fact that many regions in which the *CCL4* and *CCL4L* genes differed significantly, were unsuitable due to the formation of DNA secondary structures. The limited choice of suitable regions in which to pick primers, led to a large 4.7 kb fragment that included the core promoter region of the gene, the 5' and 3' untranslated regions (UTR), the complete gene (exons 1, 2 and 3 and introns 1 and 2) and well as regions upstream and downstream from the elements mentioned, being amplified by two sets of overlapping primers (Table 2.8.1).

**Table 2.8.1 Primer pairs used for the amplification of the *CCL4* gene using the Expand High Fidelity PCR System**

| | | Sequence | Fragment Size | Annealing Temperature^ |
|---|---|---|---|---|
| ***CCL4* Fragment 1** | Forward | 5' CATGTTGCCCAGTCTGGTATCG 3' | 2514bp | Set 1: 60 |
| | Reverse | 5' GGTCATCTGGCTTTATCCTTTGCAT 3' | | Set 2: 65 |
| ***CCL4* Fragment 2** | Forward | 5' GTTAGGTGGGAATGGATATTTGG 3' | 3010bp | Set 1: 60 |
| | Reverse | 5' CTGAGGAATAGGATGTGTGCAC 3' | | Set 2: 60 |

^ The Expand High Fidelity PCR System recommended cycling conditions involve two sets of denaturation, annealing and elongation cycles within the initial denaturation and final elongation steps. Annealing temperatures for each of the two sets are indicated.

Putative primers selected from the alignment, were analysed using the program GeneRunner (version 3.05, Hastings Software Inc, http://generunner.net/) to determine whether they met optimal suggested primer design standards and were subsequently tested for specificity using Primer-Blast (Ye et al. 2012) to confirm that there was no specificity to the *CCL4L* genes or any other regions of the genome.

Standard PCR was carried out using the Expand High Fidelity PCR system (Roche, Mannheim, Germany) and cycling conditions were as recommended by the manufacturer. DNA extracted from whole blood samples from the cohort of 55 HIV-1-uninfected individuals (2.3) was used to amplify CCL4.

### *2.8.2.Sequencing*

All PCR amplicons were purified using the QIAquick PCR Purification Kit (Qiagen, Dusseldorf, Germany). Sequencing reactions, utilising overlapping primers that were designed for specific fragments, were carried out using BigDye Terminator version 3.1 (Applied Biosystems, Foster City, CA, USA) and electrophoresed on the automated 3100 Genetic Analyzer (Applied Biosystems). (Sequencing primers are listed in Appendix A).

### *2.8.3. Sequencing analysis*

Sequences were assembled using Sequencher version 4.5/5.3 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com) and chromatographs were analysed individually for identifiable SNPs.

An alignment of individual sequences as well as the *CCL4* reference sequence (GenBank) was generated using Sequencher Software, in order to identify homologous nucleotide substitutions. Any polymorphic positions identified within the amplified region were viewed using the NCBI Sequence Viewer (version 3.6) and if they had been previously identified, SNP information and accession numbers were obtained from Database of Single Nucleotide Polymorphisms (dbSNP, Bethesda (MD): National Centre for Biotechnology Information, National Library of Medicine. dbSNP accession: rs143457996– rs17617372, (dbSNP Build ID: 142 );Available from: http://www.ncbi.nlm.nih.gov/SNP/).

### *2.8.4. Assigning the wild type (WT) allele*

It was assumed that the ancestral allele or WT allele was the most prevalent allele detected and if there were discrepancies between the two population groups, the allele most prevalent in the Black population was assigned as the WT. In cases where it was still unclear, alignment to the *Pan troglodytes* sequence (Gene ID: 454593, GenBank) was used to determine the ancestral allele.

### *2.8.5. Haplotype prediction/analysis*

For the prediction of haplotypes in both the Black and Caucasian populations, the SNP/indels across the CCL4 gene were analysed both visually as well as by using Haploview software (version 4.2) (Barrett et al. 2005) to determine linkage disequilibrium (LD) between genetic variants. Linked SNPs with $r^2 > 0.6$ were considered to be part of the haplotypes we described.

### *2.8.6.Characterization of putative insertions/deletions*

Four putative indels were identified in regions in which chromatograms indicated two different templates. In many cases, these fragments were resequenced to eliminate the possibility of a mixed template. Reverse PCR primers were designed for each indel, to generate sequence information after the point of the indel to the next sequencing primer. (indel primers are listed in Appendix A)

### *2.8.7.Hardy-Weinberg equilibrium*

Haploview software is also able to generate Hardy-Weinberg deviation values and was used to assess polymorphisms in each of the population groups independently.

### *2.8.8.Bioinformatic tools used*

The CCL4 mRNA sequence (NM_002984.3) was obtained from Genbank and analysed with Sequencher Software. The first frame protein translation displayed by Sequencher was checked against the published CCL4 protein sequence to evaluate what amino acid changes would result from exonic SNPs identified. These changes were investigated using Poly-Phen 2 (prediction of functional effects of human non-synonymous SNPs data base) Software (Adzhubei et al. 2010) to predict the structural/functional consequences of specific amino acid changes detected.

## 2.9. **STATISTICAL ANALYSIS**

The ages and male: female ratio of the South African Black and Caucasian cohorts were compared using the nonparametric Mann-Whitney test and the Fishers' exact test respectively. Correlations between the copy number of *CCL4L* and the combined copy numbers of *CCL4L1* and *CCL4L2* were calculated using the two-tailed Pearson's test. All statistical analyses were performed using GraphPad

Prism version 4.02 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). *P* values <0.05 were considered significant.

Fisher's exact test and non-parametric Mann-Whitney U-tests comparisons were used to determine differences in copy number distribution and CCL4 protein production levels between study groups. Correlations between copy number and production were also calculated using GraphPad Prism version 4.02. The same statistical software and analyses were used in the comparison of copy number and production data generated for CCL3 in the uninfected and LTNP cohorts.

Allelic and haplotype frequencies of SNPs within the CCL4 genes were calculated using Microsoft Excel 2013 (Microsoft, USA) and Haploview software was used to calculate deviations from Hardy-Weinberg equilibrium, as well as LD $r^2$ and D' values. Where LD was not complete between SNPs in respective haplotypes, calculation of haplotype frequencies only included individuals who had all SNPs within the haplotype.

# CHAPTER 3 RESULTS

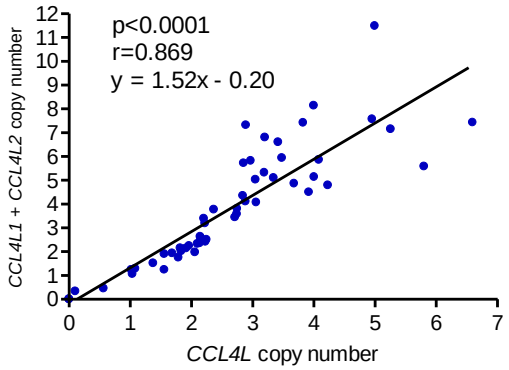
## 3.1. COMPARISON OF qPCR AND ddPCR FOR *CCL4L* COPY NUMBER DETERMINATION

### *3.1.1. Assay validation*

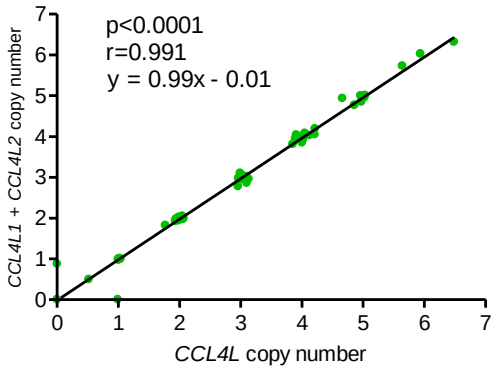
When comparing the qPCR and ddPCR techniques for CNV determination, the *CCL4L* genes are ideal because CNV can be generated for *CCL4L1* and *CCL4L2* individually as well as for the collective *CCL4L* genes as a total. The correlation of *CCL4L* gene copy number to the sum of *CCL4L1* and *CCL4L2* therefore, serves as a good way to validate the accuracy of the results generated from either assay.

While the correlations for both methods assessed in this study showed a significant positive deviation from zero (p value<0.0001, Figure 3.1.1A), the data generated from the unrounded results (raw data) of ddPCR were found to show a stronger correlation (r=0.991) than that of the qPCR assay (r=0.869). The strong correlations generated by comparing copy number data generated by ddPCR are maintained as the copy number of *CCL4L* and the sum of *CCL4L1* and *CCL4L2* increase (Figure 3.1.1Aii). It is, however, evident when examining the correlation generated by qPCR that the accuracy of the assay decreases as the copy number increases (Figure 3.1.1Ai). In fact, when copy numbers were stratified into two groups (0-3 and >3) and correlations determined separately, it was found that at copy numbers 0-3, data generated by both qPCR and ddPCR had strong correlations (r=0.88 and r=0.96, respectively; p's<0.0001, ), however at copy numbers >3, the qPCR data had a very weak correlation (r=0.44; p=0.057), while the ddPCR data maintained a very strong correlation (r=0.99; p<0.0001), highlighting the weakness of qPCR at high copy numbers. This is a major shortfall of the method, especially when investigating African populations, in which copy number ranges and means of the *CCL4L* genes (Range: 3-6, median: 4, Table 3.1) were found, as previously reported (Colobran et al. 2008), to be higher compared to Caucasian individuals (Range: 0-4, median: 2, Table 3.1). Percentage PCR efficiencies for all three primer sets were calculated to be >92%.

Ai




Aii




Bi

| CCL4L \ CCL4L1 + CCL4L2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 4 | | | | | | | | | | | | 4 |
| 1 | | 5 | | | | | | | | | | | 5 |
| 2 | | | 15 | 2 | | | | | | | | | 17 |
| 3 | | | | 2 | 3 | 4 | 2 | 3 | | | | | 14 |
| 4 | | | | | 3 | 1 | 2 | | 1 | | | | 7 |
| 5 | | | | | | | | 2 | | | | 1 | 3 |
| 6 | | | | | | 1 | 1 | | | | | | 2 |
| 7 | | | | | | | | 1 | | | | | 1 |
| 8 | | | | | | | | | | | | | 0 |
| 9 | | | | | | | | | | | | | 0 |
| 10 | | | | | | | | | | | | | 0 |
| 11 | | | | | | | | | | | | | 0 |
| Total | 4 | 5 | 15 | 4 | 6 | 6 | 4 | 6 | 1 | 0 | 0 | 1 | 53 |

Bii

| CCL4L \ CCL4L1 + CCL4L2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | | | | | | 2 |
| 1 | 2 | 5 | | | | | | 7 |
| 2 | | | 15 | | | | | 15 |
| 3 | | | | 11 | | | | 11 |
| 4 | | | | | 11 | | | 11 |
| 5 | | | | | | 6 | | 6 |
| 6 | | | | | | | 3 | 3 |
| Total | 3 | 6 | 15 | 11 | 11 | 6 | 3 | 55 |

| | CCL4L | CCL4L1 | CCL4L2 |
|---|---|---|---|
| CCR04 | 0 | 1 | 0 |
| CCR17 | 1 | 0 | 0 |
| CCR52 | 0.5 | 0.25 | 0.25 |

**Figure 3.1.1 Concordance between *CCL4L* copy number and the sum of the copy number of *CCL4L1* and *CCL4L2* generated by two assays.** (A) Scatter plots of unrounded average copy number data generated using (Ai) quantitative Real-Time PCR and (Aii) Droplet Digital PCR. (B) Numbers in the coloured blocks indicate the number of individuals in which the rounded off *CCL4L* copy number was equal to the sum of *CCL4L1* and *CCL4L2,* and numbers in the blocks that are clear indicate instances where they were not equal. Data was generated using two different assays, that is, (Bi) quantitative Real-Time PCR assay and (Bii) Droplet Digital PCR. The adjacent table shows the copy numbers obtained for the individuals that did not fall within the shaded blocks in the ddPCR assay.

**Table 3.1.1 Population specific medians and ranges generated by qPCR and ddPCR for the South African Black and Caucasian populations**

| | Black | | Caucasian | |
|---|---|---|---|---|
| | **qPCR** | **ddPCR** | **qPCR** | **ddPCR** |
| ***CCL4L*** | 4 (2-7) | 4 (3-6) | 2 (0-4) | 2 (0-4) |
| ***CCL4L1*** | 2 (1-4) | 3 (1-5) | 2 (0-3) | 2 (0-3) |
| ***CCL4L2*** | 3 (0-7) | 2 (0-4) | 0 (0-5) | 0 (0-2) |

When samples were assessed using qPCR, there were 22 individuals in which the combined rounded copy number of *CCL4L1* and *CCL4L2* did not equal to the copy number generated for *CCL4L* (n=53, Figure 3.1.1Bi). This was observed in just two individuals assayed with ddPCR and one additional individual presented with an unusual copy number pattern (n=55, Figure 3.1.1Bii). Upon repetition of these three samples, the same results were obtained and qPCR data suggested that there were 0 copies of the *CCL4L* genes in these individuals. To further investigate this, the *CCL4L* genes of the three individuals were subsequently PCR amplified and sequenced. The 0 copy number of *CCL4L* found in individual CCR04 (copy number pattern 0, 1, 0 for *CCL4L*, *CCL4L1* and *CCL4L2,* respectively) was found to be a result of a homozygous SNP in the *CCL4L* probe binding site, which would result in failure of the probe to bind in the *CCL4L* reaction resulting in an incorrect '0' copy number call. Sequencing data confirmed that individual CCR17 (copy number pattern 1, 0, 0 for *CCL4L*, *CCL4L1* and *CCL4L2,* respectively), did in fact have a *CCL4L1* gene and attributed the 0 copy number result to a 3' end homozygous SNP in the *CCL4L1* reverse primer binding site. The copy number patterns for CCR04 and CCR17 were therefore confirmed to be 1, 1, 0 for *CCL4L*, *CCL4L1* and *CCL4L2,* respectively. Although the qPCR for CCR04 and CCR17 was repeated a number of times and averaged out to 0 copies for both these individuals, when the raw data was re-examined, there were some repeats that yielded the same copy number pattern seen with ddPCR, however, these were not consistently seen in the majority of the repeats.

Interestingly, no SNPs were identified in the primer and probe binding regions of CCR52 (copy number pattern 0.5, 0.25, 0.25 for *CCL4L*, *CCL4L1* and *CCL4L2,* respectively). However, the presence of both the genes *CCL4L1* and *CCL4L2* was confirmed. It is speculated that this individual may have a SNP in a primer or probe binding site of the *BGB* reference gene. If this is the case, considering that gene copy number of the *CCL4L* genes are generated relative to *BGB,* the estimated copy number would be less than expected but as observed, would maintain the correct gene ratio, that is, the same number of *CCL4L1* and *CCL4L2* genes are present in individual CCR52.

DNA samples that had been extracted from whole blood and stored for several years were used for qPCR whereas freshly extracted DNA from PBMC's was used for ddPCR. To rule out the possibility of DNA quality contributing to the results we further tested DNA, extracted and stored in the same way as used in qPCR testing, from an additional cohort of 30 Black individuals and assessed these by ddPCR. A 100% correlation between the copy number of *CCL4L* and the sum of *CCL4L1* and *CCL4L2* was observed ($r=0.996$, $p<0.001$, Figure A in Appendix B).

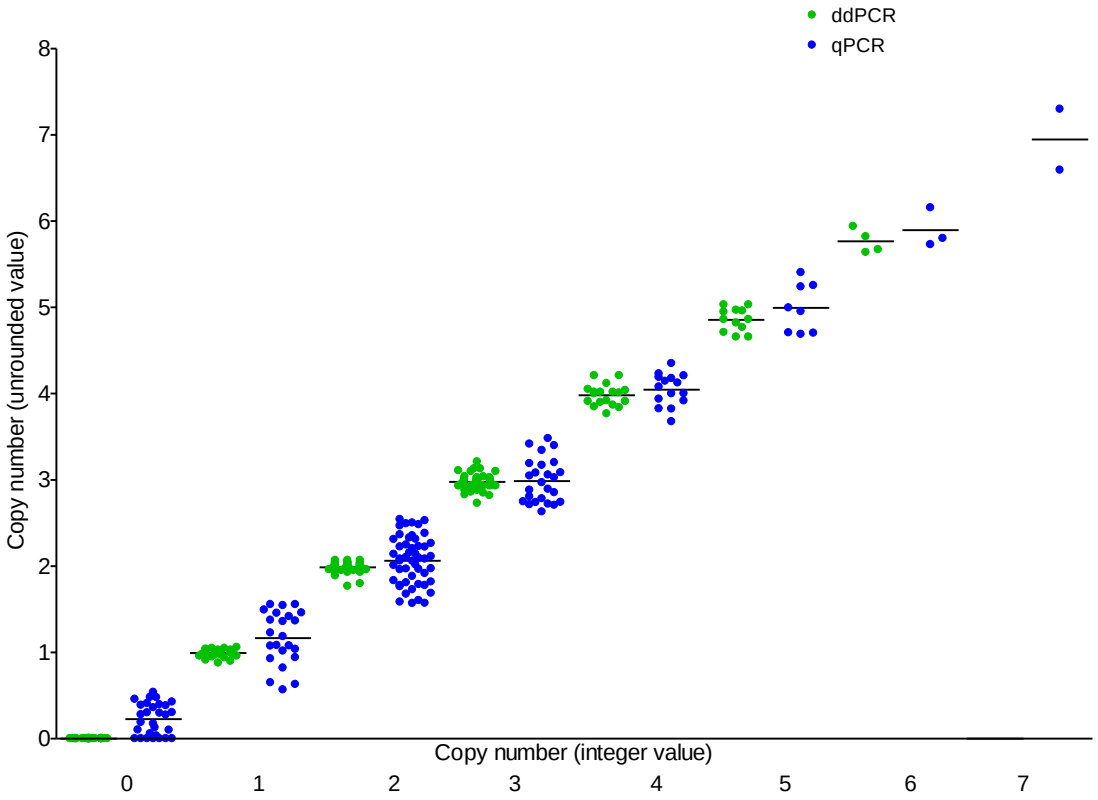



**Figure 3.1.2 Clustering of unrounded average copy number values for all three *CCL4L* genes generated by quantitative Real-Time PCR and Droplet Digital PCR around whole integer copy numbers**. Blue circles represent copy numbers generated by qPCR and green circles represent copy numbers generated by ddPCR. Y-axis: unrounded values. X-axis: integer values.

### *3.1.2. Reproducibility of results*

Out of a cohort of 55 individuals genotyped by the qPCR assay, repeats had to be performed on 87.3 % of individuals (48 samples) - when percentage deviations between replicates were greater than 10% compared to the replicate mean. Out of the 55 samples amplified using ddPCR, only 5.5% (3 samples) had to be repeated due to suspected inaccuracy, and only a single sample showed a standard deviation of above 10%. Additional repeats had to be carried out on samples assayed with qPCR as inter-experimental variation was often observed for individual samples, suggesting a low reproducibility of these results. On average, each sample had to be repeated three times. Two of the samples remained

unresolved and there were often ambiguities in the data when decimals of 0.5 were observed, as it was unclear whether to round up or round down to the nearest integer. Standard deviations between repeats, generated by ddPCR, were below 10%, except on one occasion where the copy number reads for the three repeats of the *CCL4L1* gene were 0.85, 1.10, 1.01 and could therefore easily rounded off to 1. When comparing unrounded average copy number values generated by the two methods, those generated by ddPCR showed a tighter clustering to the whole copy number values than those generated by qPCR (Figure 3.1.2).

The data from qPCR were generated by relative comparison to a standard curve which means that if the R-squared value of the standard curve is low due to experimental error, the data from the entire plate is unusable. The ddPCR method measures absolute quantifications and generation of a standard curve is not necessary.

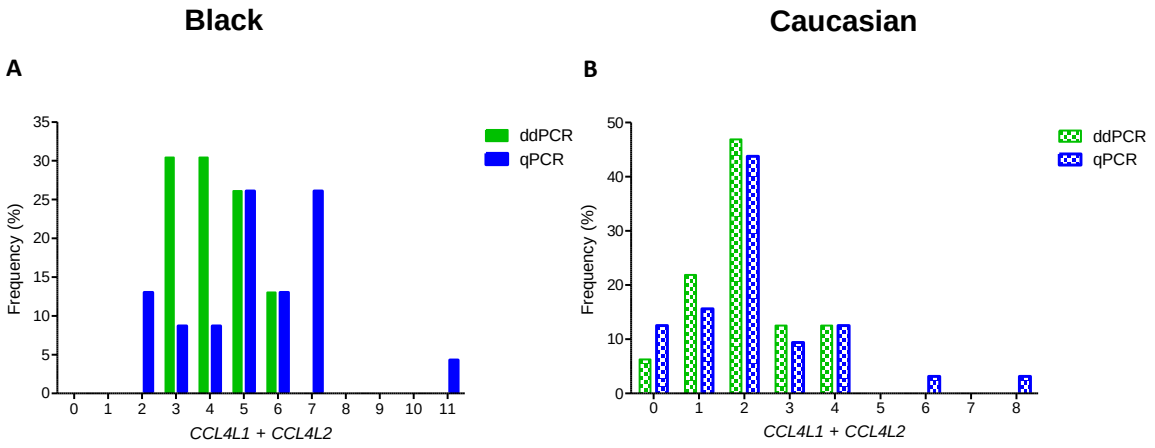



**Figure 3.1.3 Copy number frequencies of the sum of genes *CCL4L1* and *CCL4L2*.** Copy number frequencies were determined using two different methods in the South African (A) Black and (B) Caucasian populations. Green bars represent ddPCR while blue bars represent quantitative real-time PCR.

### *3.1.3. Population specific medians*

While the same copy number medians were calculated for the Caucasian cohort regardless of method used (2, 2, 0 for *CCL4L*, *CCL4L1* and *CCL4L2* respectively, Table 3.1.), and just a single difference in the copy number ranges was observed (*CCL4L2* gene: a maximum of 5 versus a maximum of 3), the inability of this assay to accurately deal with higher copy numbers was again reiterated when examining the medians and ranges in the Black individuals (Table 3.1.1, median copy numbers using qPCR 4, 2, 3 and 4, 3, 2 using ddPCR for *CCL4L*, *CCL4L1* and *CCL4L2* respectively). Furthermore, the frequency distribution of the sum of *CCL4L1* and *CCL4L2*, generated by both methods, follows a similar pattern of distribution around the population median when results from the Caucasian population are observed (Figure 3.1.3B), while the pattern of distribution differs substantially in the Black population when the two methods are compared (Figure 3.1.3A). The individual gene correlations between the copy numbers generated by ddPCR and qPCR were higher in the Caucasian population compared to the Black population (Supplementary Figure A).

## 3.2. *CCL4L* COPY NUMBER ANALYSIS

Due to the efficiency of ddPCR to determine CNV, copy number data generated by this method was carried through for *CCL4L* distribution analysis and used to compare with CCL4 production levels.

### *3.2.1.Gender and copy number distribution*

Since there was no significant difference in *CCL4L* copy number between males and females in the HIV-1-uninfected Black cohort, an additional 30 HIV-1-uninfected female individuals were added to the Black sample group to increase the cohort size, thereby allowing the determination of a more accurate representation of copy number distribution for the Black South African population group. Upon reassessment of the data including the 30 additional females, a significantly higher number of copies of *CCL4L* was observed in Black female individuals compared to Black male individuals (p=0.003, Figure 3.2.1A) but the number of *CCL4L1* and *CCL4L2* genes did not differ significantly between the two groups (Figures 3.2.1B and 3.2.1C). There was however a trend (p=0.073) of higher representation of *CCL4L2* in Black females compared to black males (Figure 3.2.1C), suggesting that this gene was the major contributor to the overall significant higher *CCL4L* copy number in Black females vs. males.

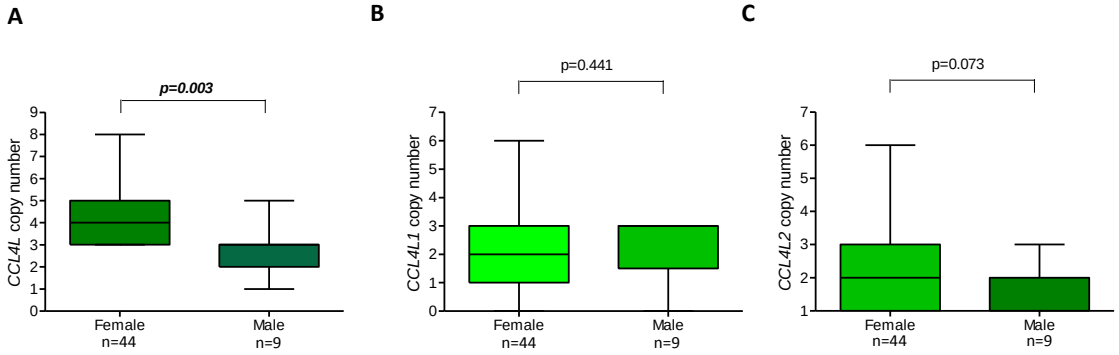



**Figure 3.2.1 Gender differences in copy number distribution between Black South African females and males for the (A) *CCL4L* genes, (B) *CCL4L1* and (C) *CCL4L2* individually.** P values and number of participants are indicated.

### *3.2.2.CCL4L copy number variation between the Black and Caucasian population groups*

As expected and reported in other studies, significant differences were observed between the copy number distribution of the genes *CCL4L*, *CCL4L1* and *CCL4L2* between the two populations (Figure 3.2.2). Copy number medians and ranges for both genes individually and collectively, in both populations were previously reported (Table 3.1.1), however with the addition of the 30 additional individuals to the Black cohort, the population median and ranges of the genes *CCL4L* (range: 3-8, median: 4), *CCL4L1* (range: 0-5, median: 2) and *CCL4L2* (range 0-6, median: 2) shifted slightly from ranges and medians for *CCL4L*, *CCL4L1* and *CCL4L2* that were originally determined in the ddPCR assay development (medians: 4, 3, 2 for *CCL4L*, *CCL4L1* and *CCL4L2* respectively). Figure 3.2.3 provides a graphical representation of the copy number distribution in the Black and Caucasian cohort and indicates the frequency of different copy number combinations. It is interesting to note that compared to Black individuals, just a single Caucasian individual had a *CCL4L2* copy number of two, and the rest of the individuals in the cohort possessed either one or no copies. Similarly, just a single Caucasian individual had a *CCL4L1* copy number of 3 while others had either 0, 1 or 2 *CCL4L1* copies in contrast to the Black cohort, in which multiple individuals had higher copies of *CCL4L1* and *CCL4L2*.

When we compared the relationship between *CCL4L1* and *CCL4L2* copy number, there was an inverse correlation between the copy number of *CCL4L1* and *CCL4L2* in the Black population ($p<0.001$, Figure 3.2.4A), that is, the more copies of *CCL4L2* an individual has, the fewer copies of *CCL4L1* they have. This relationship was not observed in the Caucasian population ($p=0.194$, Figure 3.2.4B).

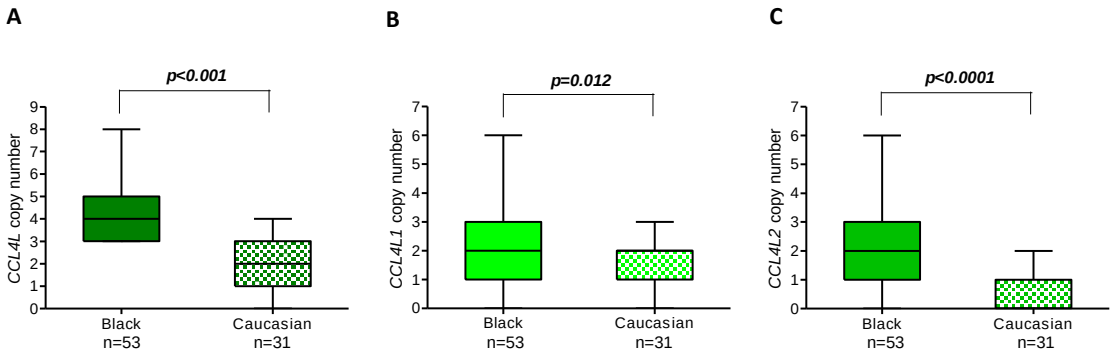



**Figure 3.2.2 Copy number variation of the *CCL4L* genes in the South African Black and Caucasian populations.** Black individuals display significantly higher copies of (A) *CCL4L,* (B) *CCL4L1* and (C) *CCL4L2* compared to Caucasian Individuals. P values and number of participants are indicated.

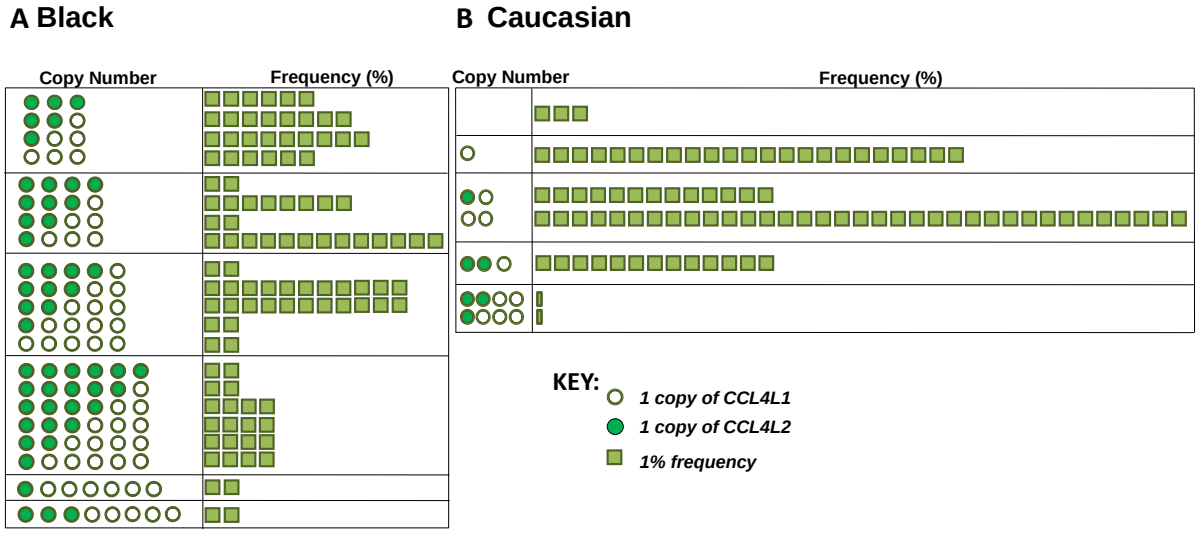



**Figure 3.2.3 Differences in distribution and frequency of *CCL4L* copy number and *CCL4L1* and *CCL4L2* copy number combinations in HIV-1-uninfected (A) Black (n=53) and (B) Caucasian (N=31) South African cohorts.** Circles represent the number of copies of *CCL4L* present, which equals to the sum of the number of copies of *CCL4L1*, represented by the open circles, and *CCL4L2*, represented by the shaded circles. Green squares represent the percentage frequency at which the various combinations occur in the two populations.

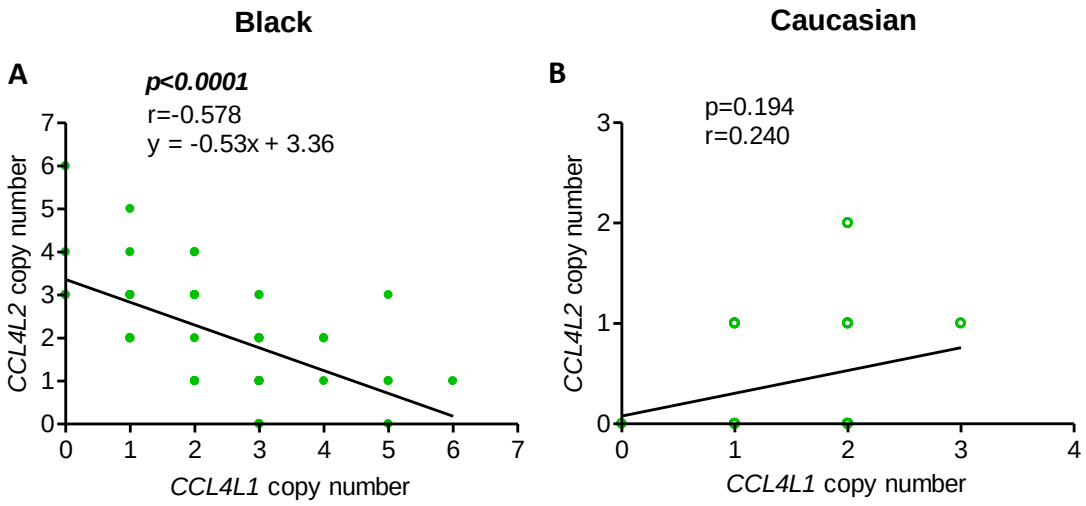



**Figure 3.2.4 Inverse significant correlation between *CCL4L1* and *CCL4L2* copy numbers in (A) Black but not in (B) Caucasian individuals.** P and r values are indicated, as well as line equation where correlation is significant**.**

## 3.3. CCL4 PROTEIN PRODUCTION AND CORRELATION TO COPY NUMBER

### *3.3.1.Population and gender differences in CCL4 production*

CCL4 produced by isolated unstimulated PBMCs and PHA-stimulated PBMCs were assessed in culture supernatants from 32 Caucasian and 23 Black individuals by ELISA. Protein levels did not differ significantly when compared across gender and age groups (age ≤35, age>35) in the whole cohort (p=0.775 and p=0.953, respectively) and individually within race groups (Black age: p=0.926, gender: p=0.136; Caucasian age: p=0.175, gender: p=0.081). Furthermore, unlike the copy number distribution, CCL4 production did not significantly differ between the race groups when comparing unstimulated PBMCs (p=0.562 Figure 3.3.16Ai) or PHA-stimulated PBMCs (p=0.335, Figure 3.3.1Aii). Black females, however, had a significantly higher level of PHA-stimulated CCL4 production than Caucasian females (Figure 3.3.1Bi, p=0.028) but this difference was not seen when PHA-stimulated CCL4 production was compared between Black and Caucasian males (Figure 3.3.1Bii, p=0.224). No

difference in CCL4 production was observed between the females of the two populations when unstimulated production data was analysed (p=0.579, data not shown). CCL4 production between Black individuals and Caucasian individuals of similar ages did not vary significantly (<35: p=0.183, ≥35 p=0.905, data not shown).

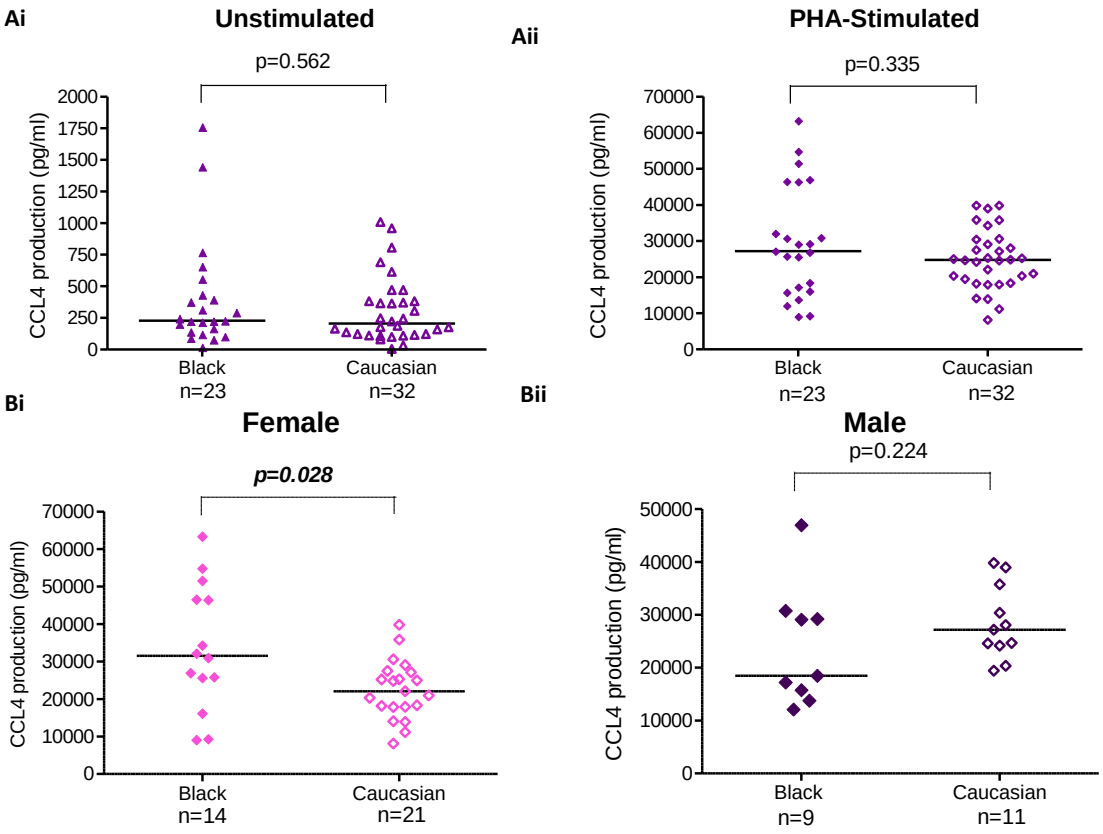



**Figure 3.3.1 No Significant differences in CCL4 protein production in two South African populations but Black females have a higher PHA-stimulated CCL4 production than Caucasian females.** CCL4 production from PBMC supernatants (Ai) prior to PHA stimulation and (Aii) post stimulation in the Black and Caucasian populations were compared. Higher levels of PHA-stimulated CCL4 production were observed in (Bi) Black females compared to Caucasian females) but there was no significant difference in protein production between (Bii) Black and Caucasian males. P values and number of participants are indicated.

### *3.3.2.Correlation of CCL4 production with CCL4L copy number*

Having a copy number of *CCL4L1* above the population median has been associated with protection against HIV-1 infection (Shostakovich-Koretskaya et al. 2009). When we compared the copy number of the *CCL4L* genes to unstimulated and PHA-stimulated CCL4 production for each population group, as well as the two groups combined, we failed to observe any significant correlations between production and copy number (PHA-stimulated protein production and *CCL4L* copy number correlations are represented in Figure 3.3.2). Subsequently, PHA-stimulated production levels below, above and equal to the population medians were examined (Figure 3.3.3). In the Black population, individuals with a *CCL4L* copy number less than the population median (median=4), displayed significantly lower PHA-stimulated CCL4 production than those with the median number of copies (p=0.018; Figure 3.3.3Ai). However, comparison of individuals with a *CCL4L* copy number less than the population median to those with *CCL4L* copies above the median, did not show a significant difference (p=0.300; Figure 3.3.3Ai). Also, compared to individuals with a median copy number, individuals with *CCL4L* copy numbers above the median did not have higher CCL4 production, in fact, the median CCL4 production was lower above the median copy number than at the median copy number, although not significantly lower (p=0.408; Figure 3.3.3Ai). Interestingly, when comparing copy number of *CCL4L1* and *CCL4L2* to CCL4 production in the Black population, although not statistically significant, inverse distributions or patterns were observed (Figures 3.3.3Aii and 3.3.3Aiii) compared to the *CCL4L* vs. CCL4 production comparison. For both *CCL4L1* and *CCL4L2*, lower production at the median copy number was observed compared to copy numbers below or above the median.

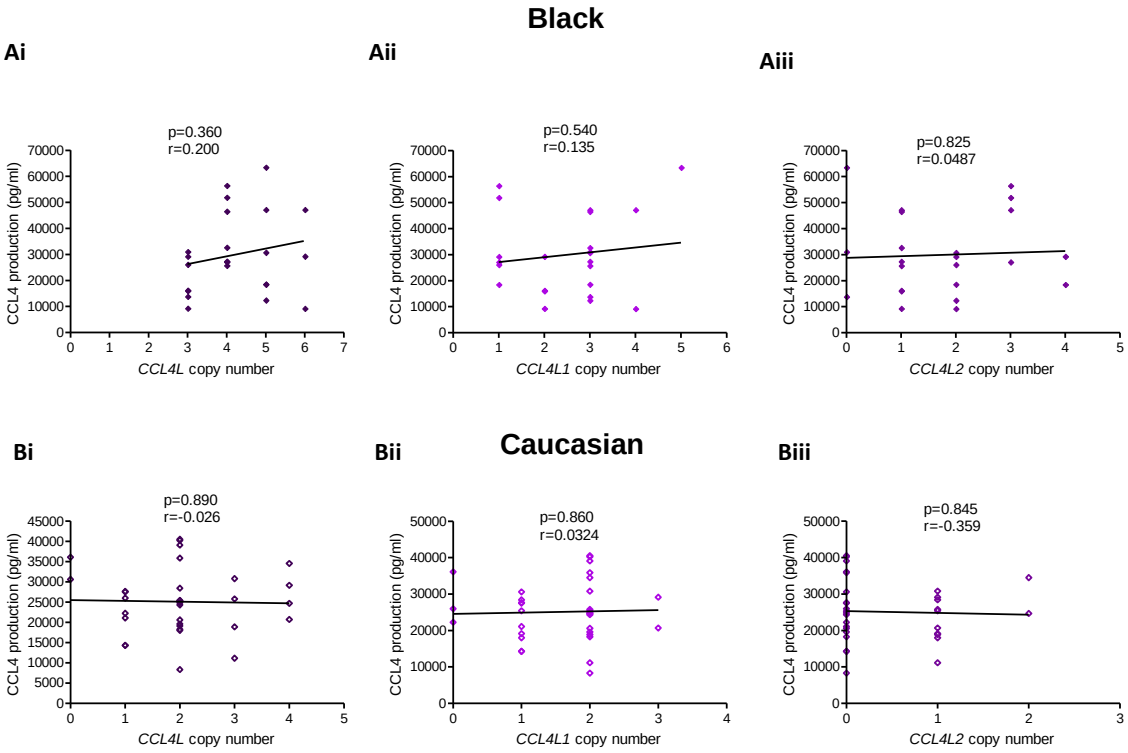



**Figure 3.3.2 Correlations of PHA-stimulated CCL4 production to (i) *CCL4L,* (ii) *CCL4L1* and (ii) *CCL4L2* copy number variants in (A) Black and (B) Caucasian South African cohorts.** P and r values are indicated.

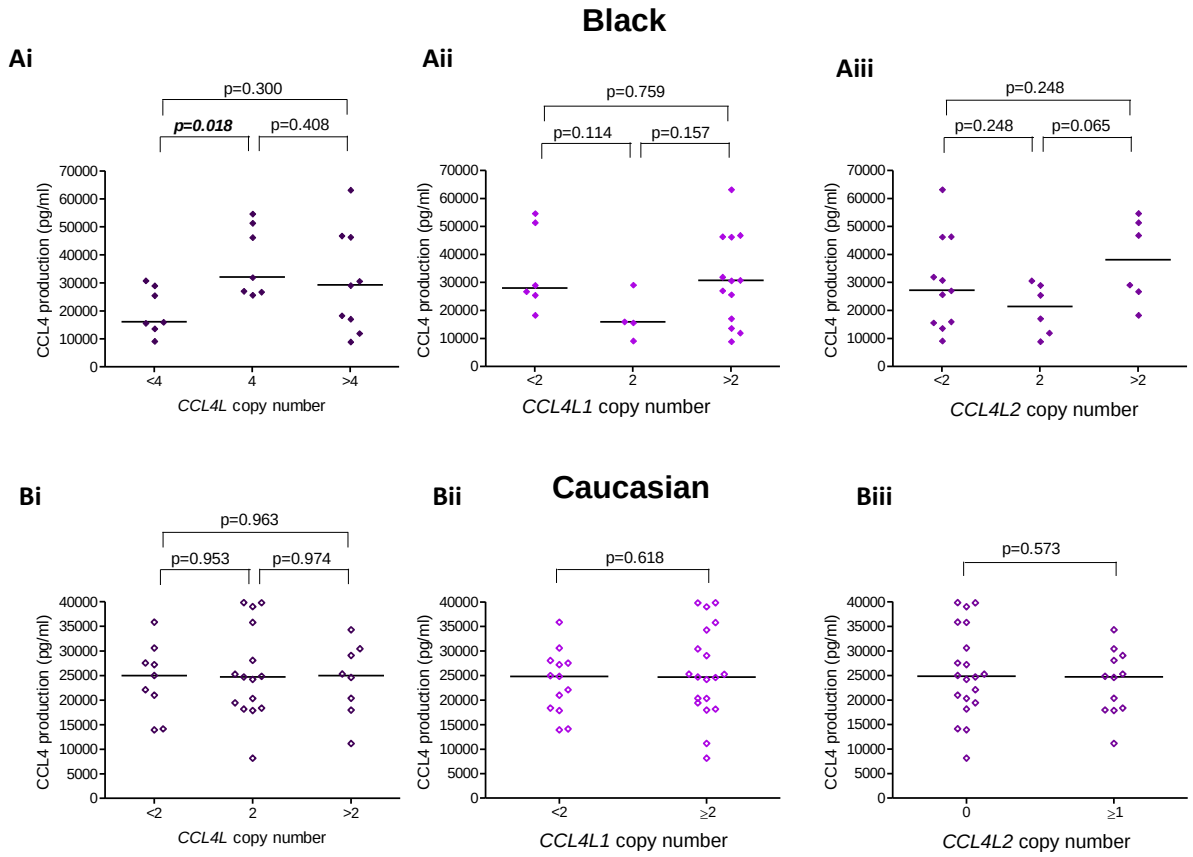



**Figure 3.3.3 PHA-stimulated CCL4 production relative to population specific median copy numbers in two South African populations.** In Black individuals, (Ai) *CCL4L,* (Aii) *CCL4L1,* (Aiii) *CCL4L2* copy numbers below, equal to and above the population specific medians, indicated on the figures, were compared to CCL4 production. CCL4 production is significantly lower below the median of 4 copies of *CCL4L* (Ai) In Caucasian individuals, CCL4 production below, equal to and above (Bi) the population median *CCL4L* copy number as well as below or equal to/above the population specific copy number medians for (Bii) *CCL4L1* and (Biii) *CCL4L2* are shown. P values are indicated.

Figure 3.3.3Bi represents the distribution of CCL4 production at, below and above the median for the *CCL4L* genes together in the Caucasian population. To examine the relationship between CCL4 production and *CCL4L1,* the cohort was divided into individuals with 0 copies of *CCL4L1* and with 1 or more copies *CCL4L1* (Figure 3.3.3Biii). To investigate the contribution of *CCL4L2* CNV in the same scenario, the cohort was stratified into groups having copy numbers at <2 or ≥2 (Figure 3.3.3Bii). This was necessary, as demonstrated by Figure 3.3.3, because of the conserved *CCL4L* copy number ranges observed in the Caucasian population. Evaluation of CCL4 production and copy number differences

of these two genes individually and collectively, in Caucasian individuals, did not reveal any significant differences or distribution patterns.

### *3.3.3.CCL4 production per CCL4L copy number*

Assuming an equal contribution of the two copies pdg *CCL4* genes to CCL4 protein production, the relationship of *CCL4L, CCL4L1* and *CCL4L2* copy number was compared with CCL4 production in an alternative way. We analysed the amount of PHA-stimulated CCL4 produced per gene copy number by dividing the CCL4 production data generated per individual by the copy number of each respective gene (i.e. *CCL4L*, *CCL4L1* and *CCL4L2*; Figure 3.3.4). Individuals with zero copies of *CCL4L*, *CCL4L1* or *CCL4L2* were therefore not included in this analysis. While the Caucasian population group had a significantly higher level of stimulated CCL4 production per *CCL4L* gene copy than the Black population group (Figure 3.3.4A; p=0.0183), this was not maintained when production per *CCL4L1* and *CCL4L2* copies were analysed individually (Figures 3.3.4B and 3.3.4C, respectively). Due to the significantly large CNV between the two population groups (with Caucasians having significantly lower copy numbers in all three genes (Figure 3.2.2)), and the fact that the CCL4 production did not differ between the two population groups (Figure 3.3.1Aii), it was not surprising that Caucasian individuals had increased CCL4 production per copy of *CCL4L* since the denominator (*CCL4L* copy number) is smaller in this group. The non-significant results for *CCL4L1* and *CCL4L2* analysed in the same manner however were more interesting. Next, to correct for the influence of the denominator, individuals from both populations groups, with either 3 or 4 copies of *CCL4L* and similar patterns of *CCL4L1* and *CCL4L2* distribution, were compared for protein production per copy of *CCL4L*, in order to establish whether the same number of genes present would have different effects on CCL4 production. Results showed that no significant differences were seen in either population (Figure 3.3.5; p=0.890).

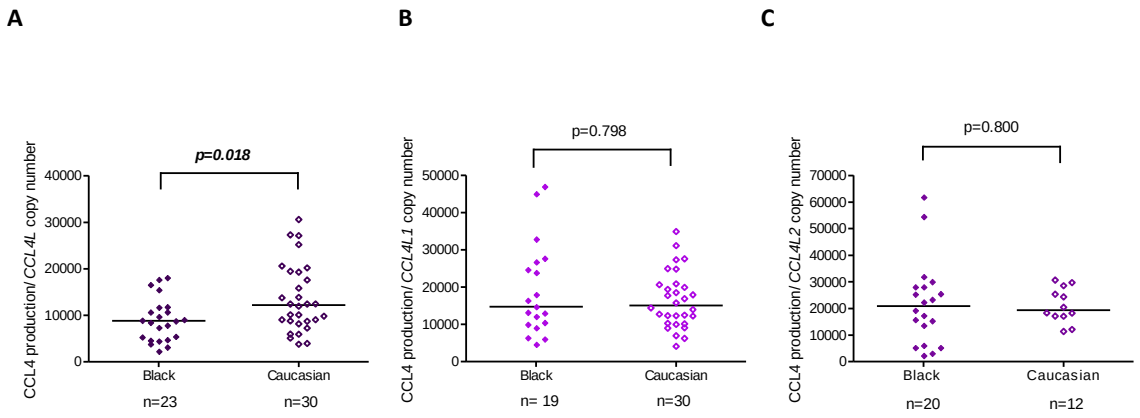



**Figure 3.3.4 PHA-stimulated CCL4 production (pg/ml) per gene copy number of the *CCL4L* encoding genes in the Black and Caucasian South African population.** The Black population has a significantly lower level of CCL4 production per (A) *CCL4L* gene copy, but the two populations did not differ greatly in the amount of CCL4 produced per copy of (B) *CCL4L1* and (C) *CCL4L2*. P values and number and sample size are indicated.

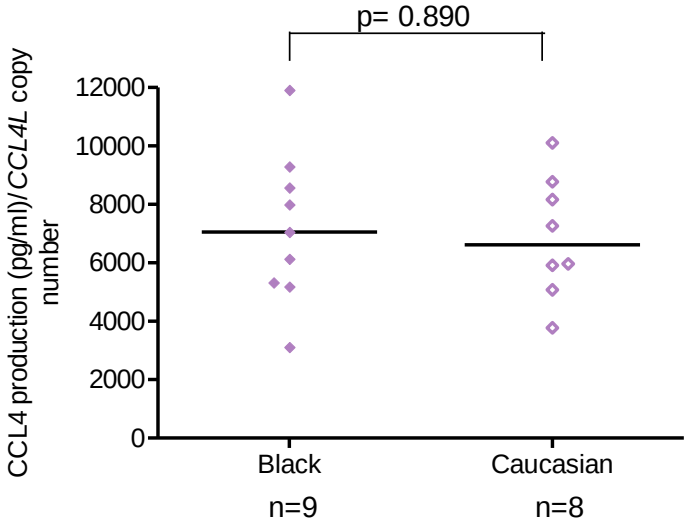



**Figure 3.3.5 Comparison of PHA-stimulated CCL4 production per *CCL4L* gene copies in Black and Caucasian individuals having equal number of copies of *CCL4L* and comparable patterns of distribution of *CCL4L1* and *CCL4L2*.**

### *3.3.4.Comparison of CCL3L and CCL4L gene copy number and CCL3 and CCL4 production*

*i. CCL3L and CCL4L copy number distribution*

Given that CCL3 and CCL4 are encoded by genes that are found closely clustered together in the 17q12 region of chromosome 17, and that in European (Caucasian) populations it is believed that *CCL3L* and *CCL4L* copy numbers correlate very strongly, having access to *CCL3* copy number and protein production data for the same cohort of 55 HIV-uninfected individuals, allowed us to investigate the relationship between these two chemokines. Significant positive correlations of *CCL3L* and *CCL4L* gene copy number (Figure 3.3.6Ai; $p<0.0001$) and *CCL3La* and *CCL4L1* copy number were observed in Black individuals (Figure 3.3.6Aii; $p=0.025$). There was however no significant correlation between *CCL3Lb* and *CCL4L2* (Figure 3.3.6Aiii). The correlations of the same genes were all significantly positive and stronger in the Caucasian cohort (Figure 3.3.6Bi; $p<0.0001$; Figure 3.3.6Bii; $p<0.0001$ and Figure 3.3.6Biii; $p=0.003$).

Since *CCL4L1* and *CCL3La* encode functional CCL4 and CCL3 proteins respectively, copy number distribution of these genes was compared in the Black South African cohort and although these gene copy numbers do significantly positively correlate (Figure 3.3.6 Aii), *CCL3La* copy number was found to be significantly higher than *CCL4L1* copy number (Figure 3.3.7A, $p<0.0001$). In the Caucasian population, however, no significant difference was observed (Figure 3.3.7B), not surprising given the stronger correlation between these two genes in the Caucasian population group compared to the Black population group (Figure 3.3.6Bii/Biii; $r=0.817$ vs. $r=0.467$, respectively).

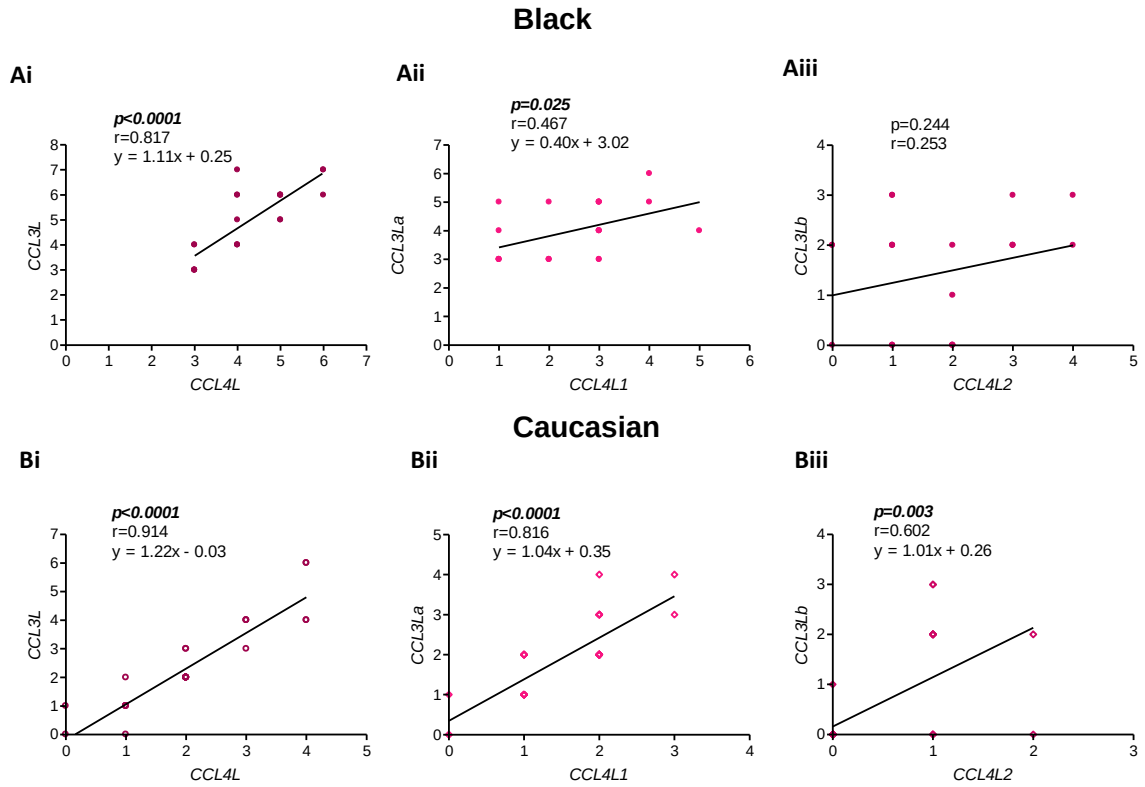



**Figure 3.3.6 Correlation of *CCL3L* and *CCL4L* copy numbers generated in individuals of the same cohorts of South African (A) Black and (B) Caucasian individuals.** Correlations of (i) *CCL3L* and *CCL4L*; (ii) *CCL3Lb* and *CCL4L1*, which encode protein forms that do not display classical CCL3 or CCL4 functionality; and (iii) *CCL3La* and *CCL4L2*, which encode functional isoforms of CCL3 and CCL4, are represented. P and r values are indicated and line equations when correlations are significant. CCL3L data provided by Picton et al. (2013).

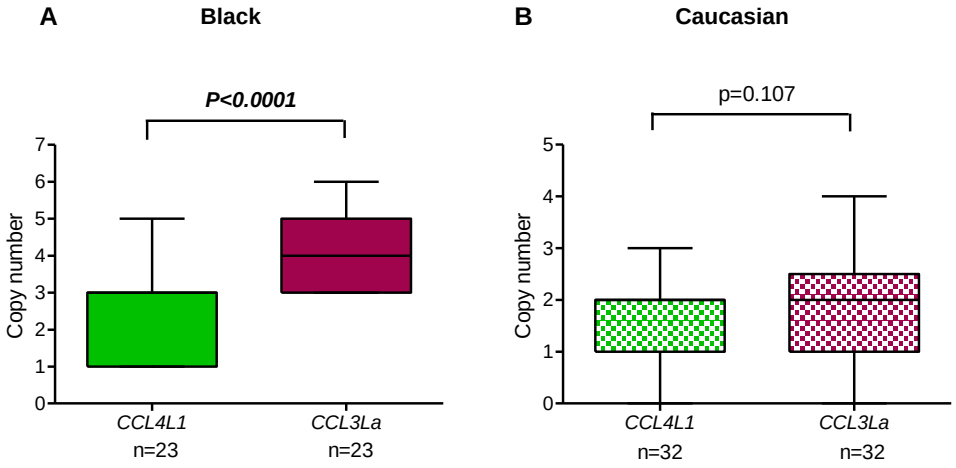



**Figure 3.3.7 Comparison of the copy number distribution of the *CCL4L1* and *CCL3La* genes (functional isoforms of the CCL4 and CCL3 genes, respectively) in a South African (A) Black and (B) Caucasian cohort.** P values and number of participants are indicated.

*ii. CCL3 and CCL4 protein production*

The correlation of CCL4 and CCL3 protein production was investigated separately in each of the population groups as well as in the groups combined (Figure 3.3.8). Results showed that in the Black population, correlation of PHA-stimulated CCL3 and CCL4 production did not reach significance (Figure 3.3.8B; p=0.119), and it is interesting to note that the copy numbers of the *CCL4L2* and *CCL3Lb* genes also did not significantly correlate in this population group (Figure 3.3.6Aiii). In contrast, when correlations between PHA-stimulated CCL3 and CCL4 production were examined in the Caucasian population, there was a significant strong correlation (Figure 3.3.8C, p=0.0001). Analysing the relationship of the two proteins in the combined population (Figure 3.3.8A) also showed a significant positive correlation (p=0.001).

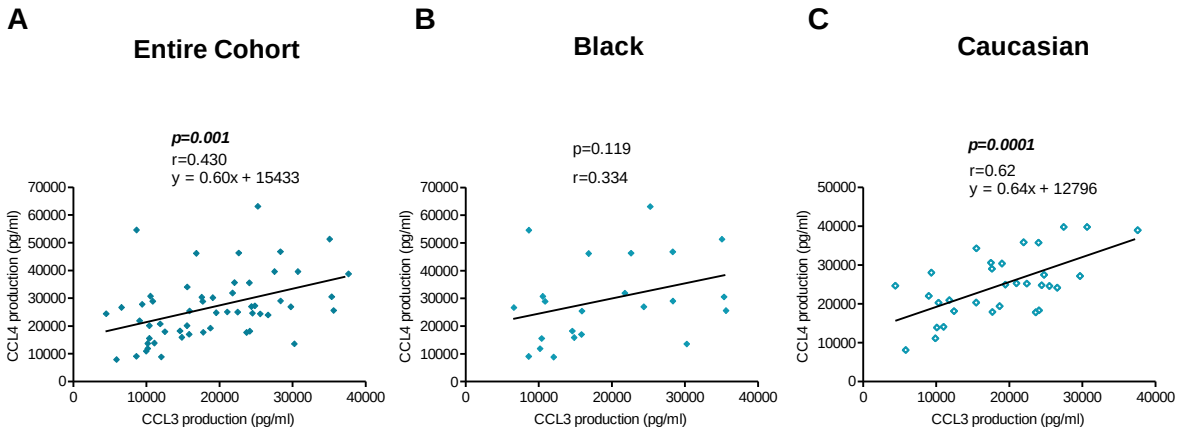



**Figure 3.3.8 Correlations of CCL4 production to CCL3 production in a cohort of (A) 55 HIV-1-uninfected individuals comprising of (B) Black and (C) Caucasian cohorts.** P values and r values are indicated; and line equations where p values are significant. CCL3 production data was generated by Picton et al. (2013).

We then compared the levels of CCL3 and CCL4 production in both unstimulated and PHA-stimulated PBMCs from the entire uninfected cohort (Figure 3.3.9A), the Black population group (Figure 3.3.9B) and the Caucasian population group (Figure 3.3.9C). We found that CCL4 production was significantly

higher than CCL3 production in both unstimulated and PHA-stimulated PBMCs in all three categories (Figure 3.3.9).

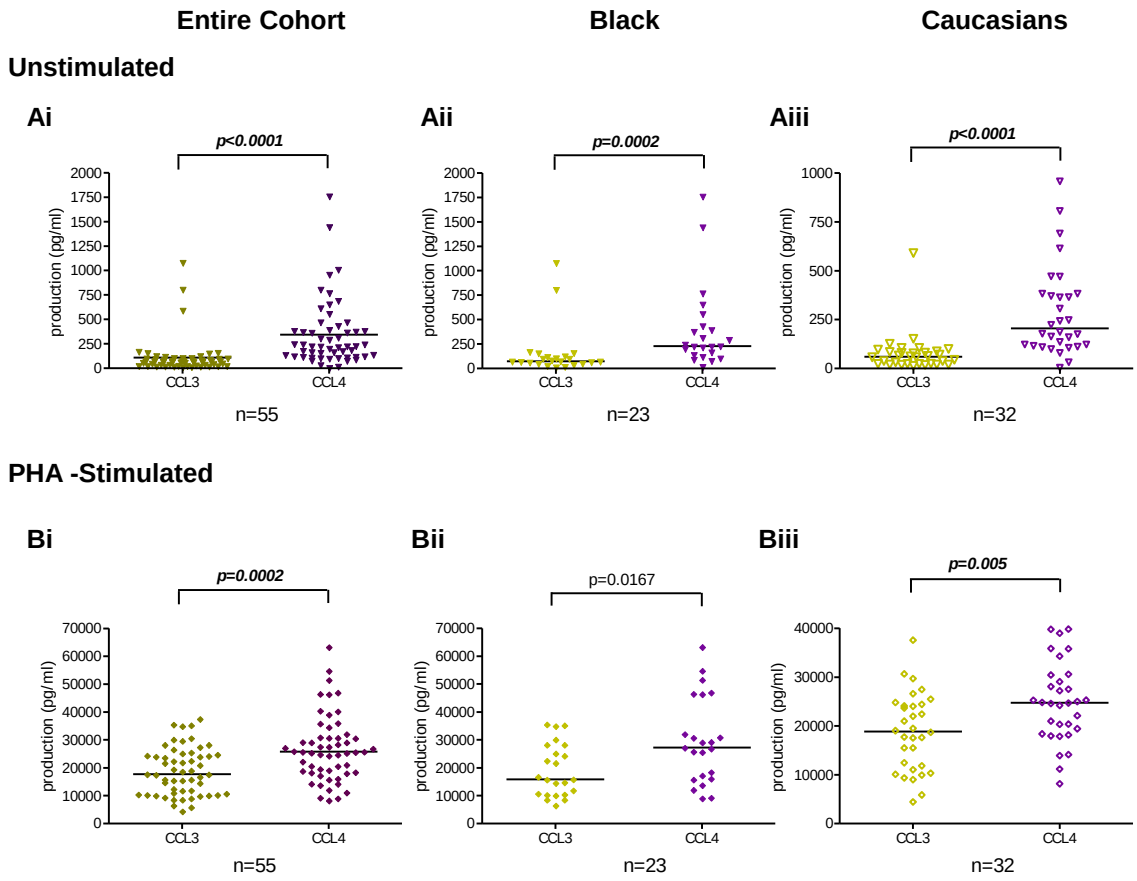



**Figure 3.3.9 CCL3 and CCL4 protein production measured from (A) unstimulated PBMC supernatants and (B) PHA-stimulated supernatants, were compared in the cohort of (i) 55 HIV-1-uninfeced South Africans, comprising of (ii) 23 Black and (iii) 32 Caucasian individuals.** P values are indicated.

## 3.4. CCL4 AND HIV-1-INFECTED LONG-TERM NONPROGRESSORS (LTNPS)

### *3.4.1.CCL4 copy number and production comparisons between uninfected Black individuals and Black LTNPs*

Since copy number distribution of the *CCL4L* genes was found to differ between the uninfected Black and Caucasian groups, comparisons of copy number distribution between the uninfected cohort and the cohort of LTNPs also had to be stratified by race. The Caucasian LTNP group however, only consisted of four individuals. Comparison of *CCL4L*, *CCL4L1* and *CCL4L2* copy number was performed between the HIV-1 uninfected Black population group and the Black LTNPs only, and as can be seen in Figure 3.4.1, no significant differences were observed in all three comparisons. Furthermore when we looked at the correlation of copy number of the CCL4-encoding genes and CCL4 production in the 9 Black LTNPs, similarly to what we saw in the uninfected Black and Caucasian populations, there were no significant correlations in all three correlations performed (Figures 3.4.2A, B and C).

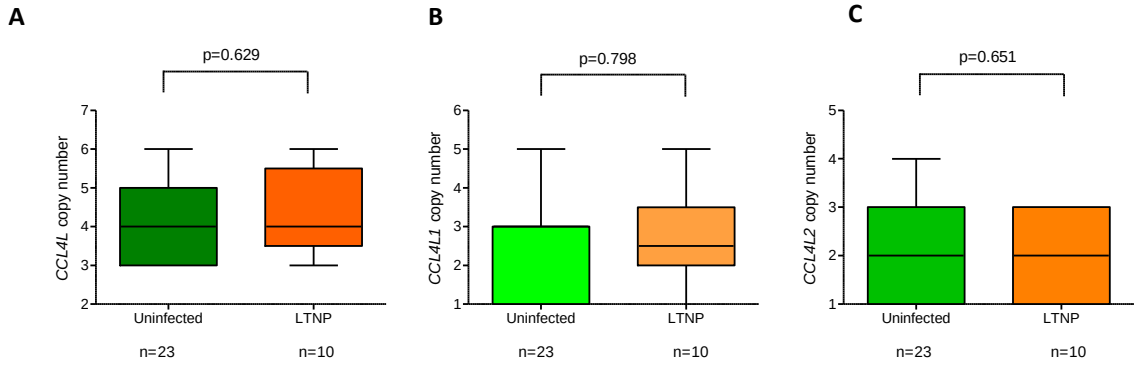



**Figure 3.4.1 Neither (A) *CCL4L*, (B) *CCL4L1* nor (C) *CCL4L2* copy number distribution differed significantly when a cohort of HIV-1-uninfected Black individuals was compared to a cohort of LTNPs.** P values and number of individuals is indicated.

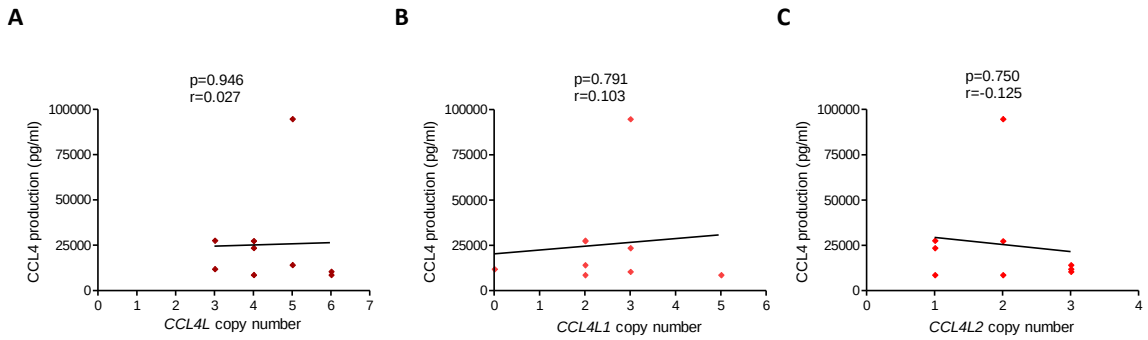



**Figure 3.4.2 *CCL4L* gene copy numbers do not correlate with CCL4 production in a cohort of 10 Black South African LTNPs.**

## *3.4.2. CCL4 production in LTNPS differs from HIV-1-uninfected individuals*

One individual in the Black LTNP cohort displayed CCL4 production levels that were much higher than the rest of the cohort (Figures 3.4.3Ai and Bi, labelled outlier) and was therefore excluded from analyses when population groups were investigated individually. When comparing CCL4 protein produced by both unstimulated and PHA-stimulated PBMCs of LTNPs and HIV-1-uninfected individuals, we found that the total LTNPs (Black and Caucasian) produced significantly lower CCL4 levels compared to the total uninfected cohort (unstimulated production: Figure 3.4.3Ai; p=0.007; p=0.002 excluding outlier; PHA-stimulated production: Figure 3.4.3Bi; p=0.003; p=0.0005 excluding outlier). This significant lower level association was maintained for both unstimulated and PHA-stimulated CCL4 production when the Black (unstimulated: Figure 3.4.3Aii; p=0.033; PHA-stimulated: Figure 3.4.3Bii; p=0.011) and Caucasian cohorts (unstimulated: Figure 3.4.3Aiii; p=0.019; PHA-stimulated: Figure 3.4.3Biii; p=0.008) were compared individually. Interestingly when similar comparisons investigating CCL3 production in LTNPs were carried out, CCL3 production was significantly lower in the LTNP cohort when PHA-stimulated levels were compared and were also maintained when Black and Caucasian cohorts were compared individually, however, unstimulated

CCL3 production did not differ significantly when whole cohorts were compared as well as when divided based on race (Picton et al, unpublished data).

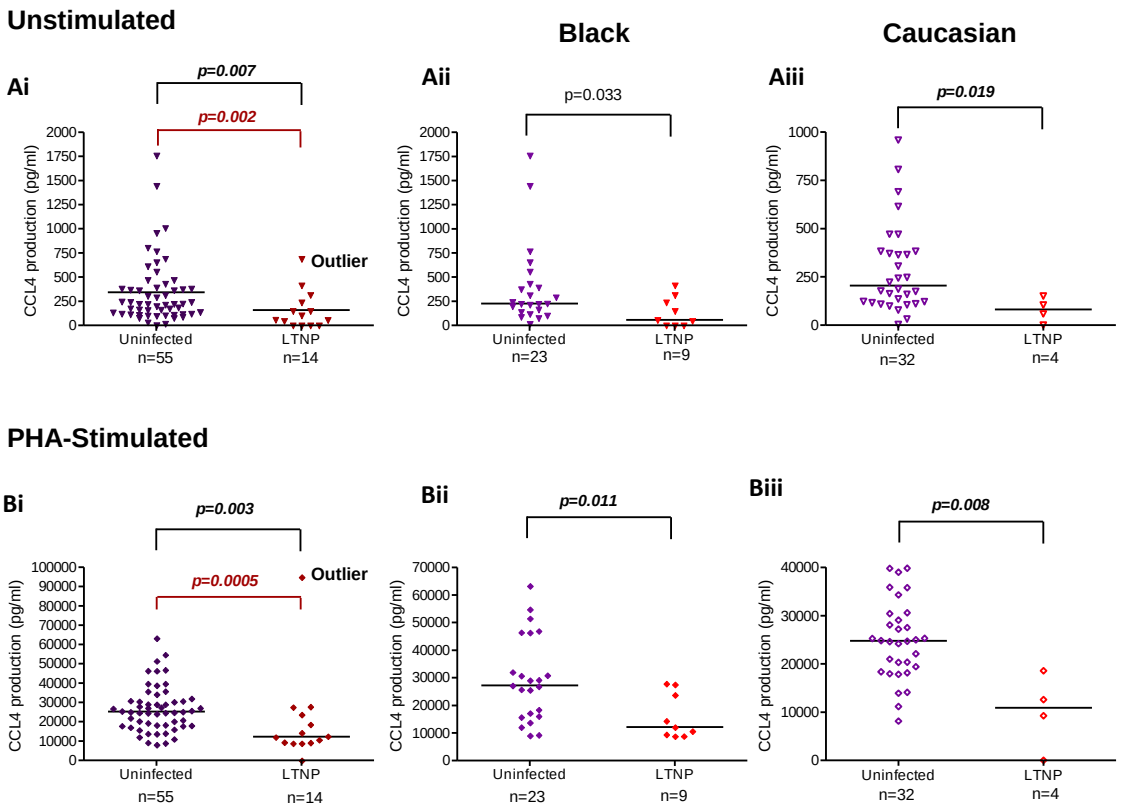



**Figure 3.4.3 CCL4 production from (A) unstimulated and (B) PHA-stimulated peripheral PBMCs isolated from HIV-1-uninfected individuals and LTNPs.** P values generated from comparisons between an (i) HIV-1-uninfected cohort and a cohort of LTNPs were calculated with both the inclusion of the production level data of an outlier, indicated in figure, and excluding the outlier (p values in red). Comparisons of uninfected and LTNP individuals stratified on race, that is, (ii) Black and (iii) Caucasian were conducted, excluding the outlier data from the Black LTNP group.

## *3.4.3.CCL4 protein production is not higher than CCL3 protein production in LTNPs*

As shown above (page 60), both unstimulated and PHA-stimulated CCL4 production was shown to be significantly higher than corresponding CCL3 production in the total uninfected cohort (Figure 3.3.9i) and when the Black (Figure 3.3.9iiB) and Caucasian (Figure 3.3.9iii) cohorts were examined

individually. This higher level of CCL4 production compared to CCL3 production has also been reported in other studies. However, in the case of LTNPs, unstimulated and PHA-stimulated CCL3 and CCL4 protein levels in the two populations do not significantly differ (Figure 3.4.4A, p=0.0663, Figure 3.4.4B, p=0.135, respectively). Furthermore, CCL3 and CCL4 protein production displays a strong positive correlation in LTNP individuals (Figure 3.4.5, p<0.0001) and although positive, the correlation between CCL3 and CCL4 was not significant in the uninfected Black cohort (Figure 3.3.6B).

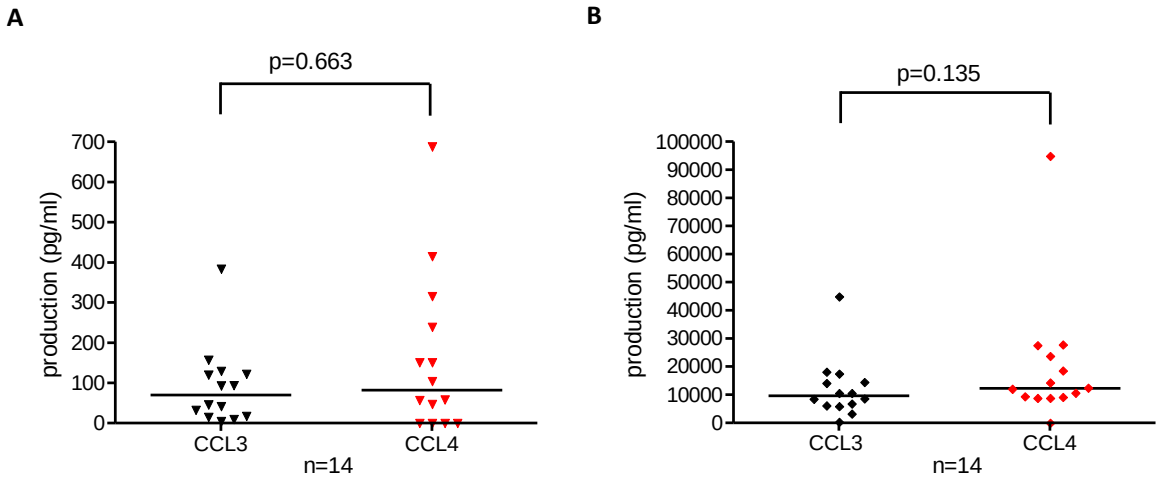



**Figure 3.4.4 No significant differences in (A) unstimulated and (B) PHA-stimulated CCL3 and CCL4 production from peripheral PBMCs.** P values and number of individuals are indicated. CCL3 production data was generated by Picton et al (unpublished).

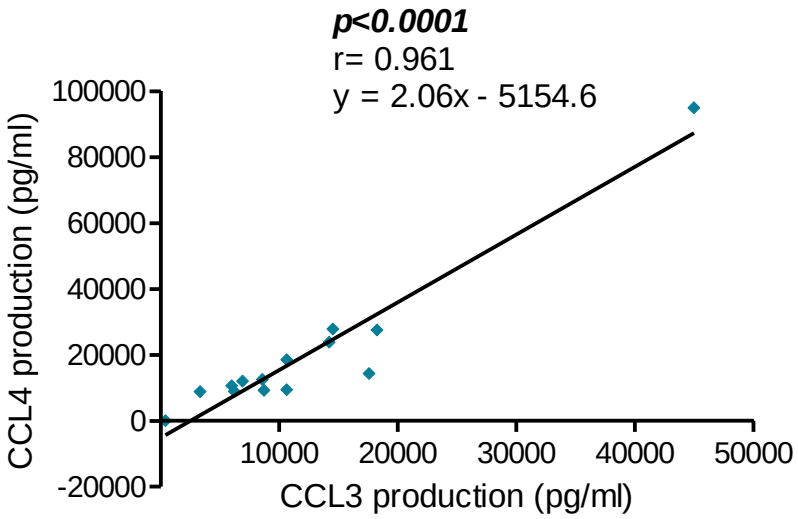



**Figure 3.4.5 Correlation of PHA-stimulated CCL4 and CCL3 production in a group of HIV-1-infected LTNPs (n=14).** CCL3 production data was generated by Picton et al (unpublished).

### *3.4.4.CCL4 and CCL3 production and HIV-1 viral load (VL)*

Stratification of the Black LTNP cohort into individuals who have VLs less than and greater than 400 RNA copies/ml and analysis of CCL4 and CCL3 production with respect to VL, revealed that in the case of unstimulated CCL3 production, Black LTNPs with lower VLs (<400 RNA copies/ml) did not significantly differ from Black LTNPs with VLs above RNA 400 copies/ml (Figure 3.4.6Ai; p=0.110). However, there is a significantly lower level of unstimulated CCL4 production in Black LTNPs with a VL of less than 400 RNA copies/ml compared to Black LTNPs with VLs greater than 400 RNA copies/ml (Figure 3.4.6Bi; p=0.016). When PHA-stimulated protein production was compared in these same groups, Black LTNPs with VLs below 400 RNA copies/ml produced significantly less CCL3 (Figure 3.4.6Aii; p=0.016) compared to Black LTNPs with VLs greater than 400 RNA copies/ml, meanwhile PHA-stimulated CCL4 production shows a strong trend (Figure 3.4.6Bii; p=0.064) of lower production in individuals with VLs less than 400 RNA copies/ml compared to greater than 400 RNA copies/ml individuals. Comparison of the VL-stratified LTNP groups to their corresponding uninfected cohort

with respect to chemokine production (i.e. CCL4 and CCL3), revealed that it is the less than 400 RNA copies/ml LTNPs that are the lower protein producers in both the unstimulated and PHA-stimulated scenarios, with significance in the unstimulated CCL4 and PHA-stimulated CCL3 and CCL4 comparisons (Figures 3.4.6Ai, Bi and Bii). The LTNPs having VLs greater than 400 RNA copies/ml, did not differ in either CCL3 or CCL4 production compared to their corresponding uninfected cohort (Figure 3.4.6).

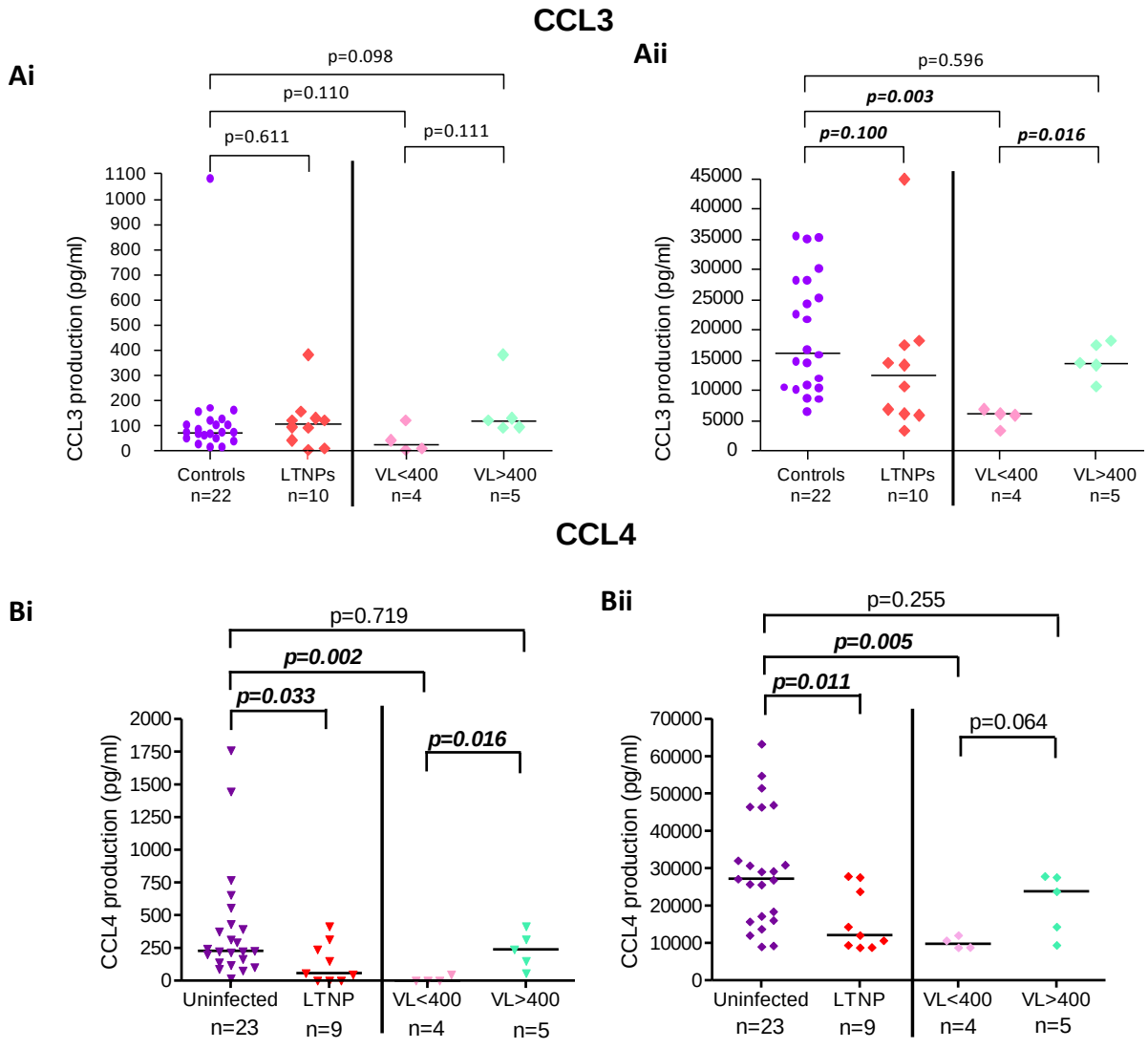



**Figure 3.4.6 Comparisons of (i) Unstimulated and (ii) PHA-stimulated (A) CCL3 and (B) CCL4 production between HIV-uninfected Black South Africans and HIV-infected Black LTNPs.** Comparisons of LTNPs are further stratified into individuals with viral loads below and above 400 RNA copies/ml. P values and sample size are indicated. Note, in Bi and Bii an outlier had been removed from the total LTNP group, i.e. n=9, whereas in Ai and Aii (CCL3), the outlier in included in the corresponding group.

## 3.5. GENETIC CHARACTERIZATION OF THE *CCL4* GENE

### *3.5.1. Single nucleotide polymorphisms*

Amplification and sequencing of the *CCL4* gene in the HIV-1-uninfected cohort allowed for the identification of a number of genetic variants. In total, 77 SNPs were identified in a 4.7kb fragment containing the *CCL4* gene, in a cohort of Black and Caucasian individuals. Sixty-six SNPs in the *CCL4* gene and its flanking regions were identified in Black South African individuals while in the Caucasian population, 38 SNPs were identified. Of these, only 22 and 12 SNPs were identified within the transcribed region of the CCL4 gene (i.e. excluding promoter and flanking regions) in Black and Caucasian individuals respectively. SNP position, nucleotide changes and minor allelic frequencies are indicated in Table 3.5.1. Exons, introns and UTRs are also indicated in Table 3.5.1 and the adjoining Figure 3.5.1A, which schematically represents *CCL4* gene structure. The percentage of SNPs unique to each population group as well as the overlapping percentages are indicated in Figure 3.5.1B.

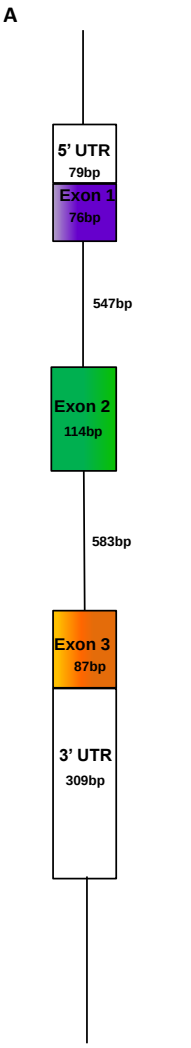
From the alignment of the 55 generated sequence assemblies to the *CCL4* reference sequence and its flanking regions retrieved from GenBank (Benson et al. 2006), the major allele from all SNPs was designated as the WT allele.

There was a single SNP locus at which the minor allele frequency was high in the Black South African group (45.65%; rs116657123, Table 3.5.1; position 663). In this case, even though one allele (G) was slightly in the majority, just to be certain, the WT allele was assigned by aligning the sequences from our population groups to the *CCL4* sequence of *P. troglodytes*. The ancestral allele (G) was thus regarded as the WT, and was interestingly more prevalent in the Caucasian population. The reported frequency of the A allele in the 1000 Genomes Project database, in the combined European

populations, is 30%, which is higher than the frequency observed in Caucasian individuals in this study (20%), while in the Yoruba population, a Sub-Saharan population which is most likely to be more genetically similar to the South African Black population than any other 1000 Genomes populations, this polymorphism occurred at a frequency of 31% (Abecasis et al. 2012) while it was observed at a frequency of 46% in black individuals in this study. The majority of SNPs were located in intronic and promoter regions, while the exons and UTRs were not rich in SNPs (Table 3.5.1). Exonic polymorphisms within the Black population included two SNPs in exon 1, a single SNP in exon 2 and two SNPs in exon 3. In Caucasian individuals, one SNP in exon 1 and two SNPs in exon 3 were identified. The two SNPs in exon 3 were common between the two populations. Only one SNP detected was not previously reported (position: -1063) and was identified in the Black population in the region upstream from the *CCL4* promoter region.

**Table 3.5.1 SNPs and indels identified in the South African Black and Caucasian populations**

| | dbSNP Acession Number | Position | Base Change (wt/mt) | Minor Allele Frequency (%) | |
|---|---|---|---|---|---|
| | | | | Black n=23 | Caucasian n=32 |
| | rs143457996 | -1095 | G/A | 4,35 | - |
| | newly identified | -1063 | C/G | 2,17 | - |
| | rs540786592 | -973 | C/T | - | 1,56 |
| | rs1160085153 | -953 | C/T | 6,52 | - |
| | rs140335931 | -914 | C/G | - | 4,69 |
| | rs10491121 | -884 | G/A | - | 34,38 |
| | rs145502799 | -832 | C/T | 6,52 | - |
| | rs1719140 | -807 | A/T | 17,39 | - |
| Promoter | rs1719141 | -770 | G/T | 10,87 | 28,13 |
| | rs9895259 | -764 | G/C | 6,52 | - |
| | rs79409766 | -706 | T/C | 15,22 | - |
| | rs79409767 | -678 | C/G | 10,87 | 28,13 |
| | rs113682499 | -665 | G/A | 2,17 | - |
| | rs143637961 | -625 | G/A | 4,35 | - |
| | rs113245686 | -603 | INDEL -/T | 6,52 | - |
| | rs139348900 | -534 | INDEL -/C | 6,52 | - |
| | rs9895812 | -502 | G/C | 6,52 | - |
| | rs9896211 | -392 | G/A | 15,22 | 6,25 |
| | rs62079738 | -319 | G/A | - | 3,13 |
| | rs1634514 | -302 | T/A | 10,87 | 28,13 |
| | rs555131559 | -274 | T/C | 2,17 | - |
| | rs191820992 | -177 | A/G | 6,52 | - |
| | rs577987679 | -164 | G/C | 2,17 | - |
| | rs143208576 | -97 | G/A | 2,17 | - |
| | rs1049746 | -9 | G/C | 10,87 | 28,125 |
| | rs147511844 | -1 | C/T | 2,17 | - |
| 5' UTR | rs1049750 | 68 | T/G | 17,39 | - |
| Exon 1 | rs149488954 | 94 | G/C | 2,17 | - |
| | rs143966312 | 141 | C/T | - | 1,56 |
| | rs150142971 | 155 | A/G | 4,35 | - |
| Intron 1 | rs138567234 | 156 | G/A | 8,70 | - |
| | rs1719144 | 184 | G/A | 4,35 | 31,25 |
| | rs4606761 | 216 | G/T | 6,52 | - |
| | rs7218357 | 367 | C/T | 6,52 | - |
| | rs116657123 | 633 | G/A | 45,65 | 20,31 |
| | rs55906345 | 665 | C/T | 13,04 | - |
| | rs184395846 | 666 | G/A | 6,52 | - |
| Exon 2 | rs1719146 | 742 | C/T | 13,04 | - |
| Intron 2 | rs541137737 | 834 | INDEL -/G | 2,17 | - |
| | rs2905542 | 870 | A/T | 2,17 | - |
| | rs150004330 | 875 | A/G | 2,17 | - |
| | rs1719147 | 906 | G/A | 36,96 | 20,31 |
| | rs72833526 | 1018 | G/A | - | 48,44 |
| | rs1634516 | 1023 | G/C | 10,87 | 28,13 |
| | rs1719148 | 1078 | A/G | 19,57 | 28,13 |
| | rs1634517 | 1184 | C/A | 23,91 | 31,25 |
| | rs9891632 | 1279 | T/C | 8,70 | - |
| | rs1719149 | 1362 | T/C | 2,17 | 28,13 |
| | rs1719150 | 1367 | A/G | 2,17 | 28,13 |
| Exon 3 | rs1049807 | 1444 | A/G | 15,22 | 28,13 |
| | rs1719152 | 1445 | T/A | 10,87 | 29,69 |
| | rs184099447 | 1653 | T/C | 2,17 | - |
| 3' UTR | rs1049833 | 1704 | G/C | 8,70 | - |
| | rs148133700 | 1725 | C/T | - | 4,69 |
| | rs1634518 | 1810 | A/G | - | 26,56 |
| | rs1719153 | 2010 | A/T | 10,87 | 28,13 |
| | rs113909899 | 2063 | C/T | 4,35 | - |
| | rs146185601 | 2164 | C/G | 8,70 | 4,69 |
| | rs3216921 | 2197 | INDEL: -TAT | - | 14,06 |
| | rs139373922 | 2319 | C/T | 2,17 | - |
| | rs187291062 | 2489 | C/A | - | 3,13 |
| | rs1634519 | 2615 | G/A | 10,87 | 28,13 |
| | rs28547364 | 2709 | G/A | 17,39 | 6,25 |
| | rs2522126 | 2720 | C/G | 4,35 | 28,13 |
| | rs2687507 | 2745 | C/T | 10,87 | 29,69 |
| | rs12939657 | 2758 | G/A | 10,87 | - |
| | rs142188747 | 2776 | G/A | 2,17 | - |
| | rs113485141 | 2836 | G/A | 2,17 | - |
| | rs9892580 | 2892 | T/C | 23,91 | - |
| | rs142438706 | 2937 | T/C | 2,17 | - |
| | rs116923441 | 2962 | C/G | - | 3,13 |
| | rs17679451 | 3099 | G/A | 4,35 | 4,69 |
| | rs373596785 | 3309 | T/G | 2,17 | - |
| | rs76690362 | 3316 | G/A | 4,35 | 4,69 |
| | rs62079742 | 3326 | T/C | - | 14,06 |
| | rs141572011 | 3364 | G/A | 4,35 | 4,69 |
| | rs145428208 | 3423 | T/A | 2,17 | - |
| | rs1619526 | 3426 | C/T | 13,04 | 28,13 |
| | rs1619600 | 3456 | A/G | 6,52 | 26,56 |
| | rs17617372 | 3480 | G/A | 2,17 | 4,69 |



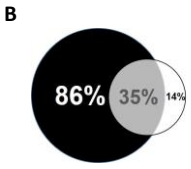




**Figure 3.5.1 (A) Schematic representation of the CCL4 gene**. (B) Venn diagram indicating the percentage of SNPs identified for the Black (black area) and Caucasian (white area) population groups and the percentage of SNPs that are shared by both race groups (grey area)

### *3.5.2. Influence of exonic SNPs on protein translation*

Of the total, six SNPs were identified in exons, three of which, were non-synonymous mutations (rs143966312, rs150142971, rs1719152). The resulting amino acid changes from these polymorphisms in the protein coding region are: Ala21Val, Met26Val and Ser80Thr (Figure 3.5.2, amino acid codes are listed in Appendix C). In order to, putatively predict possible conformational and therefore functional implications of these amino acid changes to the CCL4 protein, the three substitutions were run through the Poly-Phen 2-prediction of functional effects of human nsSNPs data base (Adzhubei et al. 2010). The Ala21Val, Met26Val and the Ser80Thr amino acid substitutions were predicted to be 'benign', i.e. not affecting protein stability and function, with scores of 0.002, 0.000 and 0.000 respectively on a scale that ranges from 0 to 1. None of the exonic SNPs occurred within the codons for the conserved cysteine residues or amino acids involved in binding (amino acids which form part of the conserved CC chemokine family structure are indicated by amino acids underlined in grey in Figure 3.5.2). The Ala21Val substitution does occur in the signal peptide binding site (indicated by amino acids in a black box in Figure 3.5.2) and the SNP was only identified in a single individual in the Caucasian cohort (Table 3.5.1; allelic frequency 1.56%).

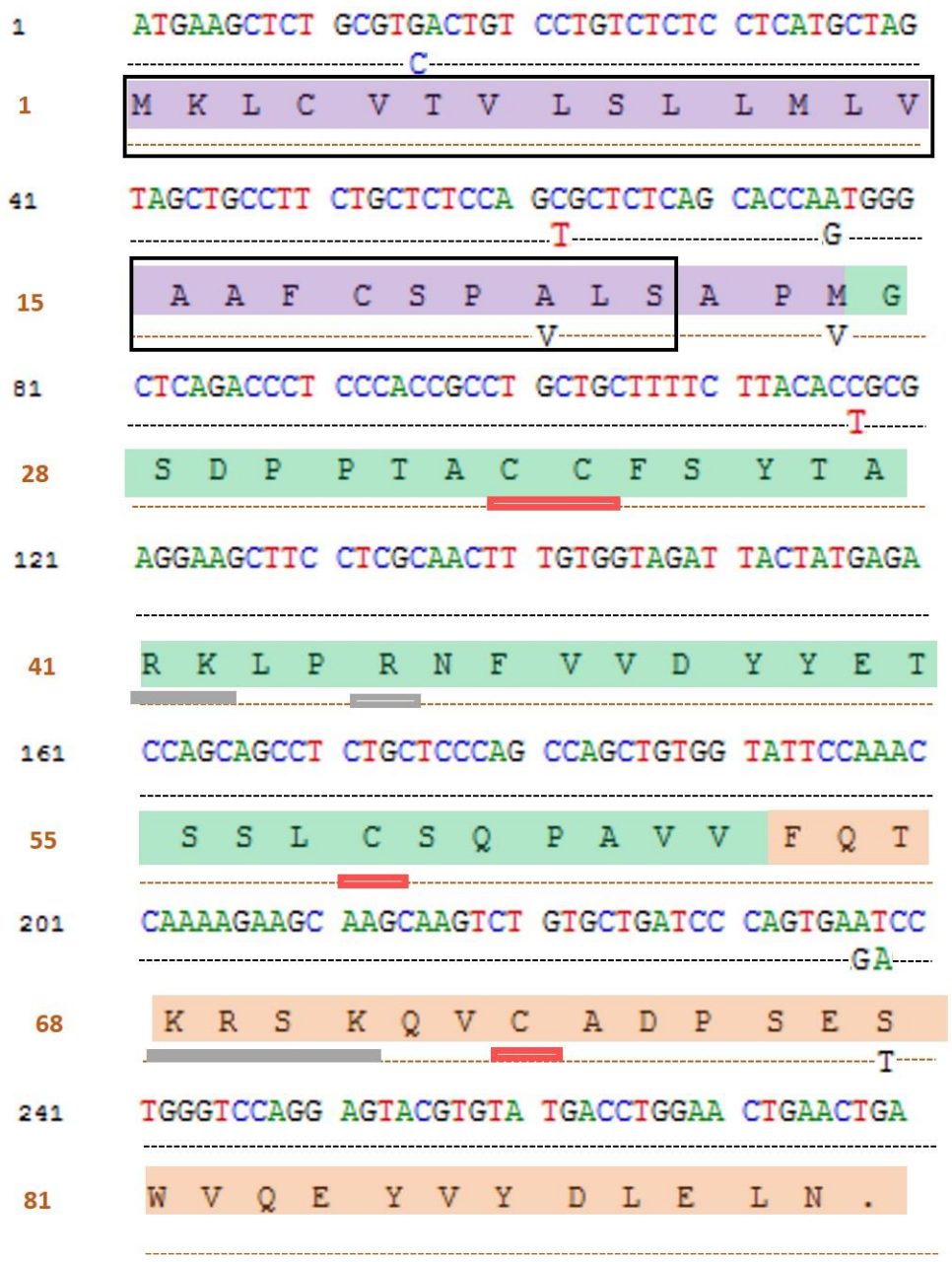



**Figure 3.5.2 Influence of CCL4 exonic SNPs on CCL4 amino acid sequence.** Lines of sequence are followed by the corresponding amino acid sequence, the '.' representing a stop codon. A dotted line under the respective sequences indicated homology and lack thereof is indicated by the base substitution at the specific site. Shading behind amino acid sequences denote the exon from which that particular sequence was transcribed, that is, purple shading: exon1, green shading: exon 2, orange shading: exon 3. The first few bases enclosed in a box represent the CCL4 signal peptide. Amino acids underlined in grey are involved in binding and the conserved cysteine residues are underlined in red.

### *3.5.3. Insertions/Deletions*

Four indels were identified when sequence data was analysed: rs113245686, rs139348900, rs541137737, rs3216921 at positions -603, -534, 834 and 2197 respectively (Table 3.5.1). Three of the four indels were identified in Black individuals only, one of which only occurred in a single individual. All the indels, which were reported as deletions, had been previously described in dbSNP.

The indel at position 2197 was in a region of AT repeats which meant that determining which bases were deleted was difficult as the repetitive nature of the region would influence sequence alignment. Even within dbSNP, there are multiple reports of a deletion in this area, overlapping by a few bases and, while most reports agree that it is a 3bp deletion, the exact nucleotide sequence of the deleted portion is unclear. One such alternate identifier for the rs3216921 deletion is rs147387560 and population data for this SNP available from the 1000 Genomes Project (Abecasis et al. 2012) suggests that in the Yoruba population, this deletion occurs at a frequency of 2%. The South African Black individuals in this study did not have this deletion at all. In European populations (1000 Genomes Project), the recorded frequency of this indel is 44%, which is strikingly lower than the 14% prevalence which was identified in South African Caucasian individuals.

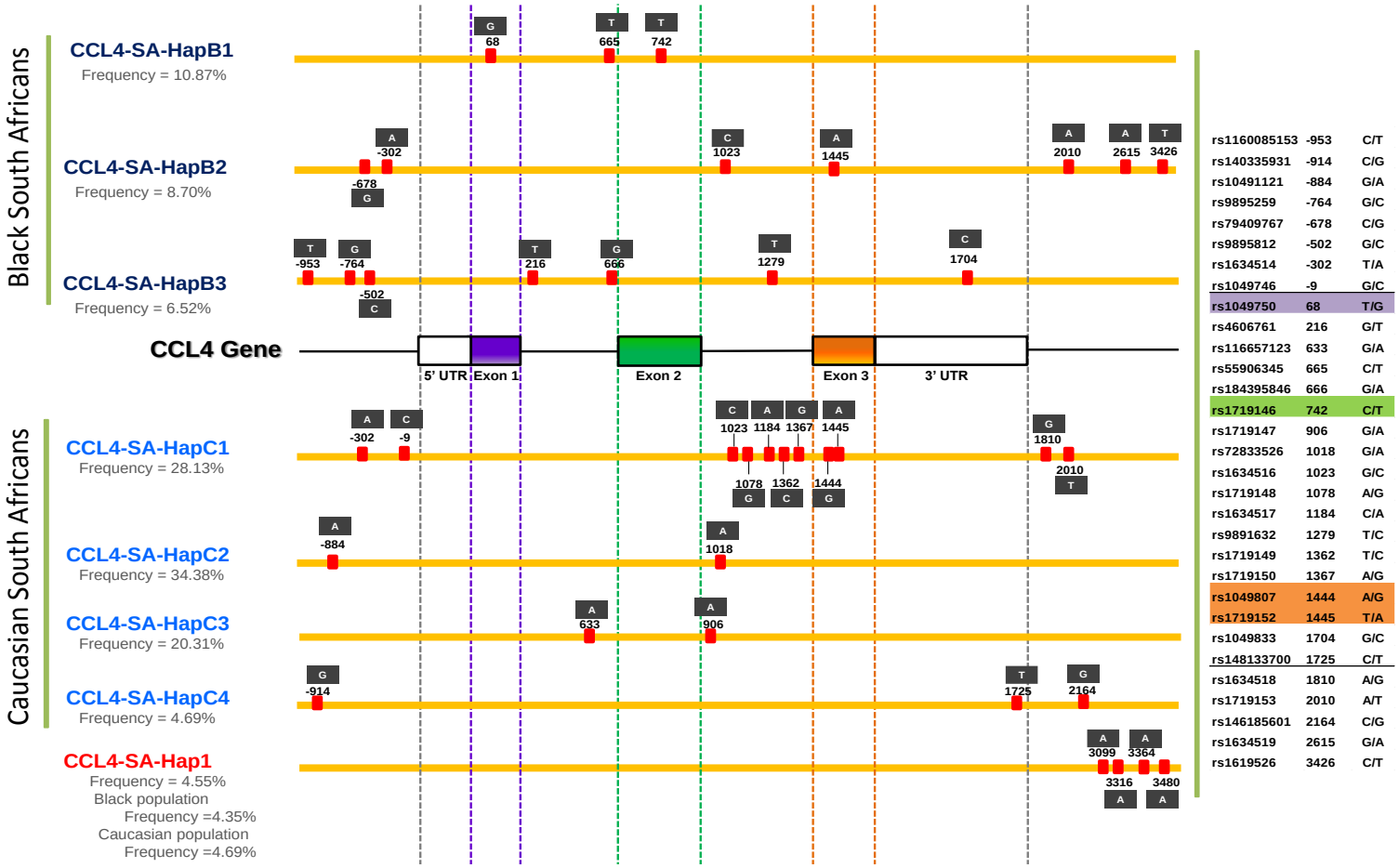



**Figure 3.5.3 Schematic representation of putative CCL4 haplotypes identified in the South African Black and Caucasian populations.** All CCL4-SA-HapB patterns were identified in Black individuals while all CCL4-SA-HapC patterns were identified in the Caucasian cohort. The haplotype CCL4-SA-Hap1 was identified in both populations. SNP positions, SNP identification numbers, base changes as well as position relative to the various gene elements are indicated in the figure and the adjacent list. Haplotype frequencies in the Black and Caucasian population are represented.

### *3.5.4. Putative haplotypic structure of CCL4 and flanking regions*

Upon visual examination of the SNP/indel data across the CCL4 gene region in both populations, we observed distinct patterns between various variations which were defined as putative haplotypes. Four putative haplotypes were identified in the Caucasian population and three haplotypes were identified in the Black population (Figure 3.5.3). Only a single haplotype (CCL4-SA-1) was shared between the Black and Caucasian population (Figure 3.5.3). Primer positioning for CCL4 meant that a very large fragment had to be amplified, which contained large upstream and downstream regions. SNP data for these regions have thus been included and we identified that haplotypes extended into these highly polymorphic regions.

Due to the large number of polymorphic loci in Black individuals, it was difficult to putatively assign haplotypes. Analysis of the Black population SNP/indel data using Haploview software shown in Figure 3.5.4, shows very few blocks of linked variants, indicated by red squares. The pattern observed in the Caucasian population, however, (Figure 3.5.5) is very different with large regions of LD. Figure 3.5.6 represents the SNPs predicted to form haplotypes CCL4-SA-HapB1 (A), CCL4-SA-HapB2 (B) and CCL4-SA-HapB3 (C). Some SNPs that were predicted to form part of these haplotypes showed partial linkage to the other SNPs in the haplotype. Haplotype frequencies in the Black population were calculated by including individuals that had all the SNPs of a given haplotype. Haplotypes occurred at frequencies ranging from 4-34% (Figure 3.5.3).

CCL4-SA-HapC1 is a large haplotype consisting of 11 SNP loci, the majority of which are located in the second intronic region of CCL4. Of the total six SNPs that were located in exons, two of these are found in this haplotype, i.e. the two SNPs that lie adjacent to each other at positions 1444 and 1445 (Table 3.5.1; rs1049807 and rs1719152). Interestingly, rs1719152 is one of the three non-synonymous

SNPs identified, leading to the Ser80Thr change (Figure 3.5.2). CCL4-SA-HapC1 occurred at a genotypic frequency of 28.13% and of the 13 individuals with this haplotype, five individuals were homozygous for all SNPs in the haplotype which may suggest that linkage disequilibrium between the SNPs is complete, however upon Haploview analysis, it became apparent that SNP rs1635417 at position 1184, showed incomplete linkage with the other SNPs in the haplotype as this was a very prominent SNP, occurring at an allelic frequency of 31.25% in the Caucasian population, suggesting that this SNP will always occur as part of this haplotype but may occur independently of the haplotype as well.

Haplotypes CCL4-SA-HapC2 and CCL4-SA-HapC3 are both haplotypes comprised of two SNPs and were both very prevalent, observed in 34.38% and 20.31% of the Caucasian cohort respectively (Figure 3.5.3). Like CCL4-SA-HapC1, when individuals with the CCL4-SA-HapC3 haplotype were homozygous for one of the SNPs, they were also homozygous for the other, implying complete linkage disequilibrium between the two SNPs. This, however was not observed for CCL4-SA-HapC2, which was the haplotype that was present in over half of the Caucasian individuals (Figure 3.5.3; 34.38%). Haploview analysis indicated that the SNP rs72833526 (position 1018), which occurred at a frequency of 48.44% in the Caucasian population, was incompletely linked with the SNP rs10491121 (position -884) with an $r^2$ value of 0.55 (Figure 3.5.5) The SNPs making up CCL4-SA-HapC3, rs116657123 (position 663) and rs1719147 (position 906) are also very prevalent in the Black cohort with allele frequencies of 45.65% and 39.69 % respectively (Table 3.5.1). A three-SNP haplotype, CCL4-SA-HapC4, displayed complete linkage between the SNPs rs140335931 (position -914), rs148133700 (position 1725) and rs146185601 (position 2164).

One three-SNP haplotype, CCL4-SA-HapB1 and two seven SNP haplotypes, CCL4-SA-HapB2 and CCL4-SA-HapB3, were identified in the Black population, occurring at genotypic frequencies of 10.87%, 8.70% and 6.52% respectively (Figure 3.5.3). The exon two SNP at position 742 (rs1719146; Table

3.5.1) forms part of CCL4-SA-HapB1 and the SNP at position 1445 (rs1719152; Table 3.5.1), which is a non-synonymous SNP in exon 3 forms part of the CCL4-SA-HapB2 haplotypes, as reported above, this SNP also forms part of the CCL4-SA-HapC1 Caucasian haplotype. While three of the four haplotypes identified in the Caucasian population had frequencies> 20%, the haplotypes in Black individuals were far less frequent, i.e., none of the three haplotypes identified had a frequency>20%, which was expected from the patterns of linkage seen in Figures 3.5.4 and 3.5.5. Some SNP loci making up the haplotypes CCL4-SA-HapB1, CCL4-SA-HapB2 and CCL4-SA-HapB3 overlapped with those that formed part of haplotypes in the Caucasian population (positions: -302, 1023, 1445, 2010; Figure 3.5.3).

A four-SNP haplotype, CCL4-SA-Hap1, was identified in the region downstream from the CCL4 3'UTR region, in both Black and Caucasian individuals (Figure 3.5.3). It was however not a very prevalent haplotype, occurring in only two Black and three Caucasian individuals, that is, with minor allele frequencies <5%.

### *3.5.5.Hardy-Weinberg* Equilibrium

All SNP positions in both the Black and Caucasian cohorts did not deviate from Hardy-Weinberg equilibrium, except for the SNP at position 1018 in the Caucasian population (rs72833526; Table 3.5.1). There is a G/A change at this position and the majority of Caucasian individuals were homozygous for the A allele. This SNP was absent in the Black population group. It is interesting to speculate whether this intronic SNP has been the focus of selective pressure and may thus play a role in splicing.

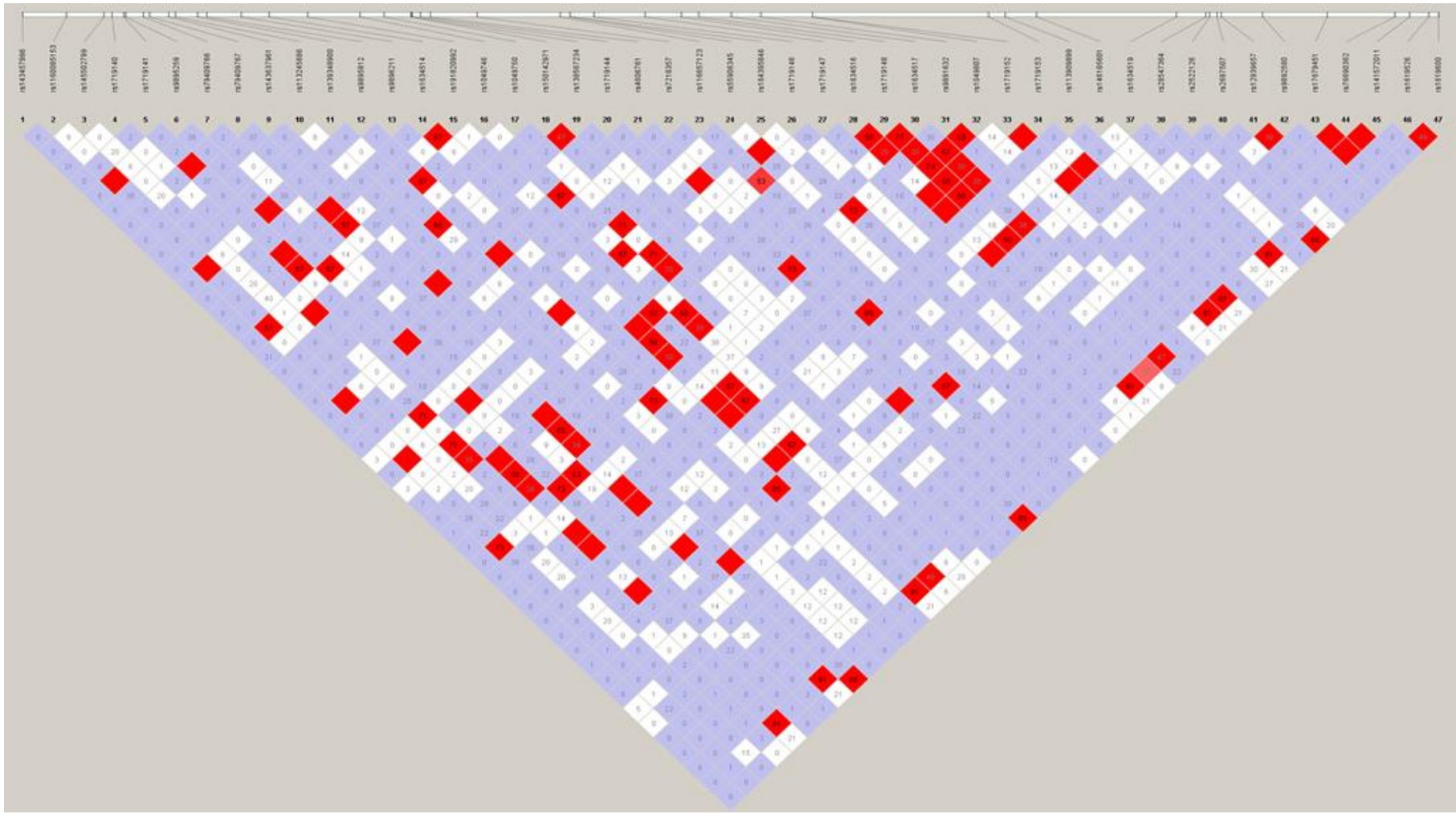


**Figure 3.5.4 Pairwise LD among SNPs (>5% prevalence) across the *CCL4* gene in South African Black individuals**. The significance of association ($r^2$) between SNP pair-wise LD values are given in each diamond. A deep red block represents maximum $r^2$ values and LD of 1. Relative positions of SNPs on the *CCL4* gene are shown.

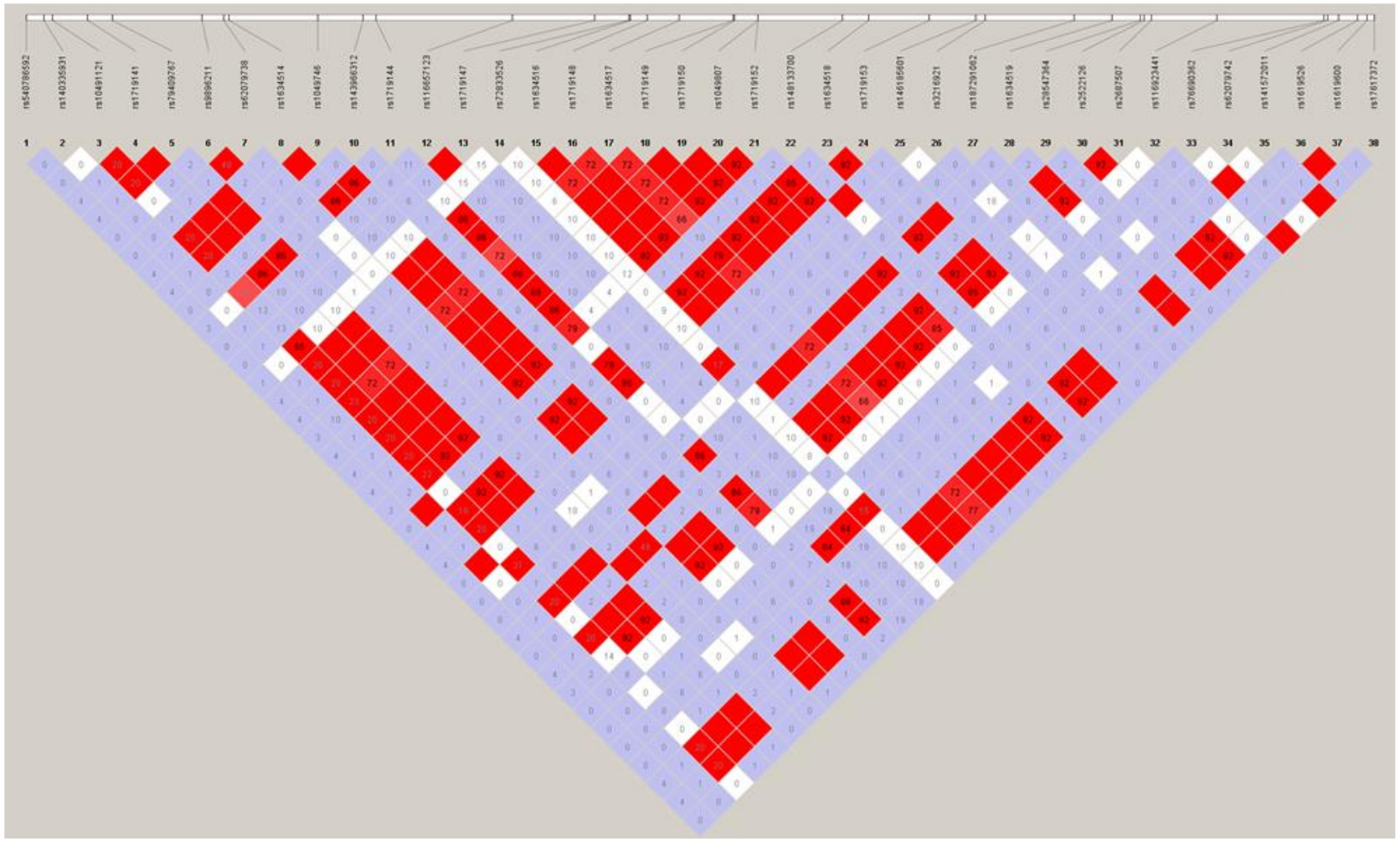



**Figure 3.5.5 Pairwise LD among SNPs (>5% prevalence) across the *CCL4* gene in South African Caucasian individuals**. The significance of association ($r^2$) between SNP pairwise LD values are given in each diamond. A deep red block represents maximum $r^2$ values and LD of 1. Relative positions of SNPs on the *CCL4* gene are shown.

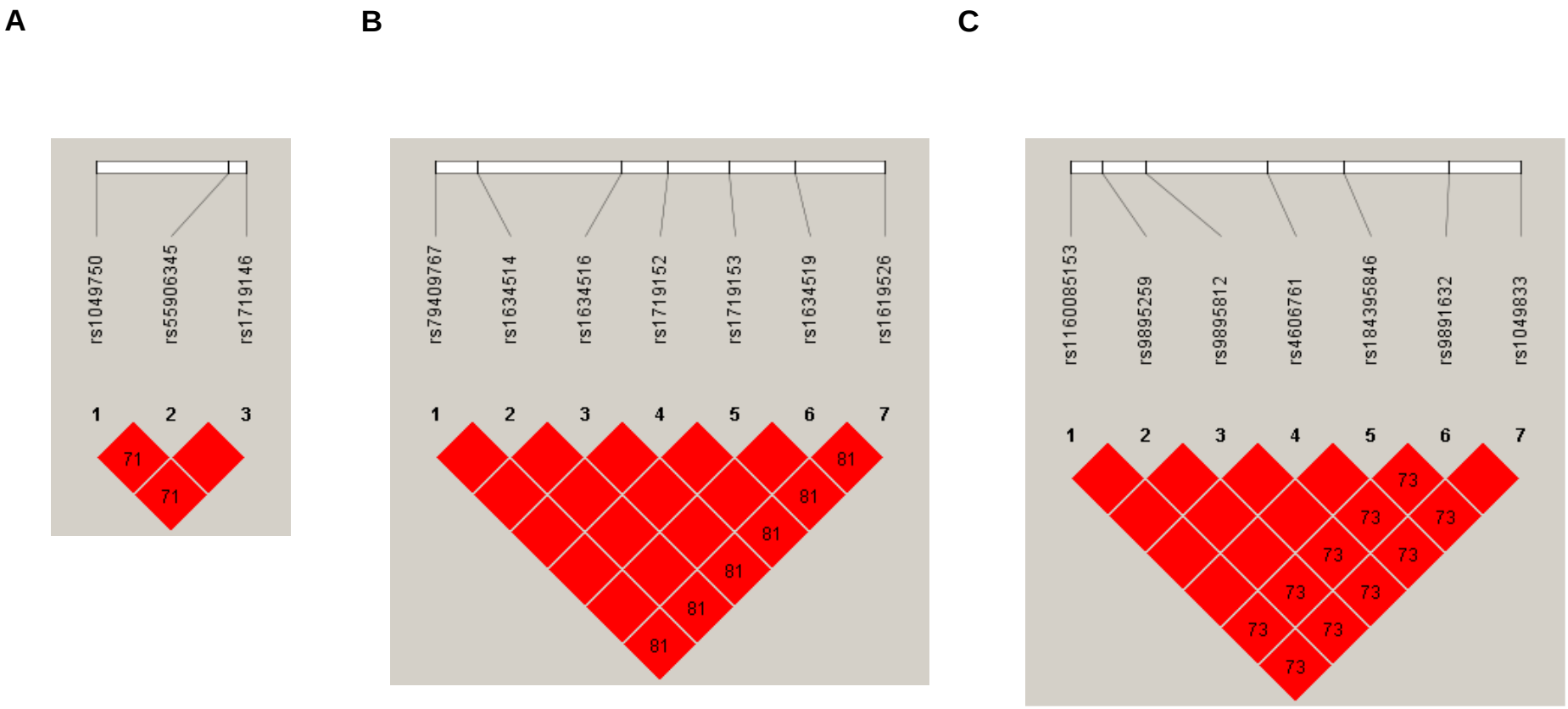



**Figure 3.5.6 Pairwise LD among SNPs (>5% prevalence) across the predicted *CCL4* haplotypes in South African Black individuals**. CCL4-SA-HapB1 (A), CCL4-SA-HapB2 (B) and CCL4-SA-HapB3 (C) are represented. The significance of association ($r^2$) between SNP pair-wise LD values are given in each diamond. A deep red block represents maximum $r^2$ values and LD of 1. Relative positions of SNPs on the *CCL4* gene in the two populations are shown.

### *3.5.6. The influence of Genetic Variation on CCL4 production*

The influence of various prevalent SNP and haplotype combinations to CCL4 protein production was investigated. No association between these genetic factors and CCL4 production were observed. Some comparisons of interesting variants were limited by sample size as the small number of individuals possessing the variant would not allow for valid statistical comparisons.

Three of the four indels that were identified were too infrequent to have any influence on production. The comparison of levels of CCL4 production in Caucasian individuals who had the deletion at position 2197 (rs3216921; Table 3.5.1) and those who did not, revealed that individuals with the deletion had a significantly higher level of PHA-stimulated CCL4 production but not unstimulated production, than those that did not possess the deletion ($p=0.024$; Figure 3.5.7). Due to the discrepancy with the actual bases that have been deleted in this region (section 3.5.3), the Figure 3.5.7 X-axis is simply labelled as 'deletion' and 'no deletion.'

This deletion is found downstream from the 3'UTR. It is interesting to speculate as to how flanking region variants may impact on transcription or if these variants are somehow involved in miRNA binding.

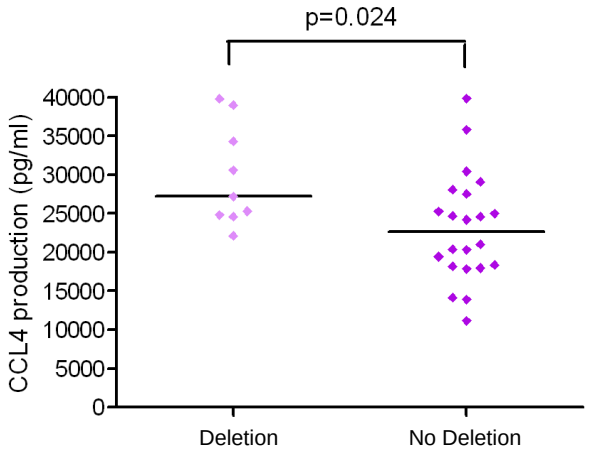



**Figure 3.5.7 Comparison of CCL4 production in Caucasian individuals with and without the rs3216921 deletion in the CCL4 gene**

# CHAPTER 4 DISCUSSION AND CONCLUSIONS

Approximately 12% of the human genome displays CNV and the functional significance of these variations fall within a wide range. Many studies have provided examples in which gene copy number variability can influence disease outcomes (Freeman et al. 2006; Redon et al. 2006; Wain et al. 2009). The downfalls of the methods, such as qPCR to accurately determine CNV, particularly at higher copy numbers, have been partly accountable for the controversy questioning whether associations generated from these methods can be considered true as they are not always reproducible (Shrestha et al. 2009; Cantsilieris and White 2013). These contradictions have had a significant influence in determining the link between copy number of the CC chemokine encoding genes and protection against HIV-1 infection and disease attenuation.

A number of studies have demonstrated the accuracy of ddPCR for a variety of applications including copy number determination of the CCL3 encoding genes (Hindson et al. 2011; Heredia et al. 2013). In this study, the comparison of qPCR and ddPCR to determine gene copy number of *CCL4L*, *CCL4L1* and *CCL4L2* genes in South African Black and Caucasian populations clearly demonstrated the differences in accuracy of these two methods.

Genotyping of the CCL4 encoding genes is ideal for comparing CNV assays because copy number of the *CCL4L1* and *CCL4L2* genes can be determined individually and together as *CCL4L*. The sum of the number of copies of *CCL4L1* and *CCL4L2* should therefore equal to the copy number generated for *CCL4L* (Shostakovich-Koretskaya et al. 2009), and the relationship between these values gives a good indication of the accuracy of an assay. Although, the correlation of these two factors maintained significance when assessed by qPCR (r=0.87) there was a stronger correlation observed when ddPCR (r= 0.99) was performed. Furthermore when one looks at the correlations below and above a *CCL4L* copy number of 3, the qPCR assay shows a very weak correlation (r=0.44) at *CCL4L* copy numbers >3, whereas the ddPCR assay is almost 100% correlated (r=0.99).The number of individuals in whom the

two values did not correspond was thus far greater when qPCR was used compared to the ddPCR assay, where just three incidences were observed. Two of the three unusual patterns observed in ddPCR were explained by SNPs in the probe binding sequence and at the 3' end of a reverse primer, which were detected by sequencing the *CCL4L* genes of the three individuals. It is suspected that the third unusual pattern observed, is perhaps a result of SNPs within the *BGB* gene which would result in an incorrect gene copy number but would still maintain a correct gene ratio. The presence of both the *CCL4L1* and *CCL4L2* genes in the third individual was confirmed by standard amplification and sequencing of these genes. Interestingly, although the copy number patterns of the individuals generated by ddPCR were seen in some of the repeats when raw qPCR data was examined, the inconsistency of the results lead to a final determination of 0 copies of the *CCL4L* genes in these individuals. In contrast, the same patterns were consistently observed between replicates within an experiment and upon repetition of the experiment with ddPCR. The consistency of these results impelled the further investigation by sequencing and the findings highlight the sensitivity and accuracy of the ddPCR assay.

Concurring with previous reports (Gonzalez et al. 2005; Picton et al. 2013), the qPCR assay had reduced accuracy at higher gene copy numbers. This is particularly disadvantageous in a South African context as African populations have previously been reported (Colobran et al. 2008), and corroborated in this study, to have higher *CCL4L* copy number medians and ranges. While the highest copy number of the CCL4 encoding genes did not exceed 6 copies, many genes that display CNV have much higher copy number ranges. The CCL3 encoding genes, for example, have been found to occur in up to 20 copies pdg (Gonzalez et al. 2005). Similar to *CCL4L*, the *CCL3L1* genes were shown to occur in higher copies in African populations. Furthermore this gene has been found to occur in even higher copies in chimpanzees with a mean copy number of 9 (Gonzalez et al. 2005). Application of the ddPCR assay could thus also be greatly beneficial to CNV studies being conducted in non-human primates.

The development of the paralogue ratio test (PRT) for *CCL3L* copy number determination and its subsequent application suggests that the duplicated region on chromosome 17, in which the *CCL3L* and *CCL4L* genes fall, will always contain the same copy number of *CCL3L* and *CCL4L* (Walker et al. 2009; Carpenter et al. 2012; Carpenter et al. 2014). However, applying this method in some African population studies and using FISH to verify results suggested that heterogeneity may exist in this region, unlike in European populations (Aklillu et al. 2013). Cardone et al. (2008) suggest that break points within this duplicated region would allow for the copy number of the *CCL3L* and *CCL4L* genes to vary from one another. Interestingly, when comparing the copy number data generated for *CCL4L* with that generated for *CCL3L* in the same cohort (Picton et al. 2013), a positive strong correlation is seen in both population groups, but as observed by Shostakovich-Koretskaya et al. (2009), the *CCL3L* copy numbers are higher than *CCL4L.* While Walker et al. (2009) suggest that the difference observed in copy number when *CCL3L* and *CCL4L* are compared is a result of the failure of studies to exclude the *CCL3L2* gene (i.e. truncated gene), a significant positive correlation is maintained when excluding *CCL3L2* (data not shown), however copy numbers were still not equal. In order to further elucidate the relationship of *CCL3L* and *CCL4L* in sub-Saharan African populations, determining both *CCL3L* and *CCL4L* copy number using ddPCR in the same cohort will be highly informative and is planned for future studies. Furthermore, as ddPCR becomes more widely implemented over time, it will be interesting to see how the PRT test performs relative to ddPCR using the same dataset or cohort.

Inaccuracy of qPCR at higher copy numbers was evident when population ranges and medians were examined. While both assays produced the same copy number medians for all three genes for the Caucasian population, calculated medians differed between the two methods for the African population. The copy number distribution around the population medians appeared to follow a very similar pattern when results from both assays were compared for the Caucasian population. This was not observed in the Black population. CCL3 and CCL4 are the natural ligands of the CCR5 receptor,

which is a major co-receptor, used by HIV-1 to infect cells, and have been studied extensively for their role in HIV disease (Cocchi et al. 1995). Incorrectly establishing median copy number has implications for association studies that aim to determine the relationship between CNV, protein production and their impact on various communicable and non-communicable diseases. Having higher copy numbers of some of the CCL3 and CCL4 encoding genes, more specifically ,above the population specific median, has been shown in several studies to influence protein production and to have protective effects against HIV-1 transmission (Gonzalez et al. 2005; Meddows-Taylor et al. 2006; Kuhn et al. 2007; Shostakovich-Koretskaya et al. 2009). The protective effects of CCL3 have also been demonstrated in mother to child transmission studies (Meddows-Taylor et al. 2006; Kuhn et al. 2007).

Although the set-up of a ddPCR experiment involves several more steps than qPCR, that is, restriction digestion, droplet generation, a conventional PCR and then the result-generating droplet reading step, and is consequently more time consuming than qPCR, only three of the samples needed to be repeated. In contrast, the results generated using qPCR showed differences between repeats for a single sample both within and between experiments. On average, samples assayed using qPCR had to be repeated three times and two samples remained unresolved. Average copy number values calculated from ddPCR replicates clustered close to whole integers, minimizing ambiguity. Data generated by the qPCR assay did not display the same type of clustering and samples were often repeated when it was unclear if the copy number was the integer above or below the average value calculated.

A minor drawback of ddPCR is that a larger quantity of DNA had to be used compared to qPCR. The amount of DNA required still remained higher than qPCR even when all the qPCR repeats were considered. This may be a disadvantage when conducting studies in which limited DNA is available.

Since the technology is relatively new, as its popularity and implementation grows, we envisage that the limitations of the assay will be improved upon.

In the evaluation of gene copy number, absolute quantification, as in ddPCR, rather than the extrapolation of data from a standard curve, as in qPCR, seems to be far more advantageous. In qPCR, the amplification efficiencies of the endogenous control and the genes of interest have to be the same as data is generated as a ratio of the quantities generated from the two reactions. While for ddPCR, the data generated is independent of the PCR efficiency as it is generated using the number of end-point positive reactions (Hindson et al. 2011). The experimental design is therefore far simpler and requires less optimization.

While our interests lie in determining the role of the CCL3 and CCL4 encoding genes in HIV-1 infection, there are several other disease studies to which this modified *CCL4L* assay can be applied. CCL4 has been implicated in susceptibility to tuberculosis and leprosy (Jamieson et al. 2004), severity of psoriasis (Pedrosa et al. 2011) and dengue fever (Bozza et al. 2008). The advantages of using ddPCR over qPCR when assaying high CNV have already been discussed in this study. Other studies have also benefited from the increased sensitivity of ddPCR and other forms of digital PCR over qPCR, for example, the detection of rare targets such as *HER2* tumour gene copies from cell free DNA of breast cancer patients which is diluted by normal DNA (Whale et al. 2012). Droplet digital PCR was also shown to be more effective than using methods such as FISH and immunohistochemistry for *HER2* detection (Heredia et al. 2013). The technology has also been used to detect and quantify HIV-1 proviral DNA in HIV-1-infected patients (Strain et al. 2013).

This study also aimed (i) to investigate the influence of *CCL4L* CNV, generated by ddPCR, on CCL4 production, in a cohort of HIV-1-uninfected individuals, (ii) to compare these findings to the closely

associated CCL3 chemokine data generated by Picton et al. (2013 and unpublished data), and (iii) to carry out a preliminary investigation of the role of CCL4 in HIV-1 control by looking at a small subset of LTNPs with respect to CCL4 copy number and production compared to HIV-uninfected individuals from the same population demographic.

As previously described above and further expanded upon now, the Black South African population displays a wider range of copy number variability, higher median copy numbers as well as a greater diversity of copy number combinations of the genes *CCL4L1* and *CCL4L2,* than the Caucasian population. Although Black females showed significantly higher *CCL4L* copy numbers than Black males, the addition of copy number data from a cohort of 30 females to the cohort may have skewed the results due to the large difference in the number of Black females compared to Black male participants, however future studies involving larger matched groups would need to be carried out to determine if this is a true association and what this could mean about the evolution of these genes.

The negative correlation between *CCL4L1* and *CCL4L2* gene copy number and the reported (Shostakovich-Koretskaya et al. 2009) opposing effects of these genes in HIV-1 infection is an intriguing phenomenon, especially since other similar gene correlations in this study were positive, including when the genes *CCL3La* and *CCL3Lb* were compared, and these genes too give rise to a functional CCL3 protein and a truncated protein, respectively (Shostakovich-Koretskaya et al. 2009). Since *CCL4L1* encodes a protein with CCL4 functionality, and the splice variants of *CCL4L2* do not seem to give rise to proteins that have 'classical' CCL4 functionality (Colobran et al. 2005), it is interesting to speculate whether the 'non-functional' *CCL4L2* gene has a post-transcriptional positive regulatory effect on CCL4 production (i.e. increases CCL4 production), and that levels of CCL4 production have been evolutionarily controlled/maintained by the inverse copy number relationship of these two genes. Regarding the relationship between *CCL3*-encoding genes and *CCL4*-encoding genes, Shao et al. (2007)

reported a positive correlation between the *CCL3L* and *CCL4L* genes, however, both Shao et al. (2007) and Shostakovich-Korestskaya et al. (2009) determined that the these gene copy numbers were not identical. Comparing *CCL4L* and *CCL3L* (Picton et al. 2013) copy number data in the Black and Caucasian cohorts in this study also a showed significantly positive correlation, except however when *CCL3Lb* and *CCL4L2* were compared in the Black population. Furthermore, the copy number of *CCL3La* was significantly higher than *CCL4L1* copy number in the Black but not in the Caucasian population. These population variations in copy number distribution were overlooked during the design of the paralogue ratio test for *CCL3L* and *CCL4L* copy number determination (Walker et al. 2009).

When assessing CCL4 protein production by ELISA, it is assumed, as has been indicated for CCL3, that the assay cannot distinguish between the contribution of protein from the *CCL4* gene, which occurs at two copies pdg, and *CCL4L1,* since the protein sequences of these genes differ by only a single amino acid. None of the *CCL4L* genes, individually or combined, correlated with protein production in either Black or Caucasian individuals. Making the broad assumption that the contribution from *CCL4* is constant, the correlation between gene copy number of the *CCL4L* encoding genes and CCL4 production was examined by dividing CCL4 production by gene copy number (i.e. comparing CCL4 production per CCL4-encoding gene copy). Although Caucasians had significantly higher CCL4 production per *CCL4L* copy number compared to Black individuals, this was not maintained when looking at *CCL4L1* (or *CCL4L2*) alone or when we corrected for the effect of *CCL4L* copy number ((i.e. compared individuals from two populations with similar *CCL4L* copy number (and distributions of *CCL4L1* and *CCL4L2*)). Several studies have suggested that the protective effects of *CCL3L* and/or *CCL4L* on HIV-1 did not show with the exact copy numbers per se, but rather when looking at copy numbers relative to the population mean. We therefore stratified CCL4 production based on the population medians generated for the Black and Caucasian populations. Similar to the pattern (albeit not significant) that was observed between CCL3 production and *CCL3La* copy number (Picton et al, 2013),

we observed a significantly lower level of CCL4 production by Black individuals with *CCL4L* copy number below the population median compared to production by individuals with median *CCL4L* copy number. This association however was not maintained when *CCL4L1* and *CCL4L2* were analysed individually in the same manner. Furthermore the inverse relationship was observed (i.e. median production by median copy number was lower than production of lower copy number individuals, although lacking significance). The negative correlation seen between *CCL4L1* and *CCL4L2* may be a possible explanation for the latter observation. The significant lower production seen at *CCL4L* copy number below the population median may indeed be reflective of the lower copy number, i.e. copies below the median all translate into protein. At the median, it seems that production is optimal, however copy number above the median does not seem to translate into an expected higher production, suggesting that at the higher copy number, post-transcriptional regulation may be "kicking" in to keep the production at an optimum threshold. It has to be kept in mind that while the 'pseudogene' of *CCL3L* produces a truncated protein, the case of CCL4 may be more complex because the most abundant *CCL4L1* transcript has been shown to produce a protein that is missing the first 5 amino acids of exon 3 which may prevent it from forming heterodimers and self-aggregates (Colobran et al. 2005). Since it is not known which epitopes the antibody in the ELISA system binds to, it cannot be totally ruled out that protein products from these spliced variants are being detected.

What is most notable when comparing the scenarios of *CCL3L* and *CCL4L,* is that in the case of *CCL3L*, associations mostly relate to *CCL3La* gene copy number (which includes the genes *CCL3L1* and *CCL3L3*, both producing a functional CCL3 protein) but in the case of CCL4, they are formed around *CCL4L* gene copy number (the combination of *CCL4L1* and *CCL4L2*). Comparing CCL4 copy number and production data to CCL3 data generated by Picton et al. (2013) in the same two cohorts, revealed a positive significant correlation of CCL3 and CCL4 protein levels in the Caucasian but not the Black population. However, CCL4 protein levels were higher than that of CCL3 in both populations.

Picton et al.'s (2013) study of the distribution of the *CCL3L* genes as well as their association with CCL3 protein production did not find marked differences in both stimulated and unstimulated production of CCL3 and neither did this study on CCL4, respectively, between the two population groups. An association between *CCL4L* and CCL4 production was not observed in a cohort of Black LTNPs in this study either, similarly, nor was there a correlation between CCL3 and *CCL3L* production in the same cohort of Black LTNPs (Picton et al., unpublished). We have however observed that LTNP individuals have lower levels of both CCL3 and CCL4 protein production compared to uninfected individuals. Examining chemokine production within the LTNP cohort of Black individuals alone, it was discovered that those individuals with viral loads less than 400 RNA copies/ml had lower levels of CCL4 in both unstimulated and PHA-stimulated PBMC supernatants than those with viral loads greater than 400 RNA copies/ml. The same was seen for PHA-stimulated CCL3 production in the same cohort (Picton et al, unpublished). Furthermore the LTNPs with viral loads less than 400 RNA copies/ml produced significantly less unstimulated and PHA-stimulated CCL4 compared to the uninfected individuals. Although these findings may be influenced by the small size of the LTNP cohort, these interesting results warrant further investigation, using larger cohorts of HIV-1 controllers and investigating whether these associations are uniquely seen in CCL3 and CCL4 or whether other chemokine production is similarly dampened and what this means in terms of the ability of these unique individuals to control HIV-1 infection. Given that these chemokines are involved in immune feedback mechanisms, lower level of production in LTNPs may lead to a less active immune response to HIV-1 infection. It would also be interesting to compare whether or not HIV-1 progressors display an opposite phenotype, that is, do they produce higher levels of CCL3 and CCL4 compared to uninfected individuals and LTNPs.

The examination of a cohort of Caucasian LTNP individuals showed that they had a higher frequency of the *CCR5Δ32* heterozygotes individuals compared to progressors and rapid progressors, (Stewart et al. 1997). This point reiterates the need to conduct population specific studies on these LTNP individuals because the majority of the LTNP individuals in our study are Black and this mutation is not prevalent in the Black population suggesting that there are perhaps a number of independent factors that can influence disease progression. While some studies are uncertain if chemokines contribute to slow progression (Gea-Banacloche et al. 2000), others suggest that perhaps studying chemokine production in areas such as the lymph nodes may reveal different results from studies that focus on the level of circulating chemokines in the blood (Pantaleo et al. 1995).

Since the contribution of the two copy pdg gene *CCL4,* to CCL4 production has not been accounted for when looking at copy number, we carried out a preliminary characterisation of the genetic variants found in this gene in the two population groups and analysed these variants with respect to their influence on CCL4 production. We identified a total of 22 SNPs within the CCL4 gene in the Black South African cohort and 12 SNPs in Caucasian individuals. Several other SNPs upstream and downstream from the gene were also identified. Just a single SNP, occurring in one Black individual, was newly identified in this study. In the same population, genetic characterization of *CCL3* yielded 25 intragenic SNPs being identified in Black individuals and 9 SNPs in Caucasian individuals (Paximadis et al. 2009). Five CCL3 SNPs were newly identified. We have described four novel haplotypes in the South African Caucasian population, three in the Black population, and a haplotype that is shared between both population groups. To our knowledge, the only other study to have described LD and haplotype structure incorporating the *CCL4* gene was done by Modi et al. (2006), only four SNPs that were incorporated in their study overlapped with the SNPs identified in our study (rs1634514, rs1719144, rs1719146, rs1719153). The SNPs rs1634514 and rs17179153 (D' =98, Modi et al. 2006) were 100% linked in both our population groups. Our study is first to have comprehensively reported the LD

pattern across this gene for both Black and Caucasian South African individuals. Again, compared to *CCL3,* three haplotypes were identified in the Black cohort while only one haplotype was characterised in the Caucasian population group (Paximadis et al. 2009). Generally, Caucasian populations tend to exhibit greater linkage disequilibrium owing to less genetic diversity and 'less time' for genetic recombination events to occur (Reich et al. 2001).

The majority of *CCL4* SNPs that were identified in both populations were positioned from the second intron onwards while the area before the second intron is rich in SNPs that are unique to the Black cohort. This can also be seen when examining the haplotypes that were predicted for each population group. The only shared haplotype, CCL4-SA-Hap1, harboured four SNPS downstream from the gene. Out of the four Caucasian haplotypes identified, in the one that was most prevalent, CCL4-SA-HapC1, the majority of SNPs fell within the second intron. The only SNP that deviated from Hardy-Weinberg equilibrium, rs72833526 (position 1018), which seems to have been selected for in Caucasians, but is not present in Black individuals in this cohort, is also situated in the second intron. This is interesting given that the genes *CCL4L1* and *CCL4L2* differ by a single SNP in the acceptor splice site in intron 2, resulting in aberrant splicing of many products that do not seem to have any chemokine activity. The ancestral chimpanzee sequence is often used as a reference in assigning human WT alleles when they are unclear. Chimpanzees however, do not have the *CCL4L2* and *CCL3L2* genes, which are both considered the 'non-functional,' non-allelic isoforms of CCL4 and CCL3 respectively. Furthermore, we found that these genes also occur in multiple copy numbers and in the case of CCL4, the *CCL4L2* genes occur at a higher copy number than *CCL4L1.* It has been suggested that a much higher genetic diversity is present in the human CCR5 gene compared to the CCR5 gene in chimpanzees and that in some cases, the predominance of certain SNPs in humans meant that they had lost certain conserved mRNAs found in other apes but had gained the ability to form novel mRNA sequences (Mummidi et

al. 2000). It is a possibility that the same response to evolutionary pressure, perhaps driven by select immunological events, which is seen in *CCR5,* is also impacting the genes encoding its ligands.

Three out of six exonic SNPs identified were non-synonymous, however, the amino acid changes that resulted were predicted to be 'benign' in terms of the way that they may impact protein function. It has been suggested that certain synonymous codons may be selected for, due to their association with exonic splicing regulatory elements (Parmley and Hurst 2007). Both the synonymous SNP rs1049807 at position 1444 and its neighbour at position 1445, the non-synonymous rs1719152, form part of the most common Caucasian haplotype that we identified (CCL4-SA-HapC1).

Two of the SNPs found in this study rs1719146 (C742T) and rs1719153 (A2010T) were found to be associated with an increased progression to AIDS in Black and Caucasian individuals respectively (Modi et al. 2006). This association was observed in individuals who were both homozygous or heterozygous for the polymorphisms. While the minor allelic frequency for rs1719146, a SNP in exon 2 of *CCL4,* in the Black populations examined in the Modi et al. (2006) study was 3%, in our South African Black cohort, it was present at a frequency of 13%. The SNP rs1719153 is found in the region just downstream from the 3'UTR and the T allele occurs at frequencies of 11% and 28% in Black and Caucasian populations investigated in this study, respectively. In the cohorts examined by Modi et al. (2006), the frequency of the allele was 0.7% and 23% for Black and Caucasian individuals respectively. Interestingly, rs1717146 forms part of the CCL4-SA-HapB1 haplotype described in the Black population and rs1719153 forms part of both the CCL4-SA-HapB2 and CCL4-SA-HapC1 haplotypes identified in the Black and Caucasian populations respectively.

Since the CCL4 gene is so polymorphic, it was very difficult to associate differences in protein production with any single genetic variant or haplotype. In the Caucasian population many individuals

had more than one of the characterised haplotypes and therefore teasing apart which individuals to compare with each other with respect to protein produced was difficult. Of the four deletions identified, rs3216921, a three base pair deletion downstream from the 3'UTR, was the most frequent deletion and the only one identified in Caucasian individuals. It was not found in the Black cohort. Individuals with this deletion had a significantly higher level of CCL4 protein production than those who did not harbour the deletion. The finding that LTNP individuals have a lower level of CCL4 production than HIV-1-uninfected individuals brings to mind the question of whether having an increased level of CCL4 production is desirable in the context of HIV-1. Perhaps an elevated level of the CC-chemokine is protective against HIV-infection but, post infection, maybe a decreased level of CCL4 production will lead to slower disease progression. It has been shown that it is only after a short period that viral levels in normally progressing individuals begin to differ from those of LTNPs, in other words, LTNPs are just better at controlling infection from that point forward (Vesanen et al. 1996). It would be interesting to investigate whether the rs3219621 indel, associated with higher CCL4 production in Caucasians, plays any role in HIV-1 susceptibility and/or control in HIV-1 infected Caucasian individuals.

Single nucleotide polymorphisms in the *CCL3L* encoding genes were suggested to be useful indicators around which to build the copy number determination assays, for example if SNPs can distinguish between the non-allelic isoforms (Perry et al. 2014). From our experience of SNPs interfering in probe and primer binding sites, we feel that this is a valid point. Secondly, it was our initial intention to separately amplify the *CCL4L1* and *CCL4L2* genes. The challenge in primer design was not only in trying to find regions that were different enough to place the primers but, because the gene is not as well studied, the reference sequences for *CCL4L2* were amended by the addition of a 'fix patch,' which corrected a difference reported between *CCL4L1* and *CCL4L2* in the region in which our primers were designed. Furthermore, other sites that did not share sequence similarity between the genes were

listed in dbSNP as polymorphic sites as some sequencing data that was submitted was probably generated by sequencing these genes together rather than separately. A similar problem was encountered by Paximadis et al. (2009) when trying to distinguish haplotypes in the CCL3L encoding genes.

One strategy that we are currently working on to get an idea of the genetic diversity of the *CCL4L* genes individually, is to use the copy number data that we generated to identify individuals with zero copies of either *CCL4L1* or *CCL4L2* and thereby sequence and study the *CCL4L* gene that is present. It is predicted that if these genes are quite polymorphic, haplotype prediction would be difficult due to the multiple copy numbers, however, from our preliminary (5 *CCL4L1* and 22 *CCL4L2* sequences) analysis it appears that there is significantly less variation in these genes compared to CCL4 (data not shown),as well as compared to the *CCL3L* genes. The low level of variations in these *CCL4L* genes may suggest differential selection for these genes.

A recent study has demonstrated that a microRNA (miR-125b) downregulates CCL4 expression and plays a part in the increased levels of CCL4 seen in older individuals (Cheng et al. 2015). The role of miR-125b in regulation of CCL4 expression was confirmed by deleting the predicted microRNA binding site in the 3'UTR of CCL4 (Cheng et al. 2015). While we did not find differences in CCL4 levels when individuals of both populations were stratified by age (perhaps relooking at this factor in a larger cohort would be beneficial), we did identify three 3'UTR SNPs in the CCL4 gene (two in Black individuals and one in Caucasian individuals). While none of these SNPs are predicted to cause a disruption in the miR-125b binding site (www.microRNA.org (Betel et al. 2008)), they are predicted to disrupt other microRNA sites (data not shown). This warrants further investigation into expression of miR-125b and the genetic characteristics of CCL4 3'UTRs in Black and Caucasian South African LTNP individuals, who have lower levels of CCL4 protein than their HIV-1-uninfected counterparts. Cheng et

al. (2015) suggest that the observed association between this microRNA and increased CCL4 expression with age may be as a result of interference with miR-125b maturation. This finding prompts the question of whether LTNP individuals are somehow able to maintain a lowered expression of CCL4 and if this may influence their progression in HIV-1-infection.

To conclude, in this study we have shown that droplet digital PCR proved to be a far superior method for the copy number assessment of the genes *CCL4L*, *CCL4L1* and *CCL4L2* compared to qPCR. While results generated by both assays were comparable for lower copy numbers, as observed in the Caucasian population, at higher copy numbers, as observed in the Black population, the qPCR assay fared poorly compared to ddPCR. The incorporation of ddPCR as the standard method of assessing gene copy number in the future would potentially generate unambiguous, accurate copy number data that would contribute to the understanding of copy number variation and its contribution to human infection and disease. We have also shown that although the CCL3- and CCL4-encoding genes are found on the same segmental duplication and bind to the same receptor, their copy numbers vary from each other and the extent of this variation is dependent on the population being studied. In addition we have highlighted differences in production of these two chemokines. We have shown that *CCL4L* gene copy number in the two South African populations does not directly influence chemokine production although, in the Black population, it was demonstrated that optimum CCL4 production occurs when an individual possessed a *CCL4L* copy number equal to the population specific median, suggesting that instead of a gene dosage mechanism, that there may be much more complex mechanisms of regulation involved. We further demonstrate the population differences exist between the South African Black and Caucasian cohorts with regards to the CCL4-encoding genes. Variation in key immune-modulating factors between populations needs to be kept in mind during the design of HIV-1 treatments or vaccines, as they may influence responses to these compounds, necessitating the conduction of drug and vaccine trials in diverse populations. By including a small

subset of LTNPs, individuals who inherently control HIV-1 infection for many years in the absence of antiretroviral therapy, we have had a glimpse into how studying these individuals can help to identify possible correlates of immunity against HIV-1. A greater understanding of host responses to HIV-1 will offer more criteria for the inclusion and exclusion of some strategies and help direct the search for treatments and cures down a better defined path. Investigations into the genetic variability of the CCL4 gene revealed the region to be very polymorphic with very clear population specific differences. We observed that variants along the entire *CCL4* gene and its flanking region were organised in distinct haplotypes in both Caucasian and Black individuals with Caucasians exhibiting much stronger LD across the gene compared to Black individuals. We found an association between a three bp deletion in the Caucasian population and increased production of CCL4. These findings highlight the need to carry out detailed genetic studies, as many novel genetic influences on protein production are still undiscovered as well as the need to keep population genetics in mind when trying to uncover these associations. This work has provided an important baseline and has influenced new directions of thought from which we can proceed to further investigate the impact of population specific variability on host immune factors and how these influence the outcomes of HIV-infection.

# REFERENCES


Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56-65.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nature methods* **7**(4): 248-249.

Agosto LM, Zhong P, Munro J, Mothes W. 2014. Highly Active Antiretroviral Therapies Are Effective against HIV-1 Cell-to-Cell Transmission. *PLoS Pathog* **10**(2): e1003982.

Ahmed RK, Biberfeld G, Thorstensson R. 2005. Innate immunity in experimental SIV infection and vaccination. *Mol Immunol* **42**(2): 251-258.

Ahmed RK, Nilsson C, Biberfeld G, Thorstensson R. 2001. Role of CD8+ cell-produced anti-viral factors in protective immunity in HIV-2-exposed but seronegative macaques resistant to intrarectal SIVsm challenge. *Scand J Immunol* **53**(3): 245-253.

Ahmed RK, Nilsson C, Wang Y, Lehner T, Biberfeld G, Thorstensson R. 1999. Beta-chemokine production in macaques vaccinated with live attenuated virus correlates with protection against simian immunodeficiency virus (SIVsm) challenge. *The Journal of general virology* **80 ( Pt 7)**: 1569-1574.

AIDSinfo. 2014. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. In *Panel on Antiretroviral Guidelines for Adults and Adolescents*. Department of Health and Hman Services.

Aklillu E, Odenthal-Hesse L, Bowdrey J, Habtewold A, Ngaimisi E, Yimer G, Amogne W, Mugusi S, Minzi O, Makonnen E et al. 2013. CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *BMC Infect Dis* **13**: 536.

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C et al. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599): 868-871.

Barrenäs F, Andersson B, Cardell LO, Langston M, Mobini R, Perkins A, Soini J, Ståhl A, Benson M. 2008. Gender differences in inflammatory proteins and pathways in seasonal allergic rhinitis. *Cytokine* **42**(3): 325-329.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**(2): 263-265.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic acids research* **34**(Database issue): D16-20.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic acids research* **36**(Database issue): D149-153.

Bhattacharya T, Stanton J, Kim EY, Kunstman KJ, Phair JP, Jacobson LP, Wolinsky SM. 2009. CCL3L1 and HIV/AIDS susceptibility. *Nat Med* **15**(10): 1112-1115.

Blanpain C, Migeotte I, Lee B, Vakili J, Doranz BJ, Govaerts C, Vassart G, Doms RW, Parmentier M. 1999. *CCR5 Binds Multiple CC-Chemokines: MCP-3 Acts as a Natural Antagonist*.

Bozza FA, Cruz OG, Zagne SM, Azeredo EL, Nogueira RM, Assis EF, Bozza PT, Kubelka CF. 2008. Multiplex cytokine profile from dengue patients: MIP-1beta and IFN-gamma as predictive factors for severity. *BMC Infect Dis* **8**: 86.

Cantsilieris S, White SJ. 2013. Correlating multiallelic copy number polymorphisms with disease susceptibility. *Hum Mutat* **34**(1): 1-13.

Cardone MF, Jiang Z, D'Addabbo P, Archidiacono N, Rocchi M, Eichler EE, Ventura M. 2008. Hominoid chromosomal rearrangements on 17q map to complex regions of segmental duplication. *Genome Biol* **9**(2): R28.

Carpenter D, Farnert A, Rooth I, Armour JA, Shaw MA. 2012. CCL3L1 copy number and susceptibility to malaria. *Infect Genet Evol* **12**(5): 1147-1154.

Carpenter D, Taype C, Goulding J, Levin M, Eley B, Anderson S, Shaw MA, Armour JA. 2014. CCL3L1 copy number, CCR5 genotype and susceptibility to tuberculosis. *BMC Med Genet* **15**(1): 5.

Catano G, Chykarenko ZA, Mangano A, Anaya JM, He W, Smith A, Bologna R, Sen L, Clark RA, Lloyd A et al. 2011. Concordance of CCR5 genotypes that influence cell-mediated immunity and HIV-1 disease progression rates. *The Journal of infectious diseases* **203**(2): 263-272.

Cheng NL, Chen X, Kim J, Shi AH, Nguyen C, Wersto R, Weng NP. 2015. MicroRNA-125b modulates inflammatory chemokine CCL4 expression in immune cells and its reduction causes CCL4 increase with age. *Aging cell* **14**(2): 200-208.

Cheruvu VK, Igo RP, Jr., Jurevic RJ, Serre D, Zimmerman PA, Rodriguez B, Mehlotra RK. 2014. African ancestry influences CCR5 -2459G>A genotype-associated virologic success of highly active antiretroviral therapy. *Journal of acquired immune deficiency syndromes* **66**(1): 102-107.

Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. 1995. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **270**(5243): 1811-1815.

Coetzee D, Hildebrand K, Boulle A, Maartens G, Louis F, Labatala V, Reuter H, Ntwana N, Goemaere E. 2004. Outcomes after two years of providing antiretroviral treatment in Khayelitsha, South Africa. *AIDS* **18**(6): 887-895.

Coffin J, Haase A, Levy JA, Montagnier L, Oroszlan S, Teich N, Temin H, Toyoshima K, Varmus H, Vogt P et al. 1986. What to call the AIDS virus? *Nature* **321**(6065): 10.

Colobran R, Adreani P, Ashhab Y, Llano A, Este JA, Dominguez O, Pujol-Borrell R, Juan M. 2005. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. *Journal of immunology* **174**(9): 5655-5664.

Colobran R, Comas D, Faner R, Pedrosa E, Anglada R, Pujol-Borrell R, Bertranpetit J, Juan M. 2008. Population structure in copy number variation and SNPs in the CCL4L chemokine gene. *Genes and immunity* **9**(4): 279-288.

Colobran R, Pedrosa E, Carretero-Iglesia L, Juan M. 2010. Copy number variation in chemokine superfamily: the complex scene of CCL3L-CCL4L genes in health and disease. *Clin Exp Immunol* **162**(1): 41-52.

Combadiere C, Ahuja SK, Tiffany HL, Murphy PM. 1996. Cloning and functional expression of CC CKR5, a human monocyte CC chemokine receptor selective for MIP-1(alpha), MIP-1(beta), and RANTES. *J Leukoc Biol* **60**(1): 147-152.

Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. 1997. Change in Coreceptor Use Correlates with Disease Progression in HIV-1–Infected Individuals. *The Journal of Experimental Medicine* **185**(4): 621-628.

Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**(5996): 763-767.

Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E et al. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**(5283): 1856-1862.

Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM et al. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**(6584): 661-666.

Donald G, McNeil J. 2009. For First Time, AIDS Vaccine Shows Some Success. In *New York Times*, New Youk.

Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP et al. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**(6584): 667-673.

Eugen-Olsen J, Iversen AK, Garred P, Koppelhus U, Pedersen C, Benfield TL, Sorensen AM, Katzenstein T, Dickmeiss E, Gerstoft J et al. 1997. Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *Aids* **11**(3): 305-310.

Field SF, Howson JM, Maier LM, Walker S, Walker NM, Smyth DJ, Armour JA, Clayton DG, Todd JA. 2009. Experimental aspects of copy number variant assays at CCL3L1. *Nat Med* **15**(10): 1115-1117.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME et al. 2006. Copy number variation: new insights in genome diversity. *Genome research* **16**(8): 949-961.

Gallo R, Salahuddin S, Popovic M, Shearer G, Kaplan M, Haynes B, Palker T, Redfield R, Oleske J, Safai B et al. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**(4648): 500-503.

Gea-Banacloche JC, Migueles SA, Martino L, Shupert WL, McNeil AC, Sabbaghian MS, Ehler L, Prussin C, Stevens R, Lambert L et al. 2000. Maintenance of large numbers of virus-specific CD8+ T cells in HIV-infected progressors and long-term nonprogressors. *Journal of immunology* **165**(2): 1082-1092.

Gonzalez E, Bamshad M, Sato N, Mummidi S, Dhanda R, Catano G, Cabrera S, McBride M, Cao XH, Merrill G et al. 1999. Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. *Proceedings of the National Academy of Sciences of the United States of America* **96**(21): 12004-12009.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**(5714): 1434-1440.

Gottlieb. 1981. Pneumocystis Pneumonia- Los Angeles. In *Epidemiologic Notes and Reports*. Centre for Disease Control and Prevention (CDC), Atlanta.

Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, Saxon A. 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**(24): 1425-1431.

Gray G, Buchbinder S, Duerr A. 2010. Overview of STEP and Phambili trial results: two phase IIb test-of-concept studies investigating the efficacy of MRK adenovirus type 5 gag/pol/nef subtype B HIV vaccine. *Curr Opin HIV AIDS* **5**(5): 357-361.

He W, Kulkarni H, Castiblanco J, Shimizu C, Aluyen U, Maldonado R, Carrillo A, Griffin M, Lipsitt A, Beachy L et al. 2009. Reply to: "Experimental aspects of copy number variant assays at CCL3L1". *Nat Med* **15**(10): 1117-1120.

Henrich TJ, Gallien S, Li JZ, Pereyra F, Kuritzkes DR. 2012. Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR. *J Virol Methods* **186**(1-2): 68-72.

Heredia NJ, Belgrader P, Wang S, Koehler R, Regan J, Cosman AM, Saxonov S, Hindson B, Tanner SC, Brown AS et al. 2013. Droplet Digital PCR quantitation of HER2 expression in FFPE breast cancer samples. *Methods* **59**(1): S20-23.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**(22): 8604-8610.

Hirashima M, Ono T, Nakao M, Nishi H, Kimura A, Nomiyama H, Hamada F, Yoshida MC, Shimada K. 1992. Nucleotide sequence of the third cytokine LD78 gene and mapping of all three LD78 gene loci to human chromosome 17. *DNA Seq* **3**(4): 203-212.

Howard OM, Turpin JA, Goldman R, Modi WS. 2004. Functional redundancy of the human CCL4 and CCL4L1 chemokine genes. *Biochem Biophys Res Commun* **320**(3): 927-931.

Irving SG, Zipfel PF, Balke J, McBride OW, Morton CC, Burd PR, Siebenlist U, Kelly K. 1990. Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. *Nucleic acids research* **18**(11): 3261-3270.

Jamieson SE, Miller EN, Black GF, Peacock CS, Cordell HJ, Howson JM, Shaw MA, Burgner D, Xu W, Lins-Lainson Z et al. 2004. Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. *Genes and immunity* **5**(1): 46-57.

Kuhn L, Schramm DB, Donninger S, Meddows-Taylor S, Coovadia AH, Sherman GG, Gray GE, Tiemessen CT. 2007. African infants' CCL3 gene copies influence perinatal HIV transmission in the absence of maternal nevirapine. *AIDS* **21**(13): 1753-1761.

Lajoie J, Poudrier J, Massinga Loembe M, Guedou F, Leblond F, Labbe AC, Alary M, Roger M. 2010. Chemokine expression patterns in the systemic and genital tract compartments are associated with HIV-1 infection in women from Benin. *Journal of clinical immunology* **30**(1): 90-98.

Lassauniere R, Shalekoff S, Tiemessen CT. 2013. A novel FCGR3A intragenic haplotype is associated with increased FcgammaRIIIa/CD16a cell surface density and population differences. *Human immunology* **74**(5): 627-634.

Lefrere JJ, Morand-Joubert L, Mariotti M, Bludau H, Burghoffer B, Petit JC, Roudot-Thoraval F. 1997. Even individuals considered as long-term nonprogressors show biological signs of progression after 10 years of human immunodeficiency virus infection. *Blood* **90**(3): 1133-1140.

Lehner T, Wang Y, Cranage M, Bergmeier LA, Mitchell E, Tao L, Hall G, Dennis M, Cook N, Brookes R et al. 1996. Protective mucosal immunity elicited by targeted iliac lymph node immunization with a subunit SIV envelope and core vaccine in macaques. *Nat Med* **2**(7): 767-775.

Levy JA. 2006. HIV pathogenesis: knowledge gained after two decades of research. *Advances in dental research* **19**(1): 10-16.

Liu C, Carrington M, Kaslow RA, Gao X, Rinaldo CR, Jacobson LP, Margolick JB, Phair J, O'Brien SJ, Detels R. 2004. Lack of associations between HLA class II alleles and resistance to HIV-1 infection among white, non-Hispanic homosexual men. *Journal of acquired immune deficiency syndromes* **37**(2): 1313-1317.

Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**(3): 333-348.

Mann JM. 1987. The World Health Organization's global strategy for the prevention and control of AIDS. *The Western journal of medicine* **147**(6): 732-734.

Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, Lee B, Doms RW, Margolick J, Buchbinder S et al. 1998. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* **282**(5395): 1907-1911.

McDougal JS, Kennedy MS, Sligh JM, Cort SP, Mawle A, Nicholson JK. 1986. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**(4736): 382-385.

Meddows-Taylor S, Donninger SL, Paximadis M, Schramm DB, Anthony FS, Gray GE, Kuhn L, Tiemessen CT. 2006. Reduced ability of newborns to produce CCL3 is associated with increased susceptibility to perinatal human immunodeficiency virus 1 transmission. *The Journal of general virology* **87**(Pt 7): 2055-2065.

Menten P, Wuyts A, Van Damme J. 2002. Macrophage inflammatory protein-1. *Cytokine & growth factor reviews* **13**(6): 455-481.

Mikhail M, Wang B, Saksena NK. 2003. Mechanisms involved in non-progressive HIV disease. *AIDS reviews* **5**(4): 230-244.

Modi WS. 2004. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. *Genomics* **83**(4): 735-738.

Modi WS, Lautenberger J, An P, Scott K, Goedert JJ, Kirk GD, Buchbinder S, Phair J, Donfield S, O'Brien SJ et al. 2006. Genetic variation in the CCL18-CCL3-CCL4 chemokine gene cluster influences HIV Type 1 transmission and AIDS disease progression. *American journal of human genetics* **79**(1): 120-128.

Mummidi S, Bamshad M, Ahuja SS, Gonzalez E, Feuillet PM, Begum K, Galvis MC, Kostecki V, Valente AJ, Murthy KK et al. 2000. Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. *The Journal of biological chemistry* **275**(25): 18946-18961.

O'Brien SJ, Nelson GW. 2004. Human genes that limit AIDS. *Nature genetics* **36**(6): 565-574.

Oravecz T, Pall M, Norcross MA. 1996. Beta-chemokine inhibition of monocytotropic HIV-1 infection. Interference with a postbinding fusion step. *Journal of immunology* **157**(4): 1329-1332.

Paintsil E, Andiman WA. 2009. Update on successes and challenges regarding mother-to-child transmission of HIV. *Current opinion in pediatrics* **21**(1): 94-101.

Palella FJ, Jr., Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ, Holmberg SD. 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *The New England journal of medicine* **338**(13): 853-860.

Pantaleo G, Menzo S, Vaccarezza M, Graziosi C, Cohen OJ, Demarest JF, Montefiori D, Orenstein JM, Fox C, Schrager LK et al. 1995. Studies in subjects with long-term nonprogressive human immunodeficiency virus infection. *The New England journal of medicine* **332**(4): 209-216.

Parmley JL, Hurst LD. 2007. Exonic splicing regulatory elements skew synonymous codon usage near intron-exon boundaries in mammals. *Molecular biology and evolution* **24**(8): 1600-1603.

Paximadis M, Mohanlal N, Gray GE, Kuhn L, Tiemessen CT. 2009. Identification of new variants within the two functional genes CCL3 and CCL3L encoding the CCL3 (MIP-1alpha) chemokine: implications for HIV-1 infection. *International journal of immunogenetics* **36**(1): 21-32.

Pedrosa E, Carretero-Iglesia L, Boada A, Colobran R, Faner R, Pujol-Autonell I, Palou E, Esteve A, Pujol-Borrell R, Ferrandiz C et al. 2011. CCL4L polymorphisms and CCL4/CCL4L serum levels are associated with psoriasis severity. *J Invest Dermatol* **131**(9): 1830-1837.

Pereyra F, Addo MM, Kaufmann DE, Liu Y, Miura T, Rathod A, Baker B, Trocha A, Rosenberg R, Mackey E et al. 2008. Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis* **197**(4): 563-571.

Perry RT, Nyaku MK, Aissani B, Shrestha S. 2014. Identification of sequence variants in the CCL3 chemokine gene family in the HapMap West African reference population. *Human immunology* **75**(8): 854-858.

Petersen DC, Kotze MJ, Zeier MD, Grimwood A, Pretorius D, Vardas E, Janse van Rensburg E, Hayes VM. 2001. Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis assay. *AIDS* **15**(2): 171-177.

Picton AC, Paximadis M, Tiemessen CT. 2012a. CCR5 promoter haplotypes differentially influence CCR5 expression on natural killer and T cell subsets in ethnically divergent HIV-1 uninfected South African populations. *Immunogenetics* **64**(11): 795-806.

-. 2013. Contribution of variable CCL3L copy number to CCL3 protein production in two ethnically divergent South African populations. *Infect Genet Evol* **14**: 347-356.

Picton AC, Shalekoff S, Paximadis M, Tiemessen CT. 2012b. Marked differences in CCR5 expression and activation levels in two South African populations. *Immunology* **136**(4): 397-407.

Popovic M, Sarngadharan M, Read E, Gallo R. 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**(4648): 497-500.

Poropatich K, Sullivan DJ, Jr. 2011. Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression. *The Journal of general virology* **92**(Pt 2): 247-268.

Raport CJ, Gosling J, Schweickart VL, Gray PW, Charo IF. 1996. Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1beta, and MIP-1alpha. *The Journal of biological chemistry* **271**(29): 17161-17166.

Reddi A, Leeper S, Grobler A, Geddes R, France KH, Dorse G, Vlok W, Mntambo M, Thomas M, Nixon K et al. 2007. Preliminary outcomes of a paediatric highly active antiretroviral therapy cohort from KwaZulu-Natal, South Africa. *BMC Pediatrics* **7**(1): 13.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al. 2006. Global variation in copy number in the human genome. *Nature* **444**(7118): 444-454.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R et al. 2001. Linkage disequilibrium in the human genome. *Nature* **411**(6834): 199-204.

Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, Pomerantz RJ. 2009. The challenge of finding a cure for HIV infection. *Science* **323**(5919): 1304-1307.

Rodes B, Toro C, Paxinos E, Poveda E, Martinez-Padial M, Benito JM, Jimenez V, Wrin T, Bassani S, Soriano V. 2004. Differences in disease progression in a cohort of long-term non-progressors after more than 16 years of HIV-1 infection. *AIDS* **18**(8): 1109-1116.

Rostene W, Kitabgi P, Parsadaniantz SM. 2007. Chemokines: a new class of neuromodulator? *Nature reviews Neuroscience* **8**(11): 895-903.

Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M. 1996a. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* **35**(11): 3362-3367.

Samson M, Stordeur P, Labbe O, Soularue P, Vassart G, Parmentier M. 1996b. Molecular cloning and chromosomal mapping of a novel human gene, ChemR1, expressed in T lymphocytes and polymorphonuclear cells and encoding a putative chemokine receptor. *Eur J Immunol* **26**(12): 3021-3028.

Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M. 1999. A view of modern human origins from Y chromosome microsatellite variation. *Genome research* **9**(6): 558-567.

Sekaly RP. 2008. The failed HIV Merck vaccine study: a step back or a launching point for future vaccine development? *The Journal of experimental medicine* **205**(1): 7-12.

Shalekoff S, Meddows-Taylor S, Schramm DB, Donninger SL, Gray GE, Sherman GG, Coovadia AH, Kuhn L, Tiemessen CT. 2008. Host CCL3L1 gene copy number in relation to HIV-1-specific CD4+ and CD8+ T-cell responses and viral load in South African women. *J Acquir Immune Defic Syndr* **48**(3): 245-254.

Shalekoff S, Schramm DB, Lassauniere R, Picton AC, Tiemessen CT. 2013. Differences are evident within the CXCR4-CXCL12 axis between ethnically divergent South African populations. *Cytokine* **61**(3): 792-800.

Shattock RJ, Moore JP. 2003. Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Micro* **1**(1): 25-34.

Shostakovich-Koretskaya L, Catano G, Chykarenko ZA, He W, Gornalusse G, Mummidi S, Sanchez R, Dolan MJ, Ahuja SS, Clark RA et al. 2009. Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. *AIDS* **23**(6): 679-688.

Shrestha S, Tang J, Kaslow RA. 2009. Gene copy number: learning to count past two. *Nat Med* **15**(10): 1127-1129.

Siegal FP, Lopez C, Hammer GS, Brown AE, Kornfeld SJ, Gold J, Hassett J, Hirschman SZ, Cunningham-Rundles C, Adelsberg BR et al. 1981. Severe Acquired Immunodeficiency in Male Homosexuals, Manifested by Chronic Perianal Ulcerative Herpes Simplex Lesions. *New England Journal of Medicine* **305**(24): 1439-1444.

Stewart GJ, Ashton LJ, Biti RA, Ffrench RA, Bennetts BH, Newcombe NR, Benson EM, Carr A, Cooper DA, Kaldor JM. 1997. Increased frequency of CCR-5 delta 32 heterozygotes among long-term non-progressors with HIV-1 infection. The Australian Long-Term Non-Progressor Study Group. *AIDS* **11**(15): 1833-1838.

Strain MC, Lada SM, Luong T, Rought SE, Gianella S, Terry VH, Spina CA, Woelk CH, Richman DD. 2013. Highly precise measurement of HIV DNA by droplet digital PCR. *PLoS One* **8**(4): e55943.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Genomes P et al. 2010. Diversity of human copy number variation and multicopy genes. *Science* **330**(6004): 641-646.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O et al. 2009. The genetic structure and history of Africans and African Americans. *Science* **324**(5930): 1035-1044.

Tomaras GD, Haynes BF. 2014. Advancing Toward HIV-1 Vaccine Efficacy through the Intersections of Immune Correlates. *Vaccines* **2**(1): 15-35.

Townson JR, Barcellos LF, Nibbs RJ. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol* **32**(10): 3016-3026.

UNAIDS. 2013a. AIDS by Numbers-A5 Brouchure Joint United Nations Programme on HIV/AIDS (UNAIDS), Geneva.

-. 2013b. Epidemiological Status. In *AIDS Info*, Vol 2015. Joint United Nations Programme on HIV/AIDS (UNAIDS).

-. 2013c. Global Report. In *UNAIDS Report on the Global AIDS Epidemic*. Joint United Nations Programme on HIV/AIDS (UNAIDS), Geneva.

-. 2014. The GAP Report. In *Joint United Nations Programme on HIV/AIDS*. Joint United Nations Programme on HIV/AIDS

Urban TJ, Weintrob AC, Fellay J, Colombo S, Shianna KV, Gumbs C, Rotger M, Pelak K, Dang KK, Detels R et al. 2009. CCL3L1 and HIV/AIDS susceptibility. *Nat Med* **15**(10): 1110-1112.

Vesanen M, Stevens CE, Taylor PE, Rubinstein P, Saksela K. 1996. Stability in controlling viral replication identifies long-term nonprogressors as a distinct subgroup among human immunodeficiency virus type 1-infected persons. *Journal of virology* **70**(12): 9035-9040.

Vila-Coro AJ, Mellado M, Martin de Ana A, Lucas P, del Real G, Martinez AC, Rodriguez-Frade JM. 2000. HIV-1 infection through the CCR5 receptor is blocked by receptor dimerization. *Proceedings of the National Academy of Sciences of the United States of America* **97**(7): 3388-3393.

Wain LV, Armour JA, Tobin MD. 2009. Genomic copy number variation, human health, and disease. *Lancet* **374**(9686): 340-350.

Walensky RP, Paltiel AD, Losina E, Mercincavage LM, Schackman BR, Sax PE, Weinstein MC, Freedberg KA. 2006. The survival benefits of AIDS treatment in the United States. *The Journal of infectious diseases* **194**(1): 11-19.

Walker S, Janyakhantikul S, Armour JA. 2009. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics* **93**(1): 98-103.

Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ. 2012. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic acids research* **40**(11): e82.

Williamson C, Loubser SA, Brice B, Joubert G, Smit T, Thomas R, Visagie M, Cooper M, van der Ryst E. 2000. Allelic frequencies of host genetic variants influencing susceptibility to HIV-1 infection and disease in South African populations. *Aids* **14**(4): 449-451.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics* **13**: 134.

Zagury D, Lachgar A, Chams V, Fall LS, Bernard J, Zagury JF, Bizzini B, Gringeri A, Santagostino E, Rappaport J et al. 1998. C-C chemokines, pivotal in protection against HIV type 1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **95**(7): 3857-3861.

Zlotnik A, Yoshie O, Nomiyama H. 2006. The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biol* **7**(12): 243.

# APPENDIX A: SUPPLEMENTARY DATA FOR CHAPTER 2

**Table: CCL4 sequencing primers**

| *Name* | *Primer Sequence 5' to 3'* |
|---|---|
| *Seq1* | *AGCAGTCTATCAGTCAACGG* |
| *Seq2* | *TCCCTCCTCACATCCAAT* |
| *Seq3* | *ACCACTTCCCTTTTCTTCTCGCCT* |
| *Seq4* | *ACTGTCCTGTCTCTCCTCAT* |
| *Seq5* | *TGTTCAAGCAAGCCCCTGT* |
| *Seq6* | *CTGCTTTTCTTACACTGCGAGG* |
| *Seq7* | *TCAGAGGACAGGGAAGCA* |
| *Seq8* | *AATCTGTCCGCTCCTTGTTC* |
| *Seq9* | *CCACTGTCACTGTTTCTCTGCT* |
| *Seq10* | *TAAGGAAACAGAGGCACAGA* |
| *Seq11* | *CCATCATAGGACCAGGCA* |
| *Seq12* | *ATGGGACACAGACTCGTAG* |
| *Seq13* | *CACTTCTCTGGGGTCAACA* |
| *Seq14* | *GTTCCTGAGTCACACTA* |

**Table: CCL4 indel primers**

| *Name* | *Primer Sequence 5' to 3'* |
|---|---|
| *Indel1* | *ACAAGCCTGACCCCAAGAAC* |
| *Indel2* | *ATCTCCCCATTCCCTGCCA* |
| *Indel3* | *CAGTAGATTGGGATTAGAGTGC* |
| *Indel4* | *GAACTGTGCTGCTCTTGGTGG* |

# APPENDIX B: SUPPLEMENTARY DATA FOR CHAPTER 3

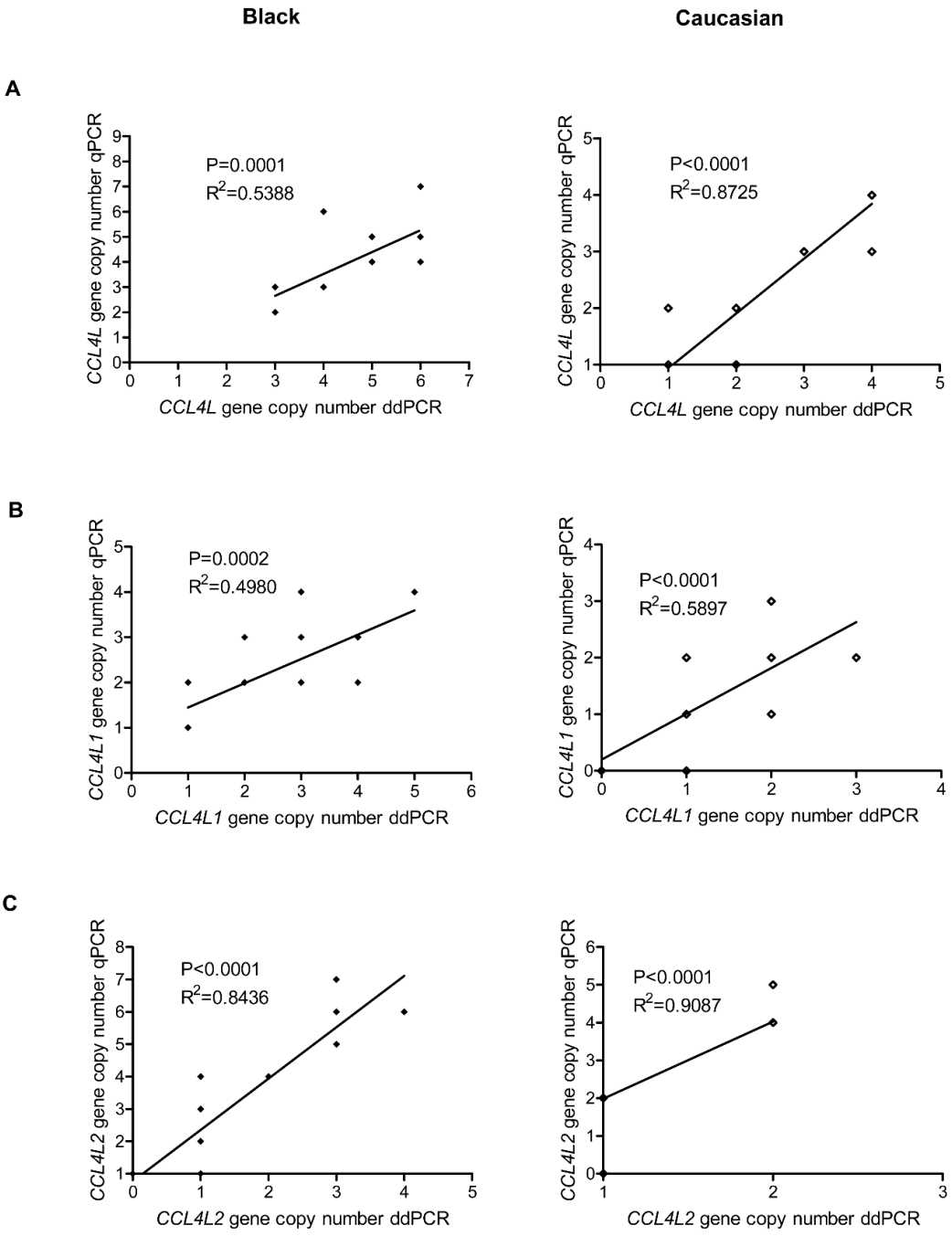



**Figure: Correlation of copy number generated by ddPCR and qPCR in the South African Black and Caucasian cohort for A: *CCL4L*, B: *CCL4L1*, C: *CCL4L2*.**

# APPENDIX C: AMINO ACID CODES

| | | |
|---|---|---|
| *A* | *alanine* | *Ala* |
| *C* | *cysteine* | *Cys* |
| *D* | *aspartate* | *Asp* |
| *E* | *glutamate* | *Glu* |
| *F* | *phenylalanine* | *Phe* |
| *G* | *glycine* | *Gly* |
| *H* | *histidine* | *His* |
| *I* | *isoleucine* | *Iso* |
| *K* | *lysine* | *Lys* |
| *L* | *leucine* | *Leu* |
| *M* | *methionine* | *Met* |
| *N* | *asparagine* | *Asp* |
| *P* | *proline* | *Pro* |
| *Q* | *glutamine* | *Glu* |
| *R* | *arginine* | *Arg* |
| *S* | *serine* | *Ser* |
| *T* | *threonine* | *Thr* |
| *V* | *valine* | *Val* |
| *W* | *tryptophan* | *Trp* |
| *Y* | *tyrosine* | *Tyr* |

# APPENDIX D: ETHICAL CLEARANCE

**HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)**

**CLEARANCE CERTIFICATE NO. M130869**

| | |
|---|---|
| **NAME: (Principal Investigator)** | Ms Avani Bharuthram |
| **DEPARTMENT:** | Centre for HIV and STIs/Virology<br>National Institute for Communicable Diseases |
| **PROJECT TITLE:** | The Role of CCl4/MIP-1B Encoding Genes in HIV Infection |
| **DATE CONSIDERED:** | Ad hoc |
| **DECISION:** | Approved unconditionally |
| **CONDITIONS:** | |
| **SUPERVISOR:** | Prof C Tiemessen |
| **APPROVED BY:** | Professor PE Cleaton-Jones, Chairperson, HREC (Medical) |

**DATE OF APPROVAL:** 04/09/2013

**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 Secretary in Room 10004, 10th floor, Senate House, University.

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 resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature

13/09/2013
M130869Date

**PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES**

# APPENDIX E: TURN IT IN REPORT

**Turnitin Originality Report**
forturnitin.docx by Avani Bharuthram
From Masters_Research
(2q76fka45jF6aS3A99AM8hSZeVKhoZZ1xoQqF1Y90M7DM97zwfR6C8dUC2X0Hvi8hJ3M3F717S3a46E39IeiD6ajZ182P3Zz0a8)

- Processed on 09-Apr-2015 4:46 PM SAST
- ID: 526411554
- Word Count: 21429

**Similarity Index 25%**
Similarity by Source
Internet Sources: 5%
Publications: 24%
Student Papers: 2%

**sources:**

**1** 19% match (publications)
Bharuthram, Avani, Maria Paximadis, Anabela C.P. Picton, and Caroline T. Tiemessen. "Comparison of a quantitative Real-Time PCR assay and droplet digital PCR for copy number analysis of the CCL4L genes", Infection Genetics and Evolution, 2014.

**2** 1% match (publications)
"Poster Sessions", Immunology, 2012.

**3** 1% match (Internet from 19-Oct-2014)
http://www.science.gov/topicpages/t/time+pcr+revealed.html

**4** < 1% match (student papers from 10-Nov-2014)
Submitted to University of Witwatersrand on 2014-11-10

**5** < 1% match (Internet from 08-Sep-2003)
http://ii1.ap.siedlce.pl/~klopotek/DAI/wekaexamples/splice.arff

**6** < 1% match (publications)
Picton, Anabela C.P., Maria Paximadis, and Caroline T. Tiemessen. "Contribution of variable CCL3L copy number to CCL3 protein production in two ethnically divergent South African populations", Infection Genetics and Evolution, 2013.

**7** < 1% match (Internet from 15-Jul-2008)
http://www.qc.dfo-mpo.qc.ca/peches/fr/statistique/pdf1997/PRODUCT_QC.pdf

**8** < 1% match ()
http://www.tali-welt.de/TW_de/stoff/humancheck/genomexample.php

**9** < 1% match (publications)
M. Paximadis. "Identification of new variants within the two functional genes *CCL3* and *CCL3L*encoding the CCL3 (MIP-1α) chemokine: implications for HIV-1 infection", International Journal of Immunogenetics, 02/2009

**10**< 1% match (Internet from 18-May-2011)
http://array.bioengr.uic.edu/~yangdai/pub/Pharmacotherapy.pdf