

https://dspace.org/

The role of CCL3 and CCL4 encoding genes in control of HIV-1 infection

Mncube, Sizanani

2017
https://hdl.handle.net/10539/24884

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

The role of CCL3 and CCL4 encoding genes in control of HIV-1 infection

Sizanani Mncube

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

Supervisors:

Professor Caroline T. Tiemessen

Dr Maria Paximadis

Johannesburg, 2017

Declaration

I, Sizanani Mncube (Student number: 541681) am a student registered for the degree of MSc (Med) Virology in the academic year 2017.

I hereby declare the following:

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

Signature: _____ Date: _____

Dedication

I dedicate this work to my brother

Thank you for all the sacrifices you have made for my education,

I will forever be grateful

Abstract

Two CC chemokine ligands that bind the CCR5 receptor (major coreceptor used by HIV-1), namely CCL3 and CCL4, are encoded by the *CCL3* and *CCL4* genes found in 2 copies per diploid genome and their non-allelic variants, *CCL3L* and *CCL4L*, respectively, which are found in variable copy numbers. *CCL3L* and *CCL4L* gene copy numbers lower than the population specific median have been associated with various HIV-1 related outcomes including increased susceptibility to HIV-1 infection, faster rate of progression to AIDS and high viral loads. These associations have however been controversial and the controversy has in part been attributed to technical challenges of accurate high copy number determination. In this study we investigated the association of the variable copy number genes, *CCL3La* and *CCL3Lb* (the sum of which is *CCL3L*) and *CCL4La* and *CCL4Lb* (the sum of which is *CCL4L*) and previously identified *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2SNP) in natural control of HIV-1 infection in the South African black population.

The study was conducted on antiretroviral (ARV) naive HIV-1-infected controllers (HICs, N=52), comprised of elite controllers (ECs, N=11), viraemic controllers (VCs, N=30) and high viral load long term non progressors (HVL LTNPs, N=11) and HIV-1-infected progressors (N=74). HIV-1 controllers were also categorised and analysed on the basis of their viral loads as individuals with less than 400 RNA copies/ml (<400 HICs, N=20) and greater than 400 RNA copies/ml (>400 HICs, N=32). Droplet digital PCR (ddPCR), which is a much more accurate method for determination of high copy numbers compared to real-time PCR, was used to determine *CCL3L* and *CCL4L* gene copy numbers. We developed and validated ddPCR assays for *CCL3L* copy number determination and used previously developed ddPCR assays for *CCL4L* copy number determination. *CCL3L* and *CCL4L* gene copy numbers were compared between HICs, HIC subgroups and progressors, individually or in combination as continuous variables or stratified around the population median. *CCL3* haplotypes were determined using previously designed C_T shift real-time PCR assays and these were compared between HICs, HIC subgroups and progressors.

We found *CCL3La* copy number to be equal to *CCL4L* copy number in all individuals (with the exception of one individual who had 8 copies of *CCL3La* and 7 copies of *CCL4L*), thus any associations involving these genes would mirror each other and were designated as *CCL3La*/*CCL4L*. *CCL3L* and *CCL4L* copy numbers compared as continuous variables did not yield any significant differences between HICs, HIC subgroups and progressors. However, a strong trend of higher representation of individuals with high *CCL3La*/*CCL4L* copy numbers was seen in VCs (p=0.07) compared to progressors. When individuals were grouped and compared according to copy number > population median (high) and ≤ population median (low), significant overrepresentations of individuals with *CCL3La*/*CCL4L* copy number > population median were

seen when comparing VCs (p=0.03, OR=0.39) and <400 HICs (p=0.01, OR=0.21) to progressors. Comparing the total controller group (HICs) to progressors, however only showed a strong trend of overrepresentation of individuals with *CCL3La*/*CCL4L* copy number > population median (p=0.05, OR=0.4). Comparison of HICs, HIC subgroups and progressors having high *CCL3L* and/or high *CCL4L* copy numbers revealed a significant difference in VCs (p=0.04, OR=0.28) compared to progressors, with VCs having a higher representation of individuals with high *CCL3L* and/or *CCL4L*. Combined associations were however more significant when comparing individuals with high *CCL3La* and/or *CCL4La* with HICs (p=0.01, OR=0.38), VCs (p=0.02, OR=0.31) and <400 HICs (p=0.01, OR=0.21) having a higher representation of individuals with high *CCL3La* and/or *CCL4La* compared to progressors.

Given that the *CCL3L* and *CCL4L* gene copy numbers occur in different combinations in different individuals, we next explored the specific patterns of copy number possession by assigning pattern identifiers (IDs) to each individual based on the number of copies of *CCL3L*, *CCL3La* and *CCL3Lb* or *CCL4L*, C*CL4La* and *CCL4Lb*. The 651 *CCL3L* pattern (i.e 6 copies of *CCL3L*, 5 copies of *CCL3La* and 1 copy of *CCL3Lb*) was overrepresented in ECs (p=0.02, OR=0.17) and <400 HICs (p=0.04, OR=0.26) whereas a 633 *CCL4L* pattern was overrepresented in HICs (p=0.03, OR=0.07) and >400 HICs (p=0.03, OR=0.06) with a combined 550541 *CCL3L* and *CCL4L* pattern being overrepresented in VCs (p=0.02, OR=0.05). Moreover, we found a *CCL3La* and *CCL3Lb* copy number difference of 2 to be underrepresented in HICs (p=0.03, OR=3.58) compared to progressors, with a copy number difference of 5 being overrepresented in VCs (p=0.02, 0.29) when compared to progressors.

No significant differences were seen when *CCL3* Hap-A1, Hap-A3 and Hap-2SNP allelic and genotypic frequencies were compared between HICs, HIC subgroups and progressors, however a trend of higher representation of Hap-A1 heterozygosity was seen in >400 HICs when compared to progressors (p=0.09, OR=0.14) or healthy HIV-1 uninfected controls (p=0.09, OR=0.20). Additionally, combined analysis of *CCL3* haplotypes and *CCL3La* copy numbers does not seem to be more significantly associated with natural control of HIV-1 infection compared to *CCL3La* copy numbers alone, with the exception of *CCL3La* and Hap-A3 which showed a slightly higher significant association in VCs (p=0.02, OR=0.11) compared to progressors.

Overall, our findings suggest that the CCL3 and CCL4 chemokines play a role in natural control of HIV-1 infection, in agreement with previous studies which found high *CCL3L* copy numbers to be protective against disease progression. However, future studies with larger sample sizes would be beneficial to further elucidate the associations of the *CCL3L* and *CCL4L* gene copy numbers with natural control of HIV-1.

Acknowledgments

I would like to express my sincere gratitude to the following people:

- Firstly, to my supervisors, Prof. Caroline Tiemessen and Dr Maria Paximadis, for their continuous guidance, encouragement and mostly their patience. I am very grateful to have been under your supervision.
- To everyone in the Tiemessen lab for being very kind and making the work environment very pleasant, and mostly for opening my eyes to new things that I never thought I would try.
- To Dr Shayne Loubser for taking his time to help me with the endless Excel related questions and so much more.
- To the study participants, for their selflessness, without them this study would not have been possible
- To my family and friends for all the love, encouragement and moral support they have given me.

I would also like to acknowledge funding from the National Research Foundation of South Africa and the University of the Witwatersrand (Postgraduate Merit Award).

Table of contents

Declaration i

Dedication ii

Abstract iii

Acknowledgments v

Table of contents vi

List of figures ix

List of tables xi

List of abbreviations xiii

Presentations xiv

CHAPTER 1: Introduction 1

1.1 Background 2

1.2 Mechanism of HIV-1 cellular entry 3

1.3 The course of HIV-1 infection 4

1.4 Natural control of HIV-1 infection 6

1.5 Host mechanisms of HIV-1 control 7

1.5.1 Variation in Human Leukocyte Antigen and Killer Immunoglobulin-like Receptor molecules 8

1.5.2 Variation in chemokine receptors 9

1.6 Chemokines 12

1.6.1 CC chemokines 12

1.6.2 Role of CCL3 and CCL4 in HIV-1 infection 18

1.6.3 Variation in the *CCL3* and *CCL4* genes 20

1.7 Study rationale 21

1.7.1 Aim of the study 22

1.7.2 Specific objectives 22

CHAPTER 2: Materials and Methods 23
2.1 Study populations 24
2.2 DNA extraction 26
2.3 *CCL3* and *CCL4* nomenclature 26
2.4 Copy number determination using droplet digital PCR 27
2.4.1 Restriction enzyme digestion 27
2.4.2 Adaptation of a *CCL3L* real-time PCR assay to ddPCR format 29
2.4.3 *CCL3L* and *CCL4L* copy number determination 30
2.5 Determination of *CCL3* haplotypes 31
2.5.1 Haplotype description 31
2.5.2 Real-time PCR assays for Hap-A1 and Hap-A3 Genotyping 32
2.6 Data analysis 33
CHAPTER 3: Results 35
3.1 The role of *CCL3L* and *CCL4L* copy numbers in HIV-1 control 36
3.1.1 *CCL3L* assay design 36
3.1.2 Assay validation 36
3.1.3 *CCL3L and CCL4L* copy number distributions in study population 38
3.1.4 Relationship between *CCL3L* and *CCL4L* gene copy numbers 41
3.1.5 Comparison of *CCL3L* and *CCL4L* copy numbers between HICs and progressors 43
3.1.6 Comparisons of stratified *CCL3L* and *CCL4L* copy numbers between HICs and progressors 45
3.1.7 Analysis of the combined effect of *CCL3L* and *CCL4L* copy numbers 46
3.2 *CCL3L* and *CCL4L* copy number pattern representation and HIV-1 control 48
3.2.1 *CCL3L* gene copy number pattern representation 49
3.2.2 *CCL4L* gene copy number pattern representation 49
3.2.3 Combined *CCL3L* and *CCL4L* gene copy number representation 50
3.3 Relationship between *CCL3La* and *CCL3Lb* and between *CCL4La* and *CCL4Lb* gene copy numbers 54

3.3.1 *CCL3La* and *CCL3Lb* copy number difference 54

3.3.2 *CCL4La* and *CCL4Lb* gene copy numbers 57

3.4 *CCL3L* and *CCL4L* copy numbers and associations with markers of HIV-1 disease progression 57

3.5 *CCL3* haplotypes in healthy and HIV-1 infected individuals and control of HIV-1 infection 60

3.5.1 Representation of *CCL3* haplotypes, Hap-A1, Hap-A3 and Hap-2SNP in the healthy South African black population and other reference population groups 60

3.5.2 Comparisons of Hap-A1, Hap-A3 and Hap-2SNP allelic and genotypic frequencies between HIV-1 controllers and healthy controls and progressors 65

3.6 *CCL3* haplotypes and *CCL3La* copy number 70

3.7 Summary of significant findings 72

CHAPTER 4: Discussion and conclusion 73

REFERENCES 88

APPENDIX A: Supplementary data for Chapter 3 98

APPENDIX B: Ethical clearance 99

APPENDIX C: Turnitin report 100

List of figures

Figure 1.1: Estimation of people living with HIV in 2015 in different regions. 2
Figure 1.2: The process of HIV-1 entry into the host cell..... 4
Figure 1.3: Typical course of HIV-1 infection. 6
Figure 1.4: Schematic representation of the *CCR5* haplotypes. The seven haplotypes are comprised of the *CCR2* SNPs (*CCR2 V64I*), CCR5 promoter region SNPs and the *CCR5*Δ*32* mutation..... 11
Figure 1.5: Structure of chemokine classes showing the conserved N terminal cysteine residues..... 12
Figure 1.6: Schematic representation of the CCL3 and CCL4 encoding genes. 15
Figure 2.1: Schematic representation of the CCL3 and CCL4 encoding genes showing relative positions of the genes on chromosome 17, nomenclature of the genes and the orientation of the genes indicated by the direction of the arrows..... 27
Figure 2.2: A flow diagram showing the basic workflow of ddPCR for copy number determination in this study. 28
Figure 2.3: Schematic representation of Hap-A1, Hap-A3 and Hap-2SNP in the *CCL3* gene. 32
Figure 3.1: One dimensional (1D) amplitude plots of the *CCL3L* assays 37
Figure 3.2: Correlation (Spearman's rank correlation) between the (A) total *CCL3L* gene copy numbers and the sum of *CCL3La* and *CCL3Lb* and (B) the total *CCL4L* gene copy numbers and the sum of *CCL4La* and *CCL4Lb*. 38
Figure 3.3: Bar graph showing representation of *CCL3L* and *CCL4L* gene copy number distributions in HICs (N=52) and progressors (N=74). 40
Figure 3.4: Correlations (Spearman's rank correlation) of *CCL3L* and *CCL4L* gene copy numbers. 42
Figure 3.5: Correlation (Spearman's rank correlation) of the *CCL3La* copy numbers to the *CCL4L* copy numbers 43
Figure 3.6: Comparisons (Mann–Whitney U test) of the distribution of the (A) *CCL3L*, *CCL3La* and *CCL3Lb* and (B) *CCL4L*, *CCL4La* and *CCL4Lb* copy numbers between HICs (N=52) and Progressors (N=74), ... 44

Figure 3.7: Comparisons (Mann–Whitney U test) of the *CCL3La/CCL4L* copy numbers between VCs (N=30) and progressors (N=74), showing strong trends of higher *CCL3La*/*CCL4L* copy numbers in VCs. 44
Figure 3.8: Copy number representations in HICs, HIC subgroups and progressors stratified around the population median. 47
Figure 3.9: A bar graph showing the representation of individuals with high *CCL3La* and/or high *CCL4La* copy numbers in HICs, HIC subgroups and progressors. 48
Figure 3.10: Schematic representation of *CCL3L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies 51
Figure 3.11: Schematic representation of *CCL4L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies 52
Figure 3.12: Schematic representation of the combined *CCL3L* and *CCL4L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies 53
Figure 3.13: Representation of the difference between *CCL3La* and *CCL3Lb* (i.e *CCL3La* – *CCL3Lb*) copy numbers in HICs, HIC subgroups and progressors..... 55
Figure 3.14: Distributions (A) and comparisons (B) of the number of individuals with CCL4La > CCL4Lb, CCL4La = CCL4Lb and CCL4La < CCL4Lb in HICs, HICs subgroups and progressors. Fisher's exact test was used for statistical analysis..... 58
Figure 3.15: Correlations (Spearman's rank correlation) of (A) *CCL3L* and (B) *CCL3La*/*CCL4L* copy numbers with viral loads for the total group of HIV-1 infected individuals (HICs and progressors, N=126)..... 59
Figure 3.16: Comparison of (A) viral load and (B) $CD4^+$ T cell count (Mann-Whitney's U test) stratified according to *CCL3La*/*CCL4L* copy numbers..... 60
Figure 3.17: Linkage disequilibrium plots of the SNPs making up Hap-A1, Hap-A3 and Hap-2SNP haplotypes..... 63
Figure 3.18: Allelic (i) and genotypic (ii) frequencies of (A) Hap-A1, (B) Hap-A3 and (C) Hap-2SNP tagSNPs 64
Figure 3.19: Bar graphs showing representation of (A) Hap-A1, (B) Hap-A3 and (C) Hap-2SNP allelic (i) and genotypic (ii) frequencies 67
Figure 3.20: Graphs showing the distribution of individuals with (A) high *CCL3La* copy numbers (> population median) and/or Hap-A1 (B) high *CCL3La* copy numbers and Hap-A3 and (C) high *CCL3La* copy number and Hap-2SNP in HICs (N=52), HIC subgroups (VCs, N=30; <400 HICs, N=20) and progressors (N=74). Fisher's exact test was used for statistical analysis. 71

List of tables

Table 2.1: Characteristics of HIV-1 infected individuals from the black South African population including HICs, HIC subgroups and progressors. 25

Table 2.2: Primers and probes designed for development of *CCL3La* assay for ddPCR. 30

Table 3.1: *CCL3L* (*CCL3La* and *CCL3Lb*) and *CCL4L* (*CCL4La* and *CCL4Lb*) gene copy number medians for the study population (HICs and progressors) as well as healthy controls (HCs). 39

Table 3.2: Table showing comparisons of *CCL3La* and *CCL3Lb* copy number differences in HICs and HIC subgroups compared with progressors. 56

Table 3.3: Total number and frequency (%) of Hap-A1, Hap-A2 and Hap-2SNP tagSNP alleles and genotypes in healthy controls, HIV-1 controllers, progressors and HIV-1 controller subgroups. 66

Table 3.4: Comparisons of Hap-A1, Hap-A3 and Hap-2SNP allelic frequencies between healthy controls and HICs, and HIC subgroups and between progressors and HICs and HIC subgroups. 68

Table 3.5: Comparisons of Hap-A1, Hap-A3 and Hap-2SNP genotypic frequencies between healthy controls and HICs, and HIC subgroups and between progressors and HICs and HIC subgroup. 69

Table 3.6: Summary of all significant findings when HICs and HIC subgroups were compared to progressors. 72

List of abbreviations

1D	One dimensional
AFR	African
AIDS	Acquired immune deficiency syndrome
ARV	Anti-retroviral
BGB	Betaglobin gene
CCL3	C–C motif chemokine ligand 3
CCL3L	C–C motif chemokine ligand 3-like
CCL4	C–C motif chemokine ligand 4
CCL4L	C–C motif chemokine ligand 4-like
CCL5	C–C motif chemokine ligand 5
CCR1	C-C chemokine receptor type 1
CCR2	C-C chemokine receptor type 2
CCR3	C-C chemokine receptor type 3
CCR5	C-C chemokine receptor type 5
CD26/DPP IV	CD26/Dipeptidyl peptidase IV
CI	Confidence interval
CTL	Cytotoxic T lymphocyte
CXCR4	C-X-C chemokine receptor type 4
ddPCR	Droplet digital PCR
DNA	Deoxyribonucleic acid
ECs	Elite controllers
EUR	European
FAM	Carboxyfluorescein
GAG	Glycosaminoglycans
gPID	Gene copy number pattern identifier
HH(A)	Human haplotype A
HICs	HIV-1 controllers
HIV	Human immunodeficiency virus
HIV-1	HIV type 1
HIV-2	HIV type 2
HLA	Human leukocyte antigen
HVL LTNPs	High viral loads LTNPs
i.e.	That is
IL	Interleukin
IQR	Interquartile range
KIR	Killer-cell Immunoglobulin-like Receptors
LNA	Locked nucleic acid
LPS	Lipopolysaccharides
LTNPs	Long term non progressors
MGB	Minor grove binding
MIP	Macrophage Inflammatory Proteins
mRNA	Messenger RNA
MT	Mutant allele

NK	Natural killer
NTC	Non template controls
OR	Odds ratio
P	Significance level
PBMC	Peripheral blood mononuclear cells
PBMCs	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
Pdg	Per diploid genome
PHA	Phytohemagglutinin
PRT	Paralogue ratio test
RANTES	Regulated upon Activation, Normal T-cell Expressed, and Secreted
RNA	Ribonucleic acid
SAA	South African population
SNP	Single nucleotide polymorphism
UNAIDS	Joint United Nations HIV and AIDS programme
VCs	Viraemic controllers
VL	Viral load
vs.	Versus
WHO	World health organisation
WT	Wildtype
YRI	Yoruba

Presentations

Oral presentation: University of the Witwatersrand, Faculty of Health Sciences, Research Day & Postgraduate Expo, 1 Sep, 2016. *CCL3L* and *CCL4L* gene copy numbers and control of HIV-1 infection.

CHAPTER 1: Introduction

1.1 Background

Acquired Immunodeficiency syndrome (AIDS), caused by the Human Immunodeficiency Virus (HIV) was first reported in homosexual males in 1981 (Gottlieb et al., 1981, Siegal et al., 1981) and has since become the most serious health pandemic in the world. It is estimated that over 78 million people have been infected with HIV since the beginning of the epidemic and approximately 35 million people have died from HIV/AIDS-related illnesses (UNAIDS, 2016). According to the Joint United Nations HIV and AIDS programme (UNAIDS), there were approximately 36.7 million people living with HIV/AIDS worldwide in 2015 (Figure 1), with 2.1 million new infections.

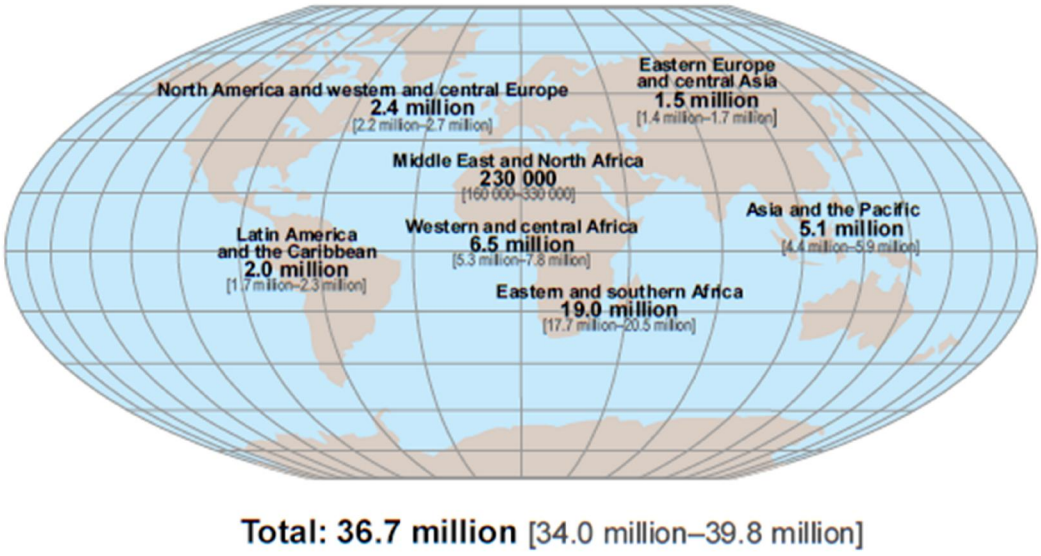


Figure 1.1: Estimation of people living with HIV in 2015 in different regions. Image source: UNAIDS (2016).

Sub-Saharan Africa is the region most affected by the HIV epidemic, with 25.6 million people living with HIV in 2015, which accounts for over 60% of people living with HIV worldwide (UNAIDS, 2016). South Africa is among the countries with the highest prevalence of HIV in the World, with approximately 7 million people living with HIV/AIDS in 2015 (UNAIDS, 2016). The HIV prevalence rate in South Africa is estimated to be 19.2% among adults aged between 15 and 49.

There are two different types of HIV which have been described, HIV type 1 (HIV-1) and HIV type 2 (HIV-2). HIV-1 and HIV-2 share similarities in their mode of transmission, replication pathways, gene arrangements and clinical consequences (Nyamweya et al., 2013), but HIV-1 has higher transmissibility and causes a more pathogenic infection with progression to AIDS being faster compared to HIV-2 infection (Gilbert et al., 2003). HIV-1 is more prevalent compared to HIV-2, accounting for 95% of the global HIV infections. HIV-2 infections are most commonly found in West African populations (Gilbert et al., 2003).

HIV infection has no cure and the disease is currently being managed with the use of anti-retroviral (ARV) drugs which have successfully reduced morbidity and mortality rates (Palella et al., 1998). Although the use of ARV therapy has been a success, it has multiple shortcomings which include severe side effects in some individuals (Currier et al., 2008, Choi et al., 2009) and rapid rebound of viral replication if treatment is interrupted (Frost et al., 2002). In addition, the health and immune system of infected individuals are not completely restored by the use of ARVs (Currier et al., 2008, Choi et al., 2009, Piot and Quinn, 2013), and some individuals, mostly in sub-Saharan Africa, do not have access to ARVs. Thus, it is important that we continue to work towards finding a cure for HIV-1 infection.

1.2 Mechanism of HIV-1 cellular entry

$CD4^+$ T lymphocytes are the cells which are preferentially infected by HIV-1, but macrophages and dendritic cells can also be infected at lower levels (Douek et al., 2002). Viral replication is initiated by the binding and entry of the virus into the target host cell which allows the delivery of the viral core to the cytoplasm (Wilen et al., 2012). The process of HIV-1 cellular entry is shown in Figure 1.2. The initial adhesion of the HIV-1 to the host cell is facilitated by the binding of viral envelope (Env) protein to the $CD4^+$ T cell receptor which is the primary HIV receptor on the host cell (Chan and Kim, 1998). The viral envelope is made up of trimers of two different subunits, namely the gp120 and gp41, which are surface and transmembrane subunits, respectively (Wilen et al., 2012).

The gp120 subunit binds with high affinity to the $CD4^+$ T cell receptors resulting in a conformational change of the subunit which allows the subsequent binding to the chemokine co-receptors, predominantly CCR5 or CXCR4 depending on the infecting virus strain. Typically, R5 viruses use the CCR5 receptors while X4 viruses use CXCR4 receptors, however some strains can use both receptors and are known as R5X4 viruses (Berger et al., 1998). The coreceptor binding initiates membrane fusion by enabling the movement of the HIV-1 virion into close

contact with the target cell which allows the insertion of the gp41 hydrophobic fusion peptide to the host cell membrane followed by the six-helix bundle formation, which is the folding of each of the three gp41 molecules (Chan and Kim, 1998). This step is followed by membrane fusion which completes the first phase of the viral replication cycle.

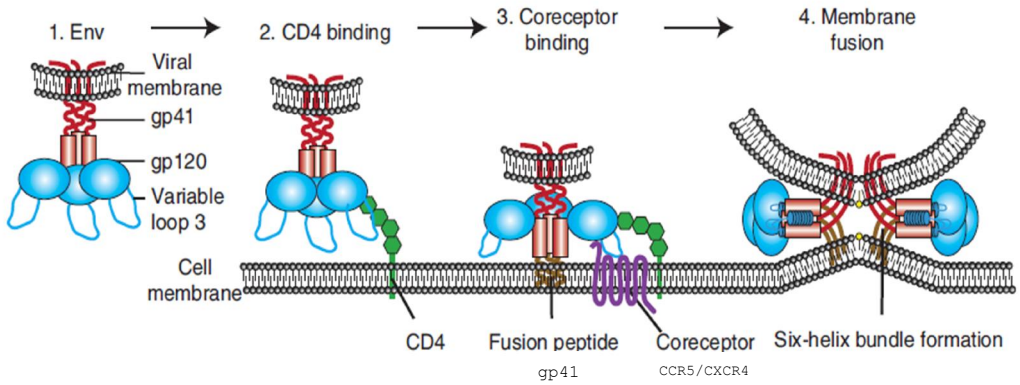


Figure 1.2: The process of HIV-1 entry into the host cell. (1) The viral envelope made up of gp120 and gp41 subunits. (2) The envelope attaches to the host cell and binds to CD4. (3) CD4 binding results in conformation change of the gp120 allowing the binding of the coreceptor (CCR5 or CXCR4). (4) The membrane fusion process is then initiated, with the insertion of the gp41 fusion peptide to the membrane and formation of the six-helix bundle. Image source: Wilen et al. (2012).

Post membrane fusion, the viral capsid is uncoated in the cytoplasm releasing HIV-1 RNA and proteins. The HIV-1 genetic material (HIV-1 RNA) is then converted to DNA using reverse transcriptase (HIV-1 enzyme) (Moss, 2013). The HIV-1 DNA is then translocated into the $CD4^+$ T cell nucleus where it is integrated into the host DNA using integrase (an HIV-1 encoded enzyme). Once integrated, transcription and translation of the virus genetic material occurs, followed by the assembling and maturation of the new viral particles which bud off from the host cell to infect new cells (Barre-Sinoussi et al., 2013).

1.3 The course of HIV-1 infection

The clinical course of HIV-1 infection is marked by progressive decline of $CD4^+$ T cell levels which result in an impaired immune system, causing susceptibility to opportunistic infections and eventually leading to death. In the absence of ARV treatment, the clinical course of HIV-1 infection is variable in different individuals, but it is typically divided into 3 different stages,

namely, acute infection stage, clinical latency and AIDS onset as shown in Figure 1.3 (An and Winkler, 2010).

The acute infection stage (HIV-1 primary infection) is the period between initial HIV-1 infection and the development of an antibody response (seroconvertion) (Kassutto and Rosenberg, 2004). This stage generally lasts between 2 and 12 weeks, but it can be as long as 8 months in some individuals (Piot and Colebunders, 1987, Touloumi and Hatzakis, 2000). The first phase of the primary HIV-1 infection is the eclipse phase with a duration of 1 to 2 weeks (Coffin and Swanstrom, 2013). The eclipse phase is characterized by uncontrollable viral replication and spreading from the initial site of infection. During the eclipse phase, both the viraemia and immune response are not detectable (Coffin and Swanstrom, 2013). The acute phase follows the eclipse phase, and in 40-90% of infected individuals it is associated with flu-like symptoms which include fever, headache, enlarged lymph nodes and fatigue (Schacker et al., 1996, Kahn and Walker, 1998). During the acute infection stage, there are high levels of viraemia of up to 10^7 viral RNA copies per ml of blood with peak viraemia reached between one and two weeks from the initial time infection (Coffin and Swanstrom, 2013). The HIV-1-specific immune response begins to appear during peak viraemia in the form of cytotoxic T lymphocytes (CTLs) and an antibody response (Weber, 2001). Cytotoxic T lymphocytes ($CD8^+$ T cells) target HIV-1 antigens presented on infected cells, while antibodies target the viral proteins (Coffin and Swanstrom, 2013). Although the presence of HIV-1 can be detected using the p24 antigen test during peak viraemia, HIV antibody tests are only positive after seroconversion.

$CD4^+$ T cell levels decline rapidly during the acute infection, the lowest point being reached at around day 9 from the initial time of infection (Vergis and Mellors, 2000). The emergence of an HIV-1 specific immune response results in both the decline of plasma HIV-1 RNA copies and the restoration of the $CD4^+$ T cell levels (Weber, 2001). Although the $CD4^+$ T cell levels are restored, they are not restored to pre-infection levels (Weber, 2001). The viral load set point is established at the end of the acute infection phase and this varies in different individuals (Kahn and Walker, 1998). The rate of disease progression has been inversely associated with the viral load set point, independent of the $CD4^+$ T cell levels (Mellors et al., 1996).

The duration of the chronic infection (clinical latency) is also variable in different individuals. This phase is characterised by gradual increase in viral loads and decrease in $CD4^+$ T cell levels (Coffin and Swanstrom, 2013). Patients are usually asymptomatic during the chronic infection, but swollen glands have been seen in some infected individuals (Moss, 2013). Although HIV-1 viral loads are lower in this stage than in the acute phase, the infected individual remains infectious and anti-HIV-1 antibodies are easily detectable in the blood (Moss, 2013).

Onset of AIDS is characterised by the depletion of the $CD4^+$ T cells to less than 200 cells/µl of blood and the manifestation of opportunistic infections (Qu et al., 2008). Without ARV treatment, the average time between the initial infection and onset of AIDS is between 8 and 10 years (Touloumi and Hatzakis, 2000), but it is variable in different individuals. The most common opportunistic infections include pulmonary tuberculosis, cryptococcal meningitis, pneumococcal pneumonia, oropharyngeal candidiasis and Kaposi's sarcoma (Moss, 2013).

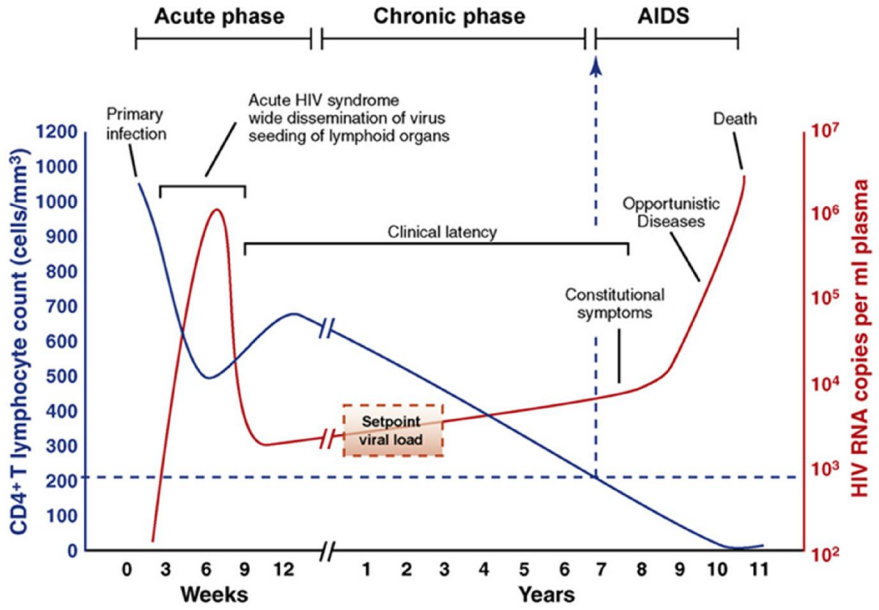


Figure 1.3: Typical course of HIV-1 infection. Image source: An and Winkler (2010).

The use of ARVs results in significant restoration of the $CD4^+$ T cell count and substantial suppression of the viral loads, resulting in reduced incidence and severity of opportunistic infections and death (Wilson and Sereti, 2013, Cockerham and Hatano, 2015). Initiation of ARV treatment during the primary infection has been shown to be more effective compared to late initiation (Zolopa et al., 2009).

1.4 Natural control of HIV-1 infection

A subset of HIV-1 infected individuals can naturally control their infection without the use of ARVs and they are collectively known as HIV-1 controllers (HICs). Within this broad group, there are individuals who exhibit different levels of control and the definitions and nomenclature

of these groups tend to vary between different studies. In this study the definitions used for the HIV-1 controllers are similar to those used in the study by Okulicz et al. (2009).

Elite controllers (ECs) account for less than 1% of HIV-1 infected individuals and they are characterised by having viral loads less than 50 RNA copies/ml of plasma (Okulicz et al., 2009). These rare individuals usually maintain normal $CD4^+$ T cell counts (Okulicz and Lambotte, 2011). Although elite controllers have good viral suppression and maintain elevated $CD4^+$ T cell counts, some elite controllers can however over time experience loss of virologic control and decline in $CD4^+$ T cell count (Okulicz and Lambotte, 2011). It has also been demonstrated that despite low levels of viral replication, some elite controllers experience progressive decline in $CD4^+$ T cell counts suggesting that even though there is viral suppression, immune activation can result in $CD4^+$ T cell depletion (Andrade et al., 2008).

Viraemic controller (VCs) are more common than elite controllers, with a reported prevalence of 3.3% and have HIV-1 viral levels of between 50 and 2000 RNA copies/ml of plasma (Okulicz et al., 2009). Viraemic control is defined by the viral load rather than duration of the infection, as low viral load set point is a good predictor of slow disease progression (Mellors et al., 1997). Unlike elite and viraemic controllers, long term non progressors (LTNPs) are more common (5 - 15% of HIV-1 infected individuals) and are defined in terms of their ability to maintain normal $CD4^+$ T cell counts regardless of the viral load (Gurdasani et al., 2014). LTNPs are described as having $CD4^+$ T cells >500 cells/µl for a period of more than 7 years after seroconversion without the use of ARVs (Gurdasani et al., 2014).

Although ECs, VCs and LTNPs are all considered HIV-1 controllers (control virus replication or disease) and likely overlap in terms of some of the mechanisms at play in their ability to control HIV-1, they are a heterogeneous group. For instance, not all ECs have high $CD4^+$ T cell count and not all LTNPs have low viral loads (Sabin and Lundgren, 2013), illustrating the importance of studying the mechanisms of control in each of these groups separately as it is possible that different mechanism are responsible for control in the different groups.

1.5 Host mechanisms of HIV-1 control

The exact mechanisms involved in viral control in HIV-1 controllers are relatively unknown and may be influenced by multiple factors including both viral and host factors. Although some individuals have been shown to be infected with a defective virus (Alexander et al., 2000), most of the natural control of HIV-1 infection seems to be at the level of the host. Some of the host factors that have been found to play a role in different HIV-1 related outcomes include: genetic

variation in genes that regulate immune response [Human Leukocyte Antigen (HLA) and Killer-cell Immunoglobulin-like Receptors (KIRs)], variation in genes that modulate HIV-1 cellular entry (chemokine receptors) and genes involved in the anti-HIV immune response (intrinsic factors) (Singh et al., 2008). Although many HIV-1 controllers have been found to be in possession of known protective host genetic factors, there are fair numbers of individuals who do not possess any known protective alleles or genotypes and some individuals with known protective host genetic factors have normal progression to AIDS, suggesting that there are other protective mechanisms and pathways which remain unknown (Cockerham and Hatano, 2015).

1.5.1 Variation in Human Leukocyte Antigen and Killer Immunoglobulin-like Receptor molecules

The HLA complex located on the short arm of chromosome 6 encodes the HLA class I and HLA class II genes. HLA class I alleles consists of *HLA-A*, *HLA-B* and *HLA-C* genes, which are the most polymorphic genes in the human genome, encoding cell surface molecules that differentially present HIV-1 antigenic epitotes to the $CD8^+$ CTLs (An and Winkler, 2010).

HLA class I alleles, the first host genetic factors identified as having an effect on AIDS have been consistently associated with different outcomes of HIV-1 infection. Heterozygosity in the *HLA-A*, *HLA-B* and *HLA-C* loci was associated with slower rate of progression to AIDS, while individuals with homozygosity in one or more loci were found to have a higher rate of progression to AIDS (Carrington et al., 1999). In Caucasian HIV-1 infected individuals, the HLA class I alleles, *HLA-B*35* and *HLA-Cw*04*, have been associated with rapid progression to AIDS (Carrington et al., 1999). In the South African black population while *HLA-B*58:*01 is associated with low viral loads, *HLA-B*58:02* is deleterious (Kiepiela, 2004).

In addition, many HIV-1 controllers have been found to be in possession of protective HLA class I alleles, specifically *HLA-B*57:01* and *HLA-B*27* which have been found to be over-represented in HIV-1 controllers compared to progressors and HIV-1 uninfected individuals (Deeks and Walker, 2007). These two alleles have been consistently associated with control of HIV-1 infection in North American and European populations (Kiepiela, 2004, Bailey et al., 2006). In African populations, the *HLA-B*57:03* allele is highly enriched among HIV-1 infected individuals able to maintain low viral levels (Kiepiela et al., 2007).

Killer Immunoglobulin-like Receptors are inhibitory and activating receptors which are expressed on the surfaces of natural killer (NK) cells and subsets of T lymphocytes. NK cells are

an important component of the innate immune response to viruses and they are involved in direct killing of the target cells and secretion of cytokines (Sowrirajan and Barker, 2011). With regards to HIV-1, NK cells are expanded during the acute phase prior to the $CD8^+$ T cell responses but HIV-1 has also evolved ways to escape NK cell recognition (Carrington and Alter, 2012). KIR encoding genes are also highly polymorphic with variable gene copy number, gene content and allelic representation (Middleton and Gonzelez, 2010).

HLA class I molecules are the ligands of KIRs (Thananchai et al., 2007). The HLA-KIR ligand interactions have been found to play a role in HIV-1 infection, with the co-occurrence of KIR3DS1 and *HLA-Bw4-80I* being associated with low viral loads and slower rate of progression to AIDS (Qi et al., 2006). In the presence of *HLA-Bw4-80I*, KIR3DL1 was also found to be associated with protection against disease progression (Martin et al., 2007). Furthermore, Pelak et al. (2011) found that in the presence of the appropriate ligand (HLA-Bw4), high *K1R3DS1* copy number was associated with low viral load set points and in the presence of *K1R3DS1* and appropriate HLA ligands, high *KIR3DL*1 copy number was also associated with low viral set point. A role of KIRs has also been seen in mother-to-child transmission of HIV-1 in the South African population, with the *KIR2DL2*/*KIR2DL*3 genotype being underrepresented in transmitting mothers (mostly intapatrum transmitting) compared to non-transmitting mothers (Paximadis et al., 2011). Moreover, Paximadis et al. (2011) also found that there was an over-representation of *KIR2DL3* homozygosity in intapatrum transmitting mothers compared to non-transmitting. The combination of the *KIR2DL3* homozygosity and HLA-C allotype heterozygosity (C1/C2) was also found to be overrepresented and underrepresented in intapatrum transmitting mothers and infected infants, respectively (Paximadis et al., 2011).

1.5.2 Variation in chemokine receptors

1.5.2.1 *CCR5* genetic variants

Polymorphisms in the *CCR5* gene, which is a co-receptor used by R5 HIV-1 strains to gain entry into the target host $CD4^+$ T cells, have been associated with resistance and natural control of HIV-1 infection. A 32 base pair deletion in the *CCR5* gene (*CCR5*Δ*32*) results in protection from HIV-1 infection in a homozygous state, while being heterozygous for the *CCR5*Δ*32* mutation results in slow progression to AIDS (Dean et al., 1996, Buseyne et al., 1998). The 32 base pair deletion introduces a premature stop codon resulting in production of a truncated CCR5 protein. The truncated protein is not expressed on the cell surface thereby restricting cellular entry of R5 HIV-1 stains in individuals homozygous for the mutation (Buseyne et al., 1998). Individuals

heterozygous for the *CCR5Δ32* mutation have reduced surface expression of the CCR5 protein compared to individuals without the deletion (Wu et al., 1997). This polymorphism is primarily found in Caucasians, with a frequency of 10 – 16% in Northern Europe (Martinson et al., 1997, Novembre et al., 2005). In African populations, the *CCR5Δ32* mutation is virtually absent (Novembre et al., 2005).

In addition to the *CCR5Δ32* mutation playing a role in natural control of HIV-1, various mutations in the *CCR5* promoter region that form distinct haplotypes have also been found to pay a role in HIV-1 infection. Gonzalez et al. (1999) identified seven distinct *CCR5* haplotypes namely, *CCR5* human haplotype (HH)-A, HHB, HHC, HHD, HHE, HHF (F*1 and F*2) and HHG (G*1 and G*2) as illustrated in Figure 1.4, with HHA being considered the wildtype/ancestral haplotype. HHG*1 and HHG*2 are differentiated by the presence of the *CCR5Δ32* mutation in the HHG*2 haplotype, while HHF*1 and HHF*2 are differentiated by a mutation in the *CCR2* gene that results in a valine to isoleucine substitution, with HHF*1 having an isoleucine. CCR2 is a chemokine receptor that can be used as an alternative co-receptor by HIV-1 and the gene is located in close proximity to *CCR5* (Kostrikis et al., 1998). The disease modifying effects of the *CCR5* haplotypes were found to be population specific. In their entire cohort consisting of Caucasians (54%), African Americans (37%), Hispanics (6%) and others (3%), Gonzalez et al. (1999) found that HHA and HHF*1 haplotypes were associated with slower rate of progression to AIDS and faster rate of progression to AIDS, respectively. The HHG*2 haplotype was associated with delayed AIDS onset and death in Caucasians most likely through its linkage disequilibrium with the *CCR5Δ32* mutation. They also found that In African Americans but not Caucasians, the HHF*2 haplotype was associated with slower rate of progression to AIDS and death while homozygosity for HHE but not heterozygosity was found to be associated with faster rate of progression to AIDS in Caucasians, but not in African Americans (Gonzalez et al., 1999). Although the effects of the *CCR5* haplotypes in HIV-1 related outcomes has not been investigated in the South African population, Picton et al. (2010) described the frequencies of these haplotypes in healthy black and Caucasian South Africans. Picton et al. (2010) found HHA and HHC to be under-and overrepresented in Caucasians, respectively. In the black South African population, HHF and HHG*1 were found to be under- and overrepresented, respectively. Additionally, Picton et al. (2010) also described a total of seven putative haplotypes in the South African populations which appeared to be extensions of the *CCR5* haplotypes described by Gonzalez et al. (1999).

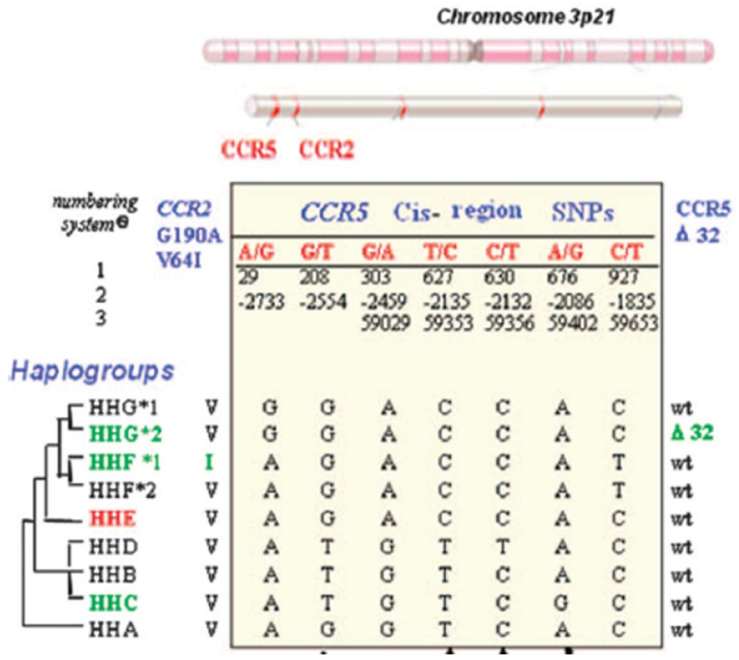


Figure 1.4: Schematic representation of the *CCR5* haplotypes. The seven haplotypes are comprised of the *CCR2* SNPs (*CCR2 V64I*), CCR5 promoter region SNPs and the *CCR5*Δ*32* mutation. Image modified from Kaur and Mehra (2009).

1.5.2.2 *CCR2* polymorphism (*CCR2 V64I*)

A polymorphism in the *CCR2* gene (*CCR2 V64I*) at position 64 within the first transmembrane region involving a valine to isoleucine substitution has been associated with slower rate of progression to AIDS (Kostrikis et al., 1998) and reduced risk of acquiring HIV-1 infection in discordant couples (Louisirirotchanakul et al., 2002), independent of the *CCR5*Δ*32* mutation. This polymorphism has also been associated with low viral load set point after seroconversion. The *CCR2 V64I* mutation is in complete linkage disequilibrium with a *CCR5* SNP (-1835T) (Kostrikis et al., 1998), as a result, there is no clarity on whether or not the association of *CCR2 V64I* with HIV-1 progression is due to the *CCR2* SNP or the *CCR5* SNP. The *CCR2 V64I* mutation is found in different racial groups with a frequency of 25% in Asians, 17% in Hispanics, 10% in Caucasians and 15% in African Americans (Smith et al., 1997).

1.6 Chemokines

Chemokines are a group of low molecular weight (8-12 kDa) chemotactic cytokines which play a role in migration of leukocytes to site of injury or infection. Functionally, chemokines are divided into homeostatic and inflammatory chemokines. Homeostatic chemokines are constitutively expressed in specific organs and tissues and they are involved in haematopoiesis, antigen sampling, immune surveillance, lymphocyte trafficking, angiogenesis and fetal development (Moser and Loetscher, 2001). Unlike homeostatic chemokines, inflammatory chemokines are induced by pro-inflammatory cytokines or contact with pathogenic agents. Inflammatory chemokines can be expressed in multiple locations by different cell types and they are mainly involved in recruitment of effector cells.

To date over 50 chemokines and over 18 chemokine receptors have been described. The chemokines are divided into 4 different subfamilies which are designated as the C, CC, CXC and CX3C chemokines according to the conserved NH_2-terminal cysteine-motifs (Figure 1.5) (Zlotnik and Yoshie, 2000, Moser and Loetscher, 2001).

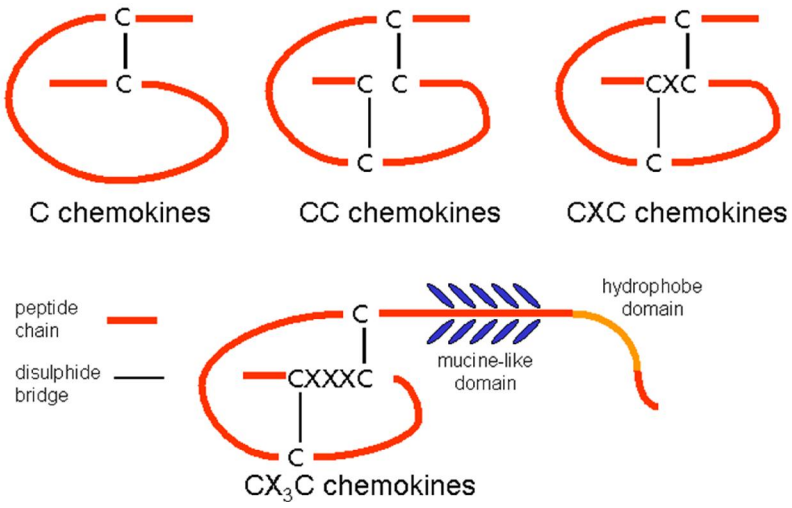


Figure 1.5: Structure of chemokine classes showing the conserved N terminal cysteine residues. Image source: (http://en.wikipedia.org/wiki/Chemokines).

1.6.1 CC chemokines

CC chemokines, previously known as β chemokines are a family of chemokines with two adjacent conserved cysteine reduces on the N-terminal and they are mostly found on a cluster of chemokines on the long arm of chromosome 17 (17q11.2) (Figure 1.6a). The CC chemokines, chemokine ligand 3 (CCL3), chemokine ligand 4 (CCL4) and chemokine ligand 5 (CCL5), also

known as MIP-1α (Macrophage Inflammatory Protein 1α), MIP-1β (Macrophage Inflammatory Protein 1β) and RANTES (Regulated upon Activation, Normal T-cell Expressed, and Secreted), respectively, are the natural ligands of the CCR5 receptor. *In vitro* these CC Chemokines have been shown to inhibit entry of HIV-1 into host $CD4^+$ T cells (Menten et al., 1999, Aquaro et al., 2001). Three mechanisms by which these CCR5 ligands inhibit entry of HIV-1 have been suggested, and they include competitive binding to CCR5 causing steric hindrance, down regulation of the receptor through desensitization and internalization and receptor dimerization (Menten et al., 2002). Prior to the identification of CCR5 as a coreceptor used by HIV-1 to gain entry into target host cells, CCL3 and CCL4 along with CCL5 were found to be the major suppressive factors produced by $CD8^+$ T cells in HIV-1 infected individuals (Cocchi et al., 1995).

1.6.1.1 The CCL3 and CCL4 Chemokines

CCL3 and CCL4 are pro-inflammatory cytokines which are inducible in most mature haematopoietic cells and they are chemoattractants of T lymphocytes, natural killer cells, dendritic cells and monocytes (Colobran et al., 2010). Both CCL3 and CCL4 recruit macrophages and lymphocytes to site of infection or inflammation, but they differ in their recruitment of specific T cells where CCL3 preferentially recruits $CD8^+$ T cells while CCL4 preferentially recruits $CD4^+$ T cells (Taub et al., 1993). Although they are inducible chemokines, low mRNA levels of CCL3 and CCL4 are constitutively expressed in monocytes and macrophages (Menten et al., 2002).

1.6.1.2 The CCL3 and CCL4 encoding genes

The CCL3 and CCL4 encoding genes are found in a hotspot for segmental duplications in a cluster of chemokine genes located on the long arm of chromosome 17 (Figure 1.6) (Hirashima et al., 1992). Segmental duplications (also known as low copy repeats) are repeated blocks of DNA sequences greater than 1kb in length which share high levels of sequence identity, usually greater than 90% (Eichler, 2001). The high sequence similarities in segmental duplications provides a substrate for homologous recombination, predisposing the region to chromosomal rearrangements (Stankiewicz and Lupski, 2002). As a result, segmental duplications have been implicated as mediators of more than 25 genomic disorders (Stankiewicz and Lupski, 2002).

CCL3 (previously known as *LD78a*, *AT464·1*, *GOS19-1*) and *CCL4* (previously known as *ACT-2*, *AT744·1*), are found in close proximity and are thought to have arisen through a duplication event of an ancestral gene (Menten et al., 2002). Both *CCL3* and *CCL4* are found in 2 copies per diploid genome (pdg) and have non-allelic copies known as *CCL3L* (previously known as *LD78b*, *AT 464·2*, *GOS19-2*) and *CCL4L* (previously known as *LAG-1*, *AT744·2*), respectively (Hirashima et al., 1992). Unlike *CCL3* and *CCL4*, *CCL3L* and *CCL4L* are found in variable copy numbers (Colobran et al., 2010).

CCL3L genes include *CCL3L1*, *CCL3L2* and *CCL3L3*. *CCL3L1* and *CCL3L3* have three exons which are identical, and they are collectively referred to as *CCL3La*. *CCL3L1* and *CCL3L3* share 95% nucleotide sequence similarity with *CCL3* which is encoded by three exons. *CCL3L2* (also referred to as *CCL3Lb*) only has 2 exons (exons 2 and 3) homologous to *CCL3L1* and *CCL3L3*, and was initially thought to be a 5'-truncated pseudogene (Hirashima et al., 1992), but novel 5' exons have recently been identified and give rise to two alternatively spliced transcripts (Shostakovich-Koretskaya et al., 2009). Although the mRNA transcripts produced by the *CCL3L2* gene have a chemokine-like domain, they are predicted not to code for a classical chemokine.

The *CCL4L* genes, *CCL4L1* and *CCL4L2* share identical sequences in the coding regions and are differentiated by a fixed mutation (A to G change) in the acceptor splice site in intron 2 of *CCL4L2* which results in production of aberrantly spliced mRNA transcripts (Colobran et al., 2005). In addition, both *CCL4* and *CCL4L1* produce alternatively spliced mRNA transcripts which lack the second exon giving rise to *CCL4Δ2* and *CCL4LΔ2* transcripts (Colobran et al., 2005). *CCL4Δ2* and *CCL4LΔ2* do not seem to have *CCL4/CCL4L1* chemokine activity (Colobran et al., 2010).

The *CCL3L* and *CCL4L* gene copy numbers are variable within different populations and range between 0 and 14 for *CCL3L* (Gonzalez et al., 2005) and 0 and 10 for *CCL4L* (Colobran et al., 2008). African populations have *CCL3L* and *CCL4L* gene copy numbers which are greater than those found in non-African populations (Gonzalez et al., 2005). The median gene copy numbers in African populations for *CCL3L* and *CCL4L* is 6 and 4, respectively. European populations have a median copy number of 2 for both *CCL3L* and *CCL4L* (Colobran et al., 2010). Both the African and European median copy numbers for *CCL3L* and *CCL4L* are comparable to the medians of the healthy South African population (Black and Caucasians) (Picton et al., 2013, Bharuthram et al., 2014). Rhesus macaques (non- human primates) have higher *CCL3L* copy numbers compared to humans and it is interesting that African populations have retained high

CCL3L copies and that Caucasians have few copy numbers, suggesting an evolutional relationship driven by differential exposure to pathogens (Gornalusse et al., 2009).

Although some studies have suggested that *CCL3L* and *CCL4L* genes are inherited as a block and that the copy number of the one gene can be used as a proxy for the other, numerous studies have however shown that although the *CCL3L* and *CCL4L* gene copy numbers are highly correlated, individuals tend to have more copies of *CCL3L* compared to *CCL4L* (Townson et al., 2002). In a Caucasian population group, Walker et al (2009) used a multiplex paralogue ratio test for *CCL3L1*/*CCL4L1* copy number determination and demonstrated that the *CCL3La* copy number (i.e. the sum of *CCL3L1* and *CCL3L3*, excluding *CCL3L2*) was equal to the total *CCL4L* copy number (sum of *CCL4L1* and *CCL4L2*), suggesting that the difference between the *CCL3L* and *CCL4L* copy number is due to the *CCL3Lb* gene which does not encode a functional chemokine. The relationship between *CCL3La* and *CCL4L* has however not been shown in South African populations (Bharuthram et al., 2014).

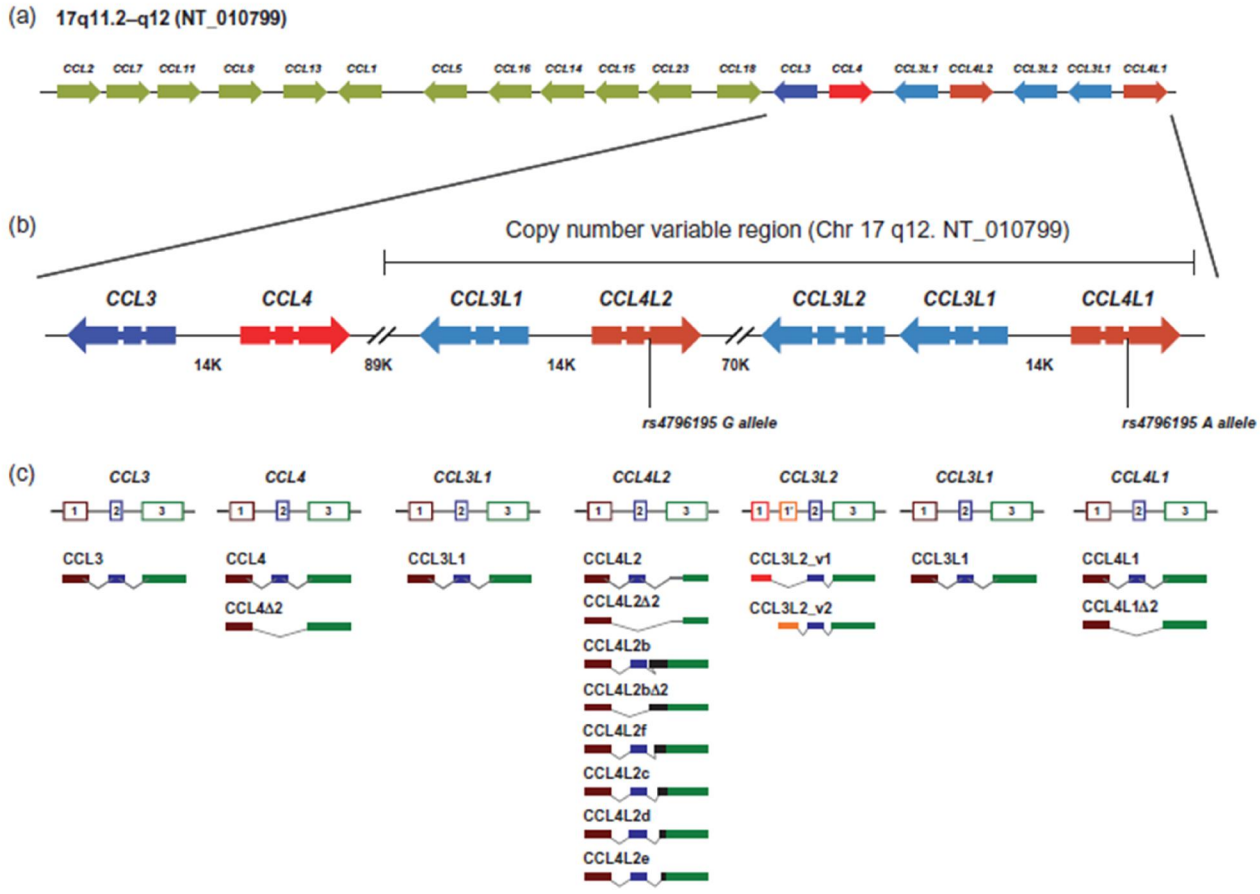


Figure 1.6: Schematic representation of the CCL3 and CCL4 encoding genes. (a) Cluster of CC chemokine encoding genes on chromosome 17, region 17q11.2 – q12. (b) Cluster of CCL3 and CCL4 encoding genes, showing directions of the genes, number of exons, distances between the genes and the SNP that differentiates *CCL4L1* and *CCL4L2*. *CCL3L1* and *CCL3L3* are both labelled as *CCL3L1* since they have identical exons. (c) Transcription patterns of the *CCL3*/*CCL3L* and *CCL4*/*CCL4L* genes, showing the mRNA transcripts produced. Image source: Colobran et al. (2010).

1.6.1.3 The CCL3 and CCL4 protein isoforms

CCL3L1 and *CCL3L3* encode for the same isoform of CCL3 (CCL3L) which differs from the *CCL3* encoded isoform by only three amino acids at positions 2, 39 and 47 (Menten et al., 1999). In position 2, CCL3L has a proline compared to serine in CCL3, while the changes at position 39 and 47 are both reciprocal swaps involving serine and glycine (Colobran et al., 2010).

Although CCL3 and CCL3L are indistinguishable using current immunological assays, they have some functional differences. In addition to using CCR5 and CCR1, CCL3L can also effectively use CCR3 as a receptor (Struyf et al., 2001). The proline residue in the second position of CCL3L results in the modification of the CCL3L by CD26/Dipeptidyl peptidase IV (DPP IV) which is a serine protease that cleaves N-terminal dipetides in proteins with proline or alanine in the second positions. The cleavage on the NH_2 – terminal in CCL3L increases its' inflammatory properties making it the most potent agonist for both CCR1 and CCR5 (Proost et al., 2000). Compared to CCL3, CCL4 and CCL5, CCL3L is the most potent ligand of the CCR5 receptor (Nibbs et al., 1999) and has the highest antiviral activity (Aquaro et al., 2001), making it the most potent chemokine in inhibiting replication of R5 strains of HIV-1 in macrophages (Menten et al., 1999, Aquaro et al., 2001). In addition, CCL3L has been shown to be much more efficient in inducing intracellular calcium signalling and chemotaxis compared to CCL3, CCL4 and CCL5 (Menten et al., 1999).

The *CCL4* and *CCL4L1* encoded protein isoforms have three amino acid differences, two in the signal peptide and one in the mature protein (Modi et al., 2001). The two different amino acids in the signal peptide involve a conservative valine to methionine change and a leucine to proline change which are both nonpolar but proline has a large aromatic ring. The difference in the mature protein is at position 47, involving a glycine to serine change which are both non-charged residues but glycine is a polar residue while serine is non-polar. It has been shown that the differences in the amino acids affect the formation of hydrogen bonds which stabilise a β loop that is important in the binding of CCL4 to glycosaminoglycans (Lodi et al., 1994, Koopmann et al., 1999), however the effect of these amino acid differences on receptor binding have not been investigated. The most abundant *CCL4L2* mRNA transcript is predicted to encode for a CCL4 isoform lacking 5 amino acids encoded by the first 15 base pairs in exon 3 (Colobran et al., 2005). Currently, no studies have looked at the functional consequences of the 5 amino acids which are absent in CCL4L2 (Colobran et al., 2010).

Due to the proline in the second position in CCL4 which results in the CD26/DPP IV modification, Guan et al. (2002) found that the naturally truncated form of $CCL4_{(3\text{-}69)}$ has the

ability to induce Ca2+ signalling though the receptors CCR1 and CCR2b, in addition to the CCR5 receptor. They also demonstrated that even though the truncation affects receptor specificity, it does not affect the anti-HIV activity. CCL4 and CCL4L1 have been shown to have functional redundancy in terms of their Ca^{2+} flux signaling assays, cell motility, competitive binding assays and anti-HIV-1 replication experiments (Townson et al., 2002, Howard et al., 2004).

In peripheral blood monocytes, CCL3 and CCL4 production can be stimulated by several inducers which include, phytohemagglutinin (PHA), interleukin-1α/β (1L-1α/β), IL-7, lipopolysaccharides (LPS), interferon-γ and lipoteichoic acids, while their expression can be inhibited by IL–4, IL-10, IL-13 and dexamethasone (Standiford et al., 1993, Berkman et al., 1995a, Berkman et al., 1995b). Several stimulation experiments have demonstrated a positive correlation between *CCL3L* copy numbers with protein production (Townson et al., 2002, Gonzalez et al., 2005, Meddows-Taylor et al., 2006). However, Carpenter et al. (2012) found no correlation between *CCL3L1* copy numbers with CCL3 expression levels in LPS-stimulated monocytes. In both HIV-1 infected and uninfected individuals, Larsen et al. (2012) did not find any correlation between total *CCL3L* copy numbers with peripheral circulating CCL3. In healthy black and Caucasians South African population, Picton et al. (2013) did not find correlations between *CCL3L* copy number and CCL3 production in PHA stimulated peripheral blood mononuclear cells (PBMCs).

It has however been shown that *CCL3L1* copy numbers are significantly correlated with *CCL3L1* mRNA expression levels, with mRNA levels increasing with higher copy number (Urban et al., 2009, Carpenter et al., 2012). This relationship suggests a possible post-transcriptional regulation of CCL3 as the effect of the copy number does not seem to be carried through to the protein level (Carpenter et al., 2012).

The effect of *CCL4L* copy numbers on CCL4 protein expression has not been as well investigated as that of CCL3. Townson et al (2002), found no correlation between *CCL4L* copy number with protein production or mRNA expression (Townson et al., 2002), however Colobran et al. (2005) found that there was higher CCL4 protein production and mRNA expression in individuals in possession of only *CCL4L1* compared to individuals in possession of only *CCL4L2*. Having a single copy of *CCL4L* has also been associated with consistently lower expression of *CCL4L1* compared to having more copies of *CCL4L* (Townson et al., 2002).

1.6.2 Role of CCL3 and CCL4 in HIV-1 infection

CCL3 and CCL4 protein expression and genetic variation in the CCL3 and CCL4 encoding genes have been associated with various diseases and infections including HIV-1. Cocchi et al. (2000) found that compared to HIV-1 uninfected individuals and individuals who progressed to AIDS (progressors), asymptomatic HIV-1 infected individuals had significantly high levels of CCL3 and CCL4 but not CCL5, which also has anti-HIV-1 activity. Walker et al. (2015), found CCL3 and CCL4 encoding genes to be upregulated in a subset of elite controllers whose $CD4^+$ T cells were resistant to R5-Tropic HIV-1. They also demonstrated that *in vitro*, the supernatant from the elite controllers with resistance to R5-Tropic HIV-1 was enough to make susceptible $CD4^+$ T cells resistant to R5 tropic HIV-1 and the resistance was reversed by adding anti-MIP antibodies.

In mother-to-child transmission of HIV-1, Meddows-Taylor et al. (2006) found that reduced production of CCL3 in cord-blood mononuclear cells from intrapartum infected infants was associated with susceptibility to HIV-1 infection. This reduced production was also evident in the matched intrapartum transmitting mothers, suggesting underlying inherent inability to produce CCL3.

1.6.2.1 *CCL3L* and *CCL4L* copy number variation and HIV-1 infection

CCL3L and *CCL4L* gene copy number variations have been implicated in various diseases and infections including hepatitis C virus infection, HCV/HIV co-infection (Grunhage et al., 2010), systemic lupus erythematosis (Mamtani et al., 2007), rheumatoid arthritis (McKinney et al., 2008), kawasaki disease (Burns et al., 2005) and acute rejection in lung transplantation (Colobran et al., 2009). In HIV-1 infection, *CCL3L* and *CCL4L* copy number variations have been associated with different HIV-1 related outcomes.

Individuals with *CCL3L1* copy number lower than the median of a specific population have been found to be at higher risk of acquiring HIV-1 (Gonzalez et al., 2005). The low *CCL3L* copy number has also been associated with faster rate of progression to AIDS, high loss of $CD4^+$ T cells (Gonzalez et al., 2005), low magnitude of HIV-1 specific $CD4^+$ and $CD8^+$ T cell responses (Shalekoff et al., 2008), high viral loads (Shalekoff et al., 2008), impaired immune recovery during antiretroviral therapy (Ahuja et al., 2008), and reduction in cell-mediated immune responses (Dolan et al., 2007).

In a study to determine the combined effect of *CCL3L* and *CCL4L* copy number in mother-to-child HIV-1 transmission in Ukrainian children, Shostakovich-Koretskaya et al. (2009) found that there was high risk of transmission if the child had less than 2 copies of *CCL3L* or *CCL4L*, with the risk being 10-fold higher if the mother was also in possession of less than 2 copies of *CCL3L* or *CCL4L*. They also found that the rate of progression to AIDS was approximately 2 times faster in HIV-1 positive children in possession of less than two copies of *CCL3L* or *CCL4L*. This study was in support of a previous study by Kuhn et al. (2007) which found that higher *CCL3L* copy numbers in infants were associated with reduced HIV-1 perinatal transmission, but the association was "attenuated" if the mother took a single dose of nevirapine or had low viral loads. In addition, Colobran et al. (2005) found that there was higher frequency of *CCL4L2* in HIV-1 infected individuals compared to controls (HIV-1 uninfected) and this was supported in another study which found that being in possession of only *CCL4L2* with no *CCL4L1* was associated with faster rate of progression to AIDS (Shostakovich-Koretskaya et al., 2009).

Although a number of studies have demonstrated the association between *CCL3L* and *CCL4L* gene copy numbers with susceptibility to HIV-1 infection, progression to AIDS and other HIV-1 related outcomes, some studies have however not been able to see similar associations (Shao et al., 2007, Bhattacharya et al., 2009, Urban et al., 2009, Larsen et al., 2012). Hence, a definitive role for CCL3 and CCL4 in HIV-1 infection control is considered to be controversial. The controversy has been linked to different factors including, difficulties in accurately measuring the gene copy numbers and generation of heterogeneous results by the assays used as well as difficulties in differentiating between the *CCL3L* genes (Shrestha et al., 2010). In addition, the cohorts in different studies are often very different and thus the study outcomes cannot always be compared.

With regards to natural control of HIV-1 infection, in a cohort of adolescents comprised mostly of female African Americans, Shao et al. (2007) did not find an association between the distribution of the *CCL3L1* and *CCL4L1* gene copy numbers with control of HIV-1 infection in HIV-1 infected controllers and intermediate controllers. In addition, Nakajima et al. (2007) did not find any significant differences in the distribution of *CCL3L* gene copy numbers (ranged from 0 to 10) between LTNPs and progressors in a cohort of HIV-1 infected and asymptomatic Japanese individuals with haemophilia.

A role of *CCL3L* has however also been seen in Rhesus macaques, a non-human model of HIV/AIDS, where low *CCL3L* copy number is associated with high rate of simian-AIDS progression in Simian immunodeficiency virus -infected macaques (Degenhardt et al., 2009). In

addition, a meta-analysis of nine *CCL3L* copy number HIV-1 association studies concluded that *CCL3L1* copy numbers are associated with HIV-1 infection, with high *CCL3L1* copy numbers associated with reduced risk of HIV-1 infection, while low *CCL3L1* copy numbers are associated with susceptibility to HIV-1 infection (Liu et al., 2010).

1.6.2.2 Copy number determination assays

Two main techniques have been used in measuring the *CCL3L* and *CCL4L* copy numbers, namely, real-time PCR and the paralogue ratio test (PRT) (Field et al., 2009). PRT is a technique that involves amplification of two products (one product from a variable copy number region of interest and another from a 2 copies pdg reference gene) using a single pair of primers (Walker et al., 2009). Capillary electrophoresis is performed on the amplified products, then copy number of the target gene is estimated by looking at the ratio of the gene of interest to the reference gene. Although real-time PCR has been the gold standard technique for measuring copy number in the CCL3 encoding genes, a number of studies have shown that real-time PCR has reduced accuracy in determining high gene copy numbers when compared to PRT and the relatively recent technology of droplet digital PCR (ddPCR) (Field et al., 2009, Bharuthram et al., 2014). Droplet digital PCR involves absolute quantification of copy numbers without the use of standard curves. In ddPCR, the sample is partitioned into thousands of droplets, and every droplet harbouring template DNA serves as a single reaction, therefore allowing numerous reactions to occur simultaneously. Bharuthram et al. (2014) clearly demonstrated that ddPCR is a much more accurate method for determining high *CCL4L* gene copy numbers compared to real-time PCR in the black South African population.

1.6.3 Variation in the *CCL3* and *CCL4* genes

SNPs and haplotypes in the genes that occur in 2 copies pdg, namely the *CCL3* and *CCL4* genes have also been studied and associated with susceptibility to HIV-1 infection and progression to AIDS. Gonzalez et al. (2001) found three *CCL3* haplotypes each comprised of two SNPs (CC, CT and TT) at positions +113 and +459 of the *CCL3* gene (rs35511254 and rs1719134) to be associated with progression and transmission of HIV-1. In European Americans, they found that the rate of progression to AIDS was faster in individuals without the wildtype haplotype which is the CC haplotype, while in African Americans the TT haplotype was associated with lower risk of acquiring HIV-1 infection. In addition, Paximadis et al. (2013) identified two seven-SNP

haplotypes exclusively on *CCL3* (i.e not present on *CCL3L*), namely Hap-A1 and Hap-A3 in the black South African population to be associated with protection to HIV-1 infection *in utero* and increased intrapartum transmission of HIV-1, respectively. The two SNP haplotype described by Gonzalez et al. (2001) form part of Hap-A1, i.e 5 additional SNPs together with the two SNPs described by Gonzalez et al. (2001) make-up Hap-A1.

Recently, a *CCL3* SNP (rs1719134) which also forms part of the Hap-A1 was found to be more frequent in healthy controls compared to HIV-1 infected individuals (da Silva et al., 2016). It is interesting to note that the rs1719134 SNP is in linkage disequilibrium with a *CCL4* SNP (rs1719153) which also has been found to be more frequent in healthy controls compared to HIV-1 infected individuals. A study investigating a total of 21 SNPs in the *CCL3*, *CCL4* and *CCL18* genes found three highly correlated *CCL3* SNPs (rs34171309, rs35511254 and rs360480177) to be more frequent in high risk exposed individuals compared to HIV-1 infected individuals from the African American population (Modi et al., 2006), two of the SNPs, rs34171309 and rs35511254 form part of the Hap-A1 haplotype described in the black South African population that was also found to be protective from *in utero* HIV-1 infection (Paximadis et al., 2013). Modi et al. (2006) also found that in European Americans, seven highly correlated SNPs across *CCL3*, *CCL4* and *CCL18* were associated with rapid progression to AIDS. One of the SNPs (rs1719130) is shared by Hap-A1 and Hap-A3 in the South African black population and another SNP (rs8951) is also part of Hap-A3. In addition, two of the SNPs rs1719134 and rs1130317 which form part of Hap-A1 did not have a protective role in European Americans (Modi et al., 2006).

In Zambian serodiscordant couples, Hu et al. (2012) found that the *CCL3* SNP rs5029410 was associated with low viral loads in seroconverters while the rs34171309 SNP was associated with an increased risk of acquiring HIV-1 infection in exposed seronegative individuals. They found the rs34171309 SNP to be linkage disequilibrium with five neighboring SNPs with similar associations (Hu et al., 2012). Interestingly, three of the SNPs also form part of Hap-A1 and like in the European Americans, these SNPs were not protective in the Zambian serodiscordant couples (Hu et al., 2012).

1.7 Study rationale

HIV-1 elite controllers represent the ideal model for understanding what constitutes HIV-1 functional cure, and studying them together with other groups of controllers and with progressors can help identify targets which can be potentially manipulated in the development of HIV-1

vaccines and novel therapeutics. Given that *CCL3* haplotypes and copy number variation in the *CCL3L* and *CCL4L* genes have been associated with susceptibility to HIV-1 infection and the rate of progression to AIDS, more in-depth investigation of these genetic variants will provide insights into the mechanisms of natural control of HIV-1. Furthermore, the association of *CCL3L* and *CCL4L* gene copy numbers with HIV-1 infection has been controversial, partly due to the reduced accuracy of real-time PCR. Improving accuracy of copy number determination through the use of ddPCR would generate more reliable results. Droplet digital PCR is a technique that provides absolute quantification with higher accuracy, precision and sensitivity. Additionally, since little is known about the combinatory representation of *CCL3L* and *CCL4L* genetic variants in our population group, this study will also describe the profiles of representation of the genes and all their variants in these cohorts.

1.7.1 Aim of the study

The main aim of this study was to investigate the role of CCL3 and CCL4 in natural control of HIV-1 infection by investigating the association of *CCL3L* and *CCL4L* copy number variations and *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2SNP) with natural control of HIV-1.

1.7.2 Specific objectives

1. To adapt a real-time PCR assay for *CCL3L* copy number determination to a ddPCR format.
2. To use the ddPCR assay to determine and compare *CCL3L* copy number between HIV-1 controllers and progressors.
3. To apply an already developed ddPCR assay to determine and compare *CCL4L* copy number between HIV-1 controllers and progressors.
4. To describe the combinatory profiles of *CCL3L* and *CCL4L* genetic variants in the context of HIV-1 control.
5. To assess the role of previously identified haplotypes in *CCL3* in HIV-1 infection control by genotyping and comparing these between HIV-1 controllers and progressors.
6. To analyse the combined effect of *CCL3L* copy number and the *CCL3* haplotypes in control of HIV-1 infection.

CHAPTER 2: Materials and Methods

2.1 Study populations

CCL3L and *CCL4L* gene copy number variations and *CCL3* haplotypes were investigated in two groups of HIV-1 infected individuals from the black South African population. The first group was comprised of HIV-1 infected controllers (HICs, N=52), recruited mainly from Johannesburg and Soweto. The inclusion criteria for HICs was based on a combination of one or more factors which include $CD4^+$ T cell count (>500 cell/µl of whole blood), viral load and duration of HIV-1 infection without the use of ARVs. HICs were divided into 3 subgroups which include: elite controllers (ECs; N=11) defined as HIV-1 infected individuals with viral loads <50 RNA copies/ml and $CD4^+$ T cell count >500 cell/µl for an unspecified length of time without the use of ARVs, selected based on at least one viral load test; viraemic controllers (VCs; N=30) defined as HIV-1 infected individuals with viral loads >50 RNA copies/ml but <2000 RNA copies/ml of plasma and $CD4^+$ T cell count >500 cell/µl, for unspecified length of time without the use of ARVs, and high viral load long term non-progressors (HVL LTNPs; N=11) defined as HIV-1 infected individuals who have been infected for >7 years with viral loads >10 000 RNA copies/ml and no demonstrable decline in $CD4^+$ T cells count without the use of ARVs. For analysis purposes, HICs were further divided based solely on their viral loads into the HICs with viral loads <400 RNA copies/ml (<400 HICs, N= 20) and viral loads >400 RNA copies/ml (>400 HICs, N=32).

The second group was comprised of HIV-1 infected progressors (N=74), defined as individuals with viral loads >10 000 RNA copies/ml and $CD4^+$ T cells declining to <350 cells/µl, requiring initiation of ARV treatment. Progressors were recruited from Chris Hani Baragwanath hospital in Soweto. Table 2.1 shows additional characteristics of our study population (HICs and progressors) including age, gender, $CD4^+$ T cell counts, viral load and time from initial HIV-1 diagnosis. The HIV-1 RNA levels (RNA copies/ml of plasma) were quantitated using the COBAS® AmpliPrep/COBAS® Taqman® HIV-1 Test, v2.0 ultrasensitive tests (<20 RNA copies/ml) (Roche Diagnostic Systems, Inc, New Jersey, USA) and $CD4^+$ T cell counts (cells/µl of blood) were determined using FACSCount System from Becton Dickinson (San Jose, CA).

Informed consent was obtained from all individuals participating in this study, and the study was approved by the Human Research on Human Subjects Committee of the University of the Witwatersrand, Johannesburg.

Table 2.1: Characteristics of HIV-1 infected individuals from the black South African population including HICs, HIC subgroups and progressors.

Study groups	N	Age (Years) Mean and range	Gender (% Female)	$CD4^+$ T cell count (Cells/µl) Median and IQR[3]	Viral load RNA copies/ml	Time from HIV-1 diagnosis (years) Median and IQR
HICs	52	37 (19 – 54)	86.5	697 (604 – 879)	1123 (127 – 2076)	8 (2 – 11)
ECs	11	44 (19 – 54)	81.8	853 (718 – 1022)	<40	11 (9 – 12)
VCs	30	36 (19 – 46)	90.0	651 (555 – 832)[1]	495 (327 – 965)[2]	3 (0 – 10.35)
HVL LTNPs	11	41 (31 – 51)	81.8	663 (635 – 749)	54 375 (13 415- 77 820)	8 (8 – 110)
<400 HICs	20	39 (19 – 54)	85.0	693 (592 – 907)	138 (<20 – 155)	10 (2 – 120
>400 HICs	32	36 (19 – 51)	87.5	707 (604 – 8650	1610 (917 – 15 709)	7 (2.25 – 9.75)
Progressors	74	43 (28 – 71)	83.8	177 (146 – 210)	38 444 (19 853 – 103 042)	6 (1 – 70

[1]$CD4^+$ T cells counts <500 cell/µl at enrolment for three of the individuals, but were included based on the length of infection without ARVs from the time of diagnosis to enrolment (≥ 9 years).

[2]One individual had VL >2000 but included in VCs due to length of infection from diagnosis without ARV treatment (9 years).

[3]Interquartile range

2.2 DNA extraction

Stored genomic DNA samples previously extracted from whole blood using the QIAamp DNA Blood Mini Kit (QIAGEN, Dusseldorf, Germany), were used for the determination of *CCL3* haplotypes/genotypes. For the determination of *CCL3L* and *CCL4L* gene copy numbers, genomic DNA was extracted either from frozen buffy coats or peripheral blood mononuclear cells (PBMCs) using the QIAamp DNA Blood Mini Kit (QIAGEN, Dusseldorf, Germany), and eluted in nuclease free water. The DNA was quantified using the Nano Drop 2000 (Thermo Fisher Scientific, Massachusetts, USA). For 16 of the samples, where the concentration of the DNA was lower than 15 ng/µl, the DNA was concentrated using a freeze dying method. This involves freezing the DNA and vacuum extraction of the water leaving the DNA behind, which is then reconstituted in an appropriate volume of nuclease-free water to make up final concentrations of 20 ng/µl.

2.3 *CCL3* and *CCL4* nomenclature

Different nomenclature has been used in various studies to refer to the CCL3 and CCL4 encoding genes (Figure 2.1), often making it difficult to compare data across studies. In this study, *CCL3* (Gene ID: 6348) refers to the allelic gene present at 2 copies per diploid genome (pdg). *CCL3L* is the total sum of the non-allelic variants *CCL3L1* (Gene ID: 6349) and *CCL3L2* (Gene ID: 390788) and *CCL3L3* (Gene ID: 414062) which are found in variable copy numbers. *CCL3L1* and *CCL3L3* have coding sequences which are identical and in this study, they are referred to as *CCL3La*, while *CCL3L2* which was previously thought to be a truncated pseudogene is referred to as *CCL3Lb*.

Similar to *CCL3*, *CCL4* (Gene ID: 6351) refers to the allelic gene present at 2 copies pdg, while *CCL4L* is the sum of the *CCL4L1* (Gene ID: 388372) and *CCL4L2* (Gene ID: 9560) genes. Some studies refer to the *CCL4L* gene immediately downstream of the *CCL4* as *CCL4L1* and the gene downstream of *CCL4L1* as *CCL4L2* (Shao *et al*., 2007). While other studies, using nomenclature which is in concordance with the annotation in the GenBank, refer to the gene immediately downstream of the *CCL4* gene as the *CCL4L2* gene and the gene downstream to *CCL4L2* as the *CCL4L1*, which encodes a functional CCL4 isoform (Colobran et al., 2010). Previous studies have also referred to the *CCL4L1* and *CCL4L2* genes as *CCL4Lb* and *CCL4La*, respectively (Shostakovich-Koretskaya et al., 2009). *CCL4L1* and *CCL4L2*, in this study will be referred to as *CCL4La* and *CCL4Lb*, respectively. As illustrated

in Figure 2.1, the gene immediately downstream of *CCL4* is *CCL4Lb* and the gene downstream of *CCL4Lb* is *CCL4La*.

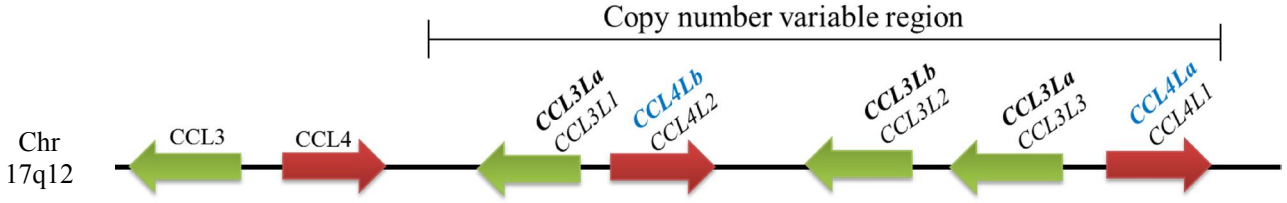


Figure 2.1: Schematic representation of the CCL3 and CCL4 encoding genes showing relative positions of the genes on chromosome 17, nomenclature of the genes and the orientation of the genes indicated by the direction of the arrows.

2.4 Copy number determination using droplet digital PCR

A previous study demonstrated that droplet digital PCR (ddPCR) was a much more accurate method for determining high *CCL4L* copy numbers compared to real-time PCR (Bharuthram et al., 2014), hence we decided to use ddPCR for copy number determination in this study. Droplet digital PCR provides absolute quantification of copy numbers without the use of standard curves. The system uses the water oil emulsion droplet technology which allows the separation of a sample into up to 20 000 droplets, and every droplet serves as a single reaction, thereby allowing thousands of reactions to be carried out for one sample. Figure 2.2 shows the basic workflow of ddPCR from DNA extraction to data analysis.

2.4.1 Restriction enzyme digestion

For ddPCR, restriction digestion of genomic DNA with a restriction enzyme/s which do not restrict within the amplified fragments of the target genes is essential for separating tandem gene copies to ensure proper partitioning into droplets and hence accurate copy number determination. Restriction enzyme analysis to identify restriction enzymes which do not restrict within the amplicons of the *CCL3L* genes was performed using SEQUENCHER (Gene Codes Corporation, Ann Arbor, MI, USA) on reference gene sequences of *CCL3La* (NG_023369.1), and *CCL3Lb* (NG_029353.2) downloaded from the GenBank sequence database. *HindIII* (Promega, Madison, Wisconsin, USA) was selected for the digestion of the *CCL3L* genes. For *CCL4L* genes, double digestion using *BamH1* and *Sac1* (Promega, Madison, Wisconsin, USA) was performed as described by Bharuthram et al. (2014).

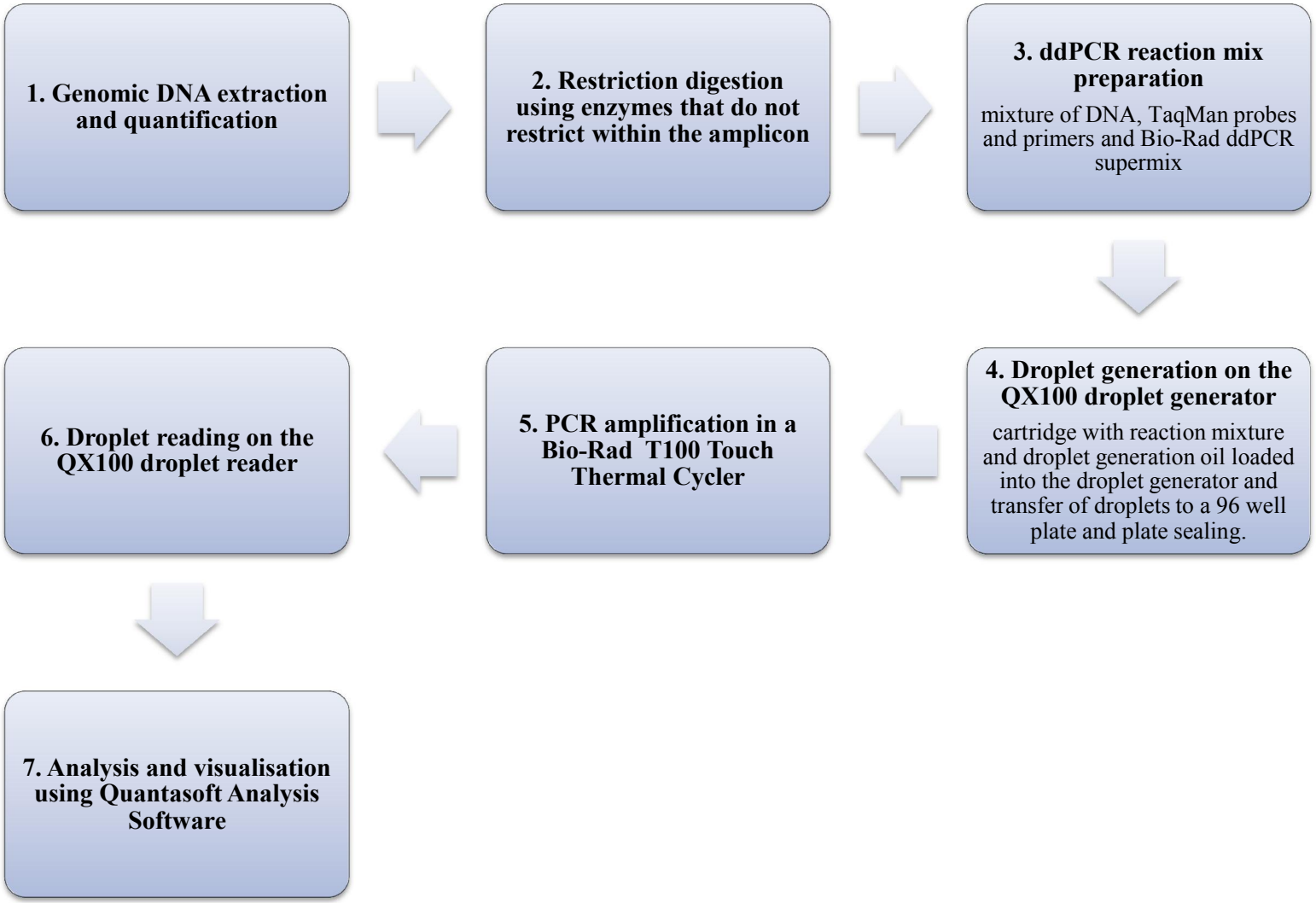


Figure 2.2: A flow diagram showing the basic workflow of ddPCR for copy number determination in this study.

DNA digestion was set up in total volumes of 60 µl consisting of 750ng of DNA, 6 µl of 1X buffer (Promega, Madison, Wisconsin, USA), 1 µl of restriction enzyme (10 U/µl) and nuclease-free water (volume dependent of the volume of template DNA). The reactions were incubated at 37°C for 2 hours followed by 65°C for 15 minutes (restriction enzyme heat inactivation) on the GeneAmp PCR System 9700 Thermocycler (Applied Biosystems, Foster City, CA, USA). The digested DNA was used directly in the ddPCR assays without undergoing a clean-up step to remove enzymes and buffer components.

2.4.2 Adaptation of a *CCL3L* real-time PCR assay to ddPCR format

In order to develop ddPCR assays for *CCL3L* copy number determination, TaqMan probes and primers used in the real-time PCR assay for the *CCL3L, CCL3La*, and *CCL3Lb* genes (Shostakovich-Koretskaya et al., 2009) were used to convert to a ddPCR assay platform. The *CCL3L*-specific primers and probe were as published by Townson et al (2002) and do not distinguish between *CCL3La* and *CCL3Lb*, while the *CCL3La*-specific and *CCL3Lb*-specific primers and probes were as described by Shostakovich-Koretskaya et al (2009). Minor grove binding (MGB) probes were used for *CCL3L* [fluorescent label: Carboxyfluorescein (FAM)], *CCL3La* (fluorescent label: FAM), *CCL3Lb* (fluorescent label: FAM) and the *betaglobin gene (BGB)* (fluorescent label: VIC) which was used an endogenous control.

The adaptation of the real-time PCR assays to the ddPCR platform was straight forward for the *CCL3L* and *CCL3Lb* genes, with a no further optimisation required. For the *CCL3La* gene however, there was insufficient separation of droplets that contained the targets (positive) and those that did not have any target (negative), with the negative droplets appearing to have high background fluorescence. We attempted different annealing temperatures to try and get a clear separation of the two droplet population however no noticeable separation was seen.

We consequently designed new sets of TaqMan probes and primers. For the design, the *CCL3La* (NG_023369.1) nucleotide sequence downloaded from GenBank was aligned to the *CCL3Lb* (NG_029353.2) sequence using SEQUENCER in order to identify the region with the least similarity between the two genes. Three sets of primers and probes (Table 2.1) were designed with the help of Primer express v.3 software (Thermo Fisher Scientific, Massachusetts, USA).

The three designed sets of primers and probes were tested on genomic DNA extracted form A431 cells which are known to have 2 copies of *CCL3L* and 2 copies of *CCL3La* pdg, with zero copies of *CCL3Lb* (Gonzalez *et al.* 2005; Townson *et al.* 2002).

Table 2.2: Primers and probes designed for development of *CCL3La* assay for ddPCR.

	Forward primers (5'-3')	Reverse primers (5'-3')	Probes (5'-3')
***CCL3La*/Set 1**	GGATTACAGGTGTGAGCGACCATG	AGGCTATTCTTAGTTGATCCCTTCTC	CTGGCTGCATAGACCAGTT
***CCL3La*/Set 2**	GAAACCCATGAAGCTAGAACCAG	CAAGAATAGCATCTCTGAGCTACTCTC	TAACTCTCAGCTCTCAACTC
***CCL3La*/Set 3**	AGGAAGCAGAATTACAGCTC	CCTATCCTTGTGGTTATAGCAGCTG	GGTCTATGCATGAACTCTCCAGCC

2.4.3 *CCL3L* and *CCL4L* copy number determination

CCL3L, *CCL3La* and *CCL3Lb* gene copy numbers were determined using the above developed ddPCR assays. Previously developed ddPCR assays (Bharuthram et al., 2014) with a few modifications, were used to determine the *CCL4L*, *CCL4La* and *CCL4Lb* gene copy numbers. We used a FAM fluorescent label for *CCL4Lb* instead of the VIC fluorescent label used by Bharuthram et al. (2014). As for *CCL3L* copy number determination assays, minor grove binding (MGB) probes were used for *CCL4L* (fluorescent label: FAM), *CCL4La* (fluorescent label: FAM), *CCL4Lb* and the *betaglobin gene (BGB)* (fluorescent label: VIC). *BGB* was used as an endogenous control for both *CCL3L* and *CCL4L* assays because, it is a housekeeping gene which is always found at 2 copies pdg. The assays were performed using the QX100 Droplet Digital PCR system (Bio-Rad, Pleasanton, CA, USA).

Reactions were carried out in triplicate for each of the *CCL3L* (*CCL3L*, *CCL3La*, and *CCL3Lb*) and *CCL4L* (*CCL4L*, *CCL4La* and *CCL4Lb*) genes, thus for each sample 18 reactions were conducted. Each reaction mixture also contained an endogenous control (*BGB*), making it a duplex reaction. The ddPCR reaction mixtures were set up in total volumes of 25 µl containing 2x ddPCR supermix for probes (no dUTP), 900 nM of each primer, 250 nM of each probe and 50 ng of the digested DNA sample. For droplet generation, 20 µl of the reaction mixtures were added into individual sample wells of the disposable Bio-Rad DG8 droplet generator cartridge with 70 µl of droplet generation oil added into the oil wells for each sample and loaded into the QX100 droplet generator. The droplet generator

applies vacuum to each of the wells to partition the sample into an emulsion (water-in-oil) of up to 20 000 droplets. Post droplet generation, 40 µl of the generated emulsified droplets were transferred to a 96 well plate (Eppendorf AG, Hamburg, Germany) and heat sealed with a pierceable foil using a Bio-Rad PX1 plate sealer. After heat sealing, the plate was subsequently placed on a Bio-Rad T100 Thermal Cycler for end-point amplification. Thermal cycling conditions were as follows: initial denaturation for 10 minutes at 95°C, followed by 40 cycles of 30 seconds denaturation at 94°C and annealing at 60°C for 60 seconds and final elongation for 10 minutes at 98°C. After PCR amplification, the plate was loaded into the QX100 Droplet Reader (Bio-Rad, Hercules, CA, USA). In the droplet reader, the droplets are passed through a two-colour optical detection system in order to identity and count droplets as either positive (target present) or negative (target absent) based on the fluorescence measurements. Bio-Rad consumables and reagents including DG8 cartridges, gaskets, droplet generation oil and droplet reader oil were used.

Quantasoft Analysis Software version 1.7.4.0917 (Bio-Rad, Hercules, CA, USA) was used for data analysis and visualization. The number of copies present in each sample was determined by calculating the ratio of the target gene concentration to the endogenous control (*BGB*) concentration, multiplied by two (number of *BGB* copies pdg). The basic ddPCR work flow for copy number determination is shown in Figure 2.2.

2.5 Determination of *CCL3* haplotypes

2.5.1 Haplotype description

Paximadis et al. (2013) identified two seven-SNP haplotypes, namely HapA1 and HapA3, in the *CCL3* gene to be associated with protection from *in utero* HIV-1 infection and increased intrapartum transmission, respectively. The SNPs and minor alleles for Hap-A1 and Hap-A3 relative to the *CCL3* gene are shown in Figure 2.3. Hap-A1 and Hap-A3 overlap at 2 SNP positions, A/G p + 1245 (rs1719130) and C/G p + 1728 (rs1063340) which are referred to as Hap-2SNP as can be seen on Figure 2.3.

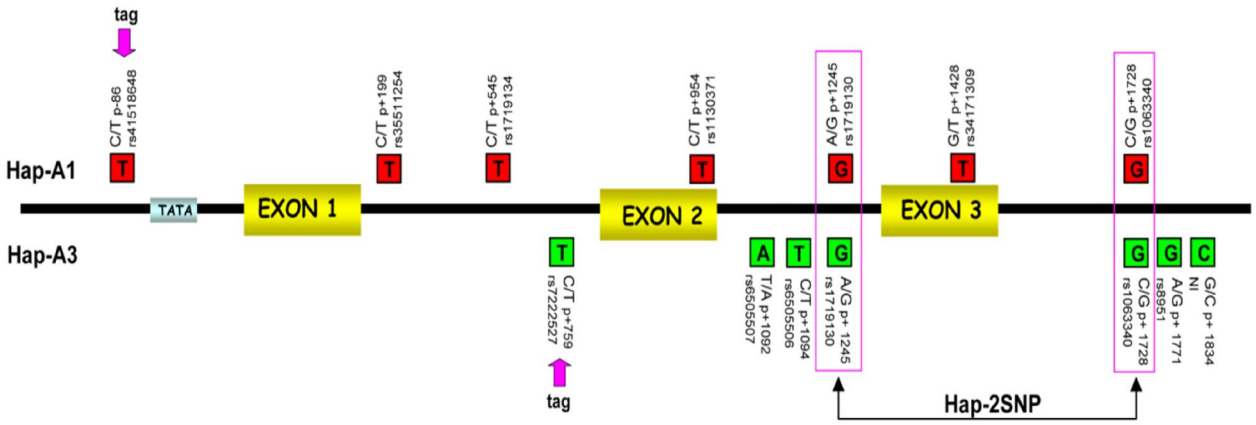


Figure 2.3: Schematic representation of Hap-A1, Hap-A3 and Hap-2SNP in the *CCL3* gene. Hap-A1 and Hap-A3 minor alleles are indicated in red and green boxes, respectively with SNP gene positions and dbSNP accession numbers given above or below the boxes. Pink rectangles indicate the two SNPs shared by Hap-A1 and Hap-A3, labeled Hap-2SNP. Hap-A1 and Hap-A3 tag SNPs are indicated with pink arrows labeled 'tag'. Image source: Paximadis et al. (2013).

2.5.2 Real-time PCR assays for Hap-A1 and Hap-A3 Genotyping

In order to investigate the role of the *CCL3* haplotypes in natural control of HIV-1 infection, tag SNPs for Hap-A1 and Hap-A3 were genotyped using previously developed SYBR green real-time C_T shift assays (Paximadis et al., 2013). C_T shift assays detect SNPs using allele-specific PCR with two primers, one specific to the wildtype (WT) and one specific to the mutant allele (MT), with one common reverse or forward primer with the orientation dependent on the orientation of the allele specific primers. The genotypes are determined from C_T shifts which result from the mispriming of the nonspecific primers.

For each sample, two PCR reactions were conducted, one with the primer to detect the WT allele and the other with the primer to detect the MT allele. Hap-A1 allele specific primers were forward primers with their 3' ends modified with locked nucleic acid (LNA). LNAs are nucleic acid analogs which increase specificity and template binding strength when incorporated into the primers. Standard reverse primers were used in Hap-A1 assay. The Hap-A3 allele specific primers were a combination of 3' end LNA modified WT allele reverse primer and a standard reverse MT allele primer with a common forward primer.

The reactions were set up in 96 well plates in total volumes of 10 µl containing 1 x SYBR green PCR master mix, 10 pmol of each of the reverse and forward primer and 10-20 ng DNA

template. The assays were performed in the Applied Biosystems 7500 Real-Time PCR. Amplification conditions were as follows: initial 10 minutes denaturation at 95°C, followed by 40 cycles of 15 seconds denaturation at 95°C, annealing at 60°C for 20 seconds and extension at 72° for 1 minute. Non template controls (NTC) were also included.

Hap-2SNP genotypes were determined by counting individuals heterozygous for either Hap-A1 or Hap-A3 as heterozygous for Hap-2SNP, and individuals heterozygous for both Hap-A1 and Hap-A3 as homozygous for Hap-2SNP. Individuals with homozygosity for either Hap-A1 or Hap-A3 were also considered homozygous for Hap-2SNP.

2.6 Data analysis

In order to determine if *CCL3L* and *CCL4L* copy numbers and/or *CCL3* haplotypes were associated with natural control of HIV-1, we stratified and analysed the data in a number of ways:

1. *CCL3L* and *CCL4L* gene copy numbers as continuous variables were compared between HICs and progressors and HIC subgroups and progressors. In addition, copy numbers were correlated to markers of disease progression (VL and $CD4^+$ T cell count).
2. Gene copy numbers of *CCL3L* and *CCL4L* were stratified around the population specific median and compared between HICs, HIC subgroups and progressors in two ways:
    a) According to previous studies (Gonzalez et al., 2005, Shostakovich-Koretskaya et al., 2009) as copy number greater or equal to the population median (≥ population median) and less than the population median (< population median).
    b) According to Liu et al. (2010), as copy number less than or equal to the population median (≤ population median) and greater than the population median (> population median).

    CCL3L and *CCL4L* copy numbers stratified according to > population median and ≥ median were also compared to VL and $CD4^+$ T cell counts.
3. The combined effect of *CCL3L* and *CCL4L* copy numbers stratified by copy numbers ≤ population median and > population median was assessed by comparing HICs, HIC subgroups and progressors as having either high *CCL3L* or *CCL4L* copy numbers (i.e one or the other) or high *CCL3L* and high *CCL4L* (i.e having both).

4. Profiles of *CCL3L* and *CCL4L* copy numbers (assigned patterns IDs) and combination of both profiles were compared between HICs, HIC subgroups and progressors.
5. The *CCL3La* and *CCL3Lb* copy number differences (i.e *CCL3La – CCL3Lb*) were compared between HICs, HIC subgroups and progressors.
6. The number of individuals with *CCL4La* > *CCL4Lb*, *CCL4La* = *CCL4Lb* and *CCL4La* > *CCL4Lb* were compared between HICs, HIC subgroups and progressors.
7. *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2-SNP) representations, individually and each in combination with *CCL3La* copy numbers were compared between HICs, HIC subgroups and progressors.

All statistical analyses were performed using GraphPad Prism version 5.0 for Windows (Graphpad Software, San Diego, CA, USA). The statistical significance level was set at $P < 0.05$ and odds ratios with 95% confidence intervals. P values greater than 0.05 but less than 0.1 ($0.05 < p < 0.01$) were regarded as a trend of significance. Mann-Whitney non parametric U tests were used to compare continuous variables, with the Fisher's exact test being used to compare categorical data. Due to the exploratory nature of this study, corrections for multiple testing were not performed. Correlations were performed using Spearman's rank correlation. Haploview software (Broad Institute of MIT and Harvard) was used to determine the linkage disequilibrium and in addition, Haploview also measured deviation from the Hardy-Weinberg equilibrium using the exact method described by Wigginton et al. (2005).

CHAPTER 3: Results

3.1 The role of *CCL3L* and *CCL4L* copy numbers in HIV-1 control

In order to determine if *CCL3L* and *CCL4L* gene copy numbers associate with natural control of HIV-1, we determined the *CCL3L* (*CCL3L*, *CCL3La*, *CCL3Lb)* and *CCL4L* (*CCL4L*, *CCL4La*, *CCL4Lb*) gene copy numbers for 126 HIV-1 infected black South Africans, which included 52 HIV-1 controllers (HICs) and 74 progressors and compared these between HICs, the three HIC subgroups (ECs, VCs and HVL LTNPs) and progressors.

3.1.1 *CCL3L* assay design

Droplet digital PCR assays for *CCL3L* gene copy number determination were adapted from real-time PCR assays as outlined in Section 2.4.2. Figure 3.1A shows 1 dimensional (1D) amplitude plots of the *CCL3L*, *CCL3La* and *CCL3Lb* gene copy number determination assays. As can be seen, while the *CCL3L* and *CCL3Lb* assays showed sufficient separation of negative and positive droplets (Figure 3.1A i and iii), the separation of positive and negative droplets was insufficient for the *CCL3La* assay (Figure 3.1A ii). Consequently, we designed three new sets of TaqMan probes and primers (sets 1 – 3) for the *CCL3La* assay and the 1D amplitude plots of the *CCL3La* assay with the three sets of probes and primers are shown in Figure 3.1B. Although both set 1 and set 2 (Figure 3.1B i and ii), resulted in sufficient separation of positive and negative droplets, the difference in the fluorescence of the negative and positive droplets was greater for set 2, and hence was selected for use in our *CCL3La* assay. Set 3 of the probes and primers failed to result in the detection of any positive droplets (Figure 3.1B iii).

3.1.2 Assay validation

Our copy number determination assays were designed to determine the total *CCL3L* and total *CCL4L* gene copy numbers and to determine the copy numbers for the individual *CCL3L* (i.e *CCL3La* and *CCL3Lb*) and *CCL4L* (i.e *CCL4La* and *CCL4Lb*) genes. Thus, correlating the sum of *CCL3La* and *CCL3Lb* in addition to the sum of *CCL4La* and *CCL4Lb* copy numbers to the total copy numbers of *CCL3L* or *CCL4L,* respectively served as good internal validations of the assays. The results of the correlations are shown in Figure 3.2. Unrounded copy numbers were used in the correlations, and both the *CCL3L* and *CCL4L* assays performed exceptionally well each with r values of 0.99 ($p<0.0001$).

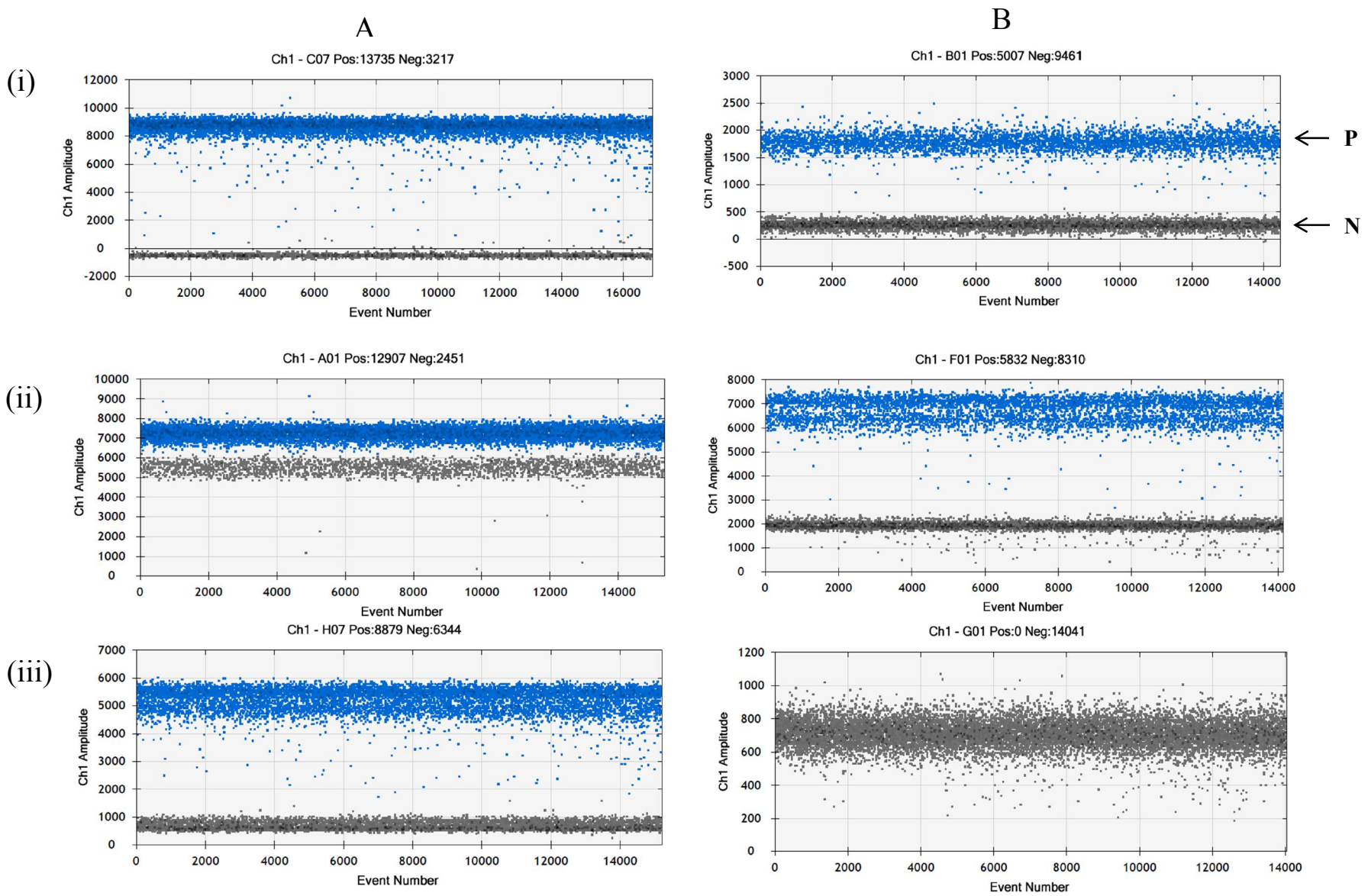


Figure 3.1: One dimensional (1D) amplitude plots of the *CCL3L* assays,(A) showing the 1D plots for *CCL3L* (i), *CCL3La* (ii) and *CCL3Lb* (iii) using TaqMan probes and primers adapted from *CCL3L* real-time PCR assays (Townson et al., 2002, Shostakovich-Koretskaya et al., 2009) and (B) showing the 1D plots for the *CCL3La* assay when using the newly designed sets of probes and primers (i) set 1, (ii) set 2 and (iii) set 3. Blue dots represents positive (P) droplets, grey dots are negative (N) droplets as indicated in B(i).

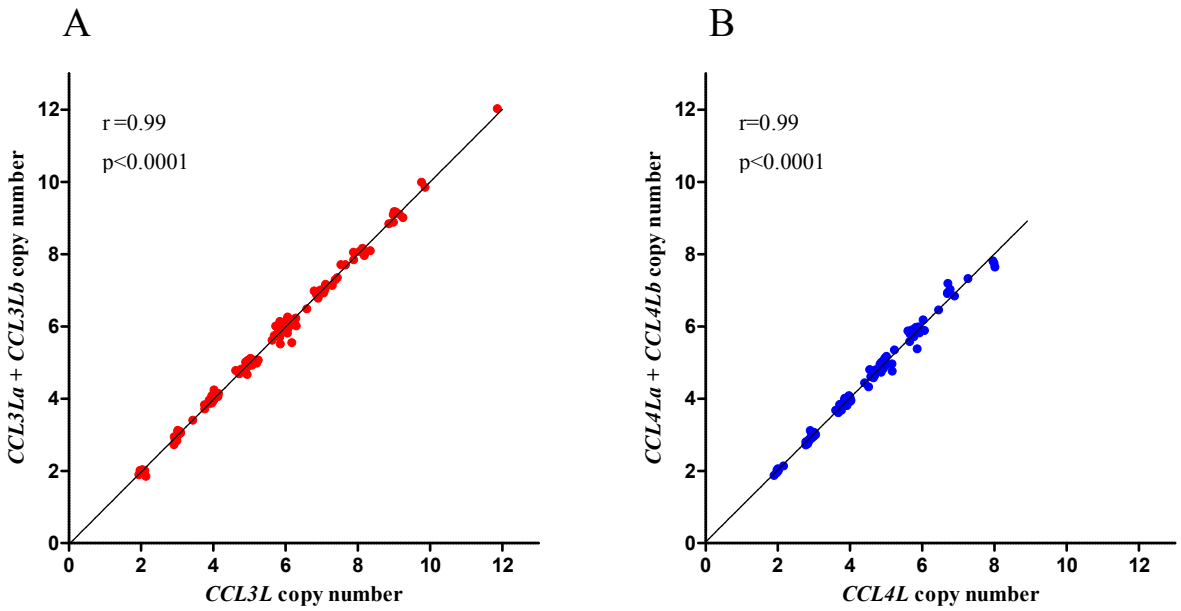


Figure 3.2: Correlation (Spearman's rank correlation) between the (A) total *CCL3L* gene copy numbers and the sum of *CCL3La* and *CCL3Lb* and (B) the total *CCL4L* gene copy numbers and the sum of *CCL4La* and *CCL4Lb*. Raw data (unrounded values) were used in the correlations**.**

3.1.3 *CCL3L and CCL4L* copy number distributions in study population

The *CCL3L* and *CCL4L* gene copy number medians and ranges in our study population are listed in Table 3.1. Copy number medians for *CCL3L* and *CCL4L* for HIV-1-uninfected healthy controls (HCs) from the black South African population were sourced from Picton et al. (2013) and Bharuthram (2015), generated using real-time PCR and ddPCR, respectively.

HIV-1 controllers (HICs) had a median *CCL3L* copy number of 6, and the median copy numbers for *CCL3La* and *CCL3Lb* were 5 and 1, respectively. The median *CCL3L*, *CCL3La* and *CCL3Lb* copy numbers in progressors were 5, 4 and 1, respectively, comparable to the medians for healthy controls. Thus, for HICs, only the *CCL3Lb* copy number median was comparable to that of progressors and healthy controls, while the *CCL3L* and *CCL3La* median copy numbers were higher than those in progressors and healthy controls.

The median *CCL4L*, *CCL4La* and *CCL4Lb* copy numbers were 5, 3 and 2 in HICs and 4, 2 and 2 in progressors, respectively. The *CCL4Lb* median was thus the same for HICs and progressors as well as healthy controls. While the *CCL4L* median was the same for progressors and healthy controls, it was higher in HICs. The *CCL4La* median was comparable between HICs and healthy controls and was lower in progressors.

Table 3.1: *CCL3L* (*CCL3La* and *CCL3Lb*) and *CCL4L* (*CCL4La* and *CCL4Lb*) gene copy number medians for the study population (HICs and progressors) as well as healthy controls (HCs).

	Median (Ranges)		
Gene	**HCs[1]**	**HICs (N=52)**	**Progressors (N=74)**
CCL3L	5 (2 – 11)	6 (2 – 12)	5 (2 – 10)
CCL3La	4 (2 – 9)	5 (2 – 8)	4 (2 – 8)
CCL3Lb	1 (0 – 4)	1 (0 – 4)	1 (0 – 4)
CCL4L	4 (3 – 6)	5 (2 – 8)	4 (2 – 8)
CCL4La	3 (0 – 4)	3 (0 – 5)	2 (0 – 6)
CCL4Lb	2 (0 – 4)	2 (0 – 5)	2 (0 – 6)

[1]Data for *CCL3L* (N=143) and *CCL4L* (N=55) in healthy controls was obtained from Picton et al. (2013) and Bharuthram (2015), generated using real-time PCR and ddPCR, respectively.

Figure 3.3 shows the representation of the *CCL3L* and *CCL4L* gene copy number distributions in HICs and progressors. The copy number distributions for *CCL3L*, *CCL3La* and *CCL3Lb* seemed comparable between HICs and progressors, however it was interesting to note that for both *CCL3La* and *CCL4L* there was a distinct shift at copy number >4, where at frequencies ≤4, progressors had higher frequencies compared to HICs. However, at a copy number of 5, there was a clear shift with higher frequencies of this copy number in HICs compared to progressors. For *CCL4Lb*, a similar pattern was seen where at *CCL4Lb* copy number of 2, progressors had a noticeably higher frequency compared to HICs, but at a copy number of 3 there was a shift with HICs having a higher frequency than progressors.

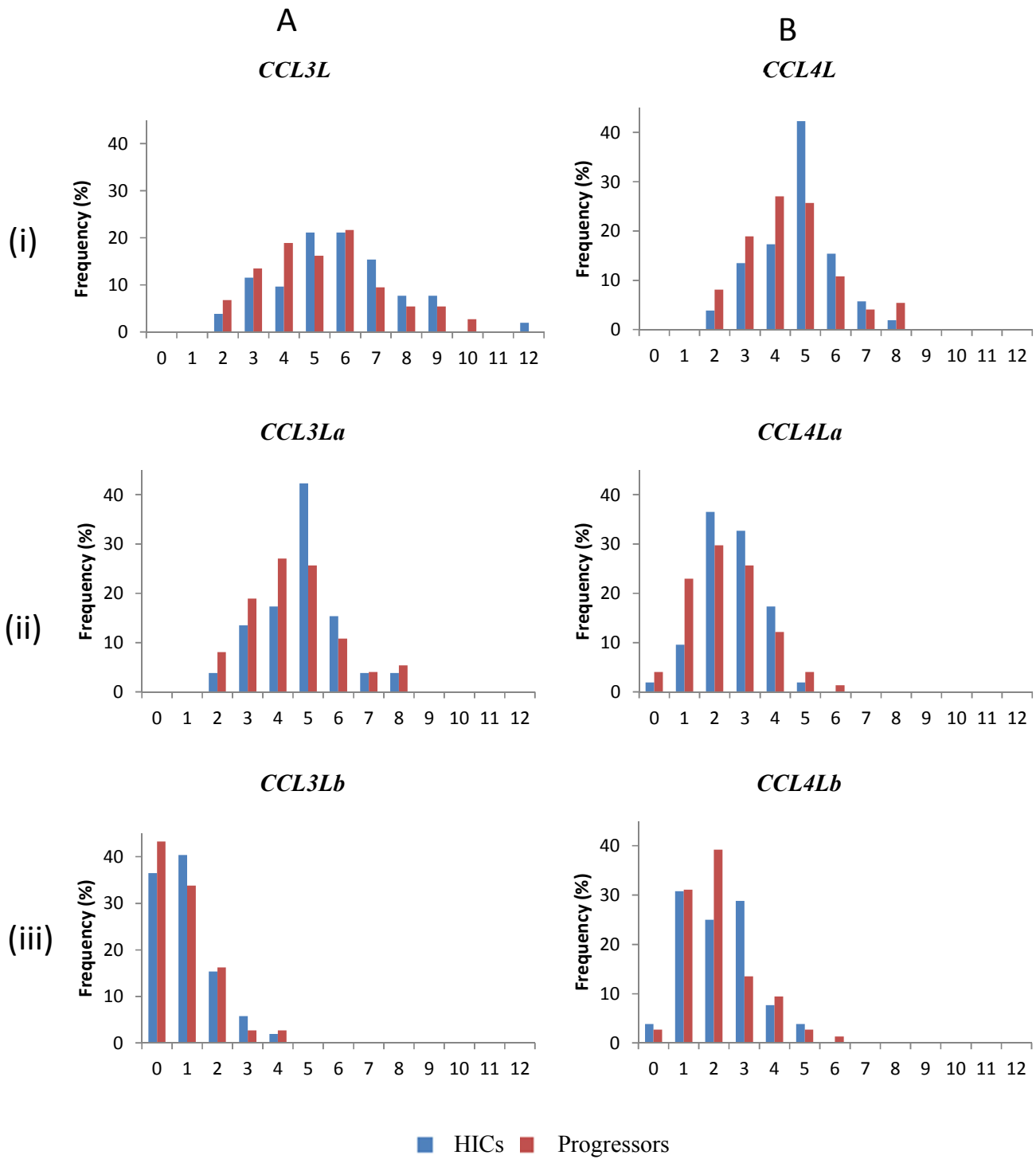


Figure 3.3: Bar graph showing representation of *CCL3L* and *CCL4L* gene copy number distributions in HICs (N=52) and progressors (N=74). (A) Distribution of the (i) *CCL3L*, (ii) *CCL3La* and (iii) *CCL3Lb* gene copy numbers. (B) Distribution of the (i) *CCL4L*, (ii) *CCL4La* and (iii) *CCL4Lb* gene copy numbers.

3.1.4 Relationship between *CCL3L* and *CCL4L* gene copy numbers

Figure 3.4 shows the correlations of *CCL3L* and *CCL4L* copy numbers relative to each other. Strong statistically significant positive correlations were seen between the total *CCL3L* copy numbers with *CCL3La* and *CCL3Lb* copy numbers (Figure 3.4Ai and Aii). *CCL3La* and *CCL3Lb* copy numbers also showed a significant positive correlation (Figure 3.4Aiii). Although the total *CCL4L* gene copy numbers showed significant positive correlations to both *CCL4La* and *CCL4Lb* gene copy numbers, a significant weak negative correlation was seen between *CCL4La* and *CCL4Lb* (Figure 3.4B), as previously reported (Shostakovich-Koretskaya et al., 2009, Bharuthram, 2015).

When we looked at the relationship between the *CCL3L* and *CCL4L* genes, we found that all three correlated significantly and positively, i.e *CCL3L* and *CCL4L* (Figure 3.4Ci), *CCL3La* and *CCL4La* (Figure 3.4Cii) and *CCL3Lb* weakly with *CCL4Lb* (Figure 3.4Ciii).

Interestingly, we found that the *CCL3La* copy number was identical to the *CCL4L* (*CCL4La* + *CCL4Lb*) copy number in both the HICs and progressors, with the exception of one individual who had 8 copies for *CCL3La* and 7 copies for *CCL4L*. Figure 3.5, shows the correlation of the *CCL3La* and *CCL4L* copy numbers. This relationship between *CCL3La* and *CCL4L* has been previously demonstrated in a Caucasian population and in an East African population (Walker et al., 2009, Nordang et al., 2012).

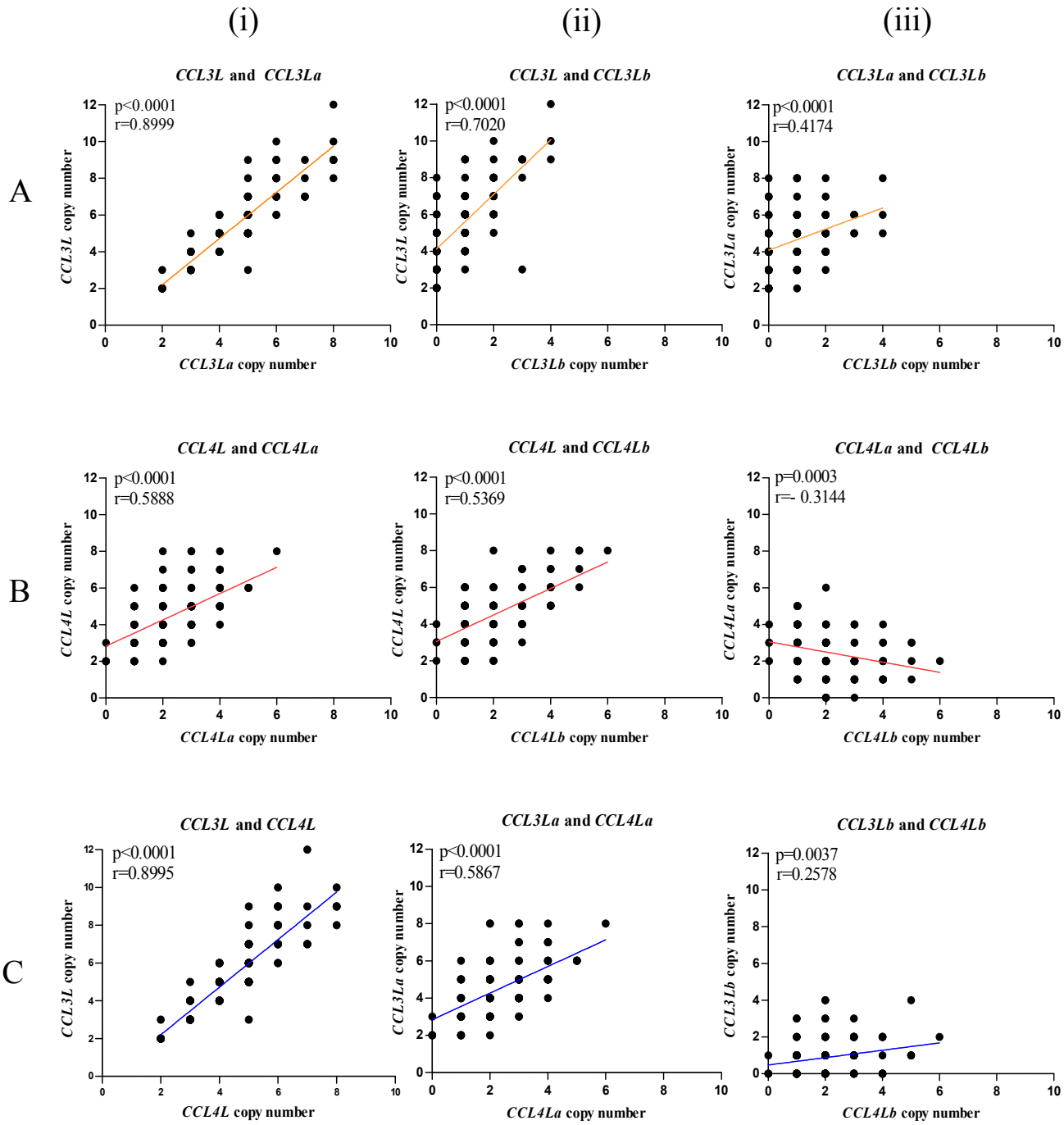


Figure 3.4: Correlations (Spearman's rank correlation) of *CCL3L* and *CCL4L* gene copy numbers. (A) Correlation of (i) *CCL3L* and *CCL3La*, (ii) *CCL3L* and *CCL3Lb* and (iii) *CCL3La* and *CCL3Lb* copy numbers showing significantly positive correlations. (B) Correlation of (i) *CCL4L* and *CCL4La*, (ii) *CCL4L* and *CCL4Lb* and (iii) *CCL4La* and *CCL4Lb* copy numbers showing significantly positive correlations between *CCL4L* and *CCL4La* and *CCL4L* and *CCL4Lb* with a negative correlation between *CCL4La* and *CCL4Lb*. (C) Correlation of (i) *CCL3L* and *CCL4L*, (ii) *CCL3La* and *CCL4La,* (iii) *CCL3Lb* and *CCL4Lb***.** For these correlations the rounded values were plotted.

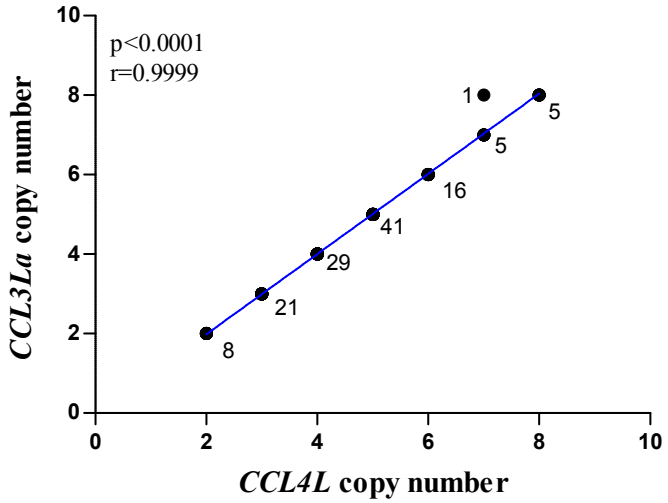


Figure 3.5: Correlation (Spearman's rank correlation) of the *CCL3La* copy numbers to the *CCL4L* copy numbers, showing that the *CCL3La* copy numbers are identical to *CCL4L* copy numbers for both HICs and progressors with the exception of one individual. The numbers above or next to the data points represent the number of individuals with the particular copy numbers. For this correlation the rounded values were plotted.

3.1.5 Comparison of *CCL3L* and *CCL4L* copy numbers between HICs and progressors

To determine if the distributions of the *CCL3L* and *CCL4L* gene copy numbers were significantly different between HICs, HIC subgroups (VCs, ECs, HVL LTNPs, <400 HICs and >400 HICs) and progressors, we compared copy numbers between the respective groups. No significant differences were seen between HICs and progressors for any of the *CCL3L* or *CCL4L* genes.

Figure 3.6A and B shows the comparisons between the HICs and progressors for *CCL3L* and *CCL4L* genes, respectively. No significant differences were also seen when comparing HIC subgroups and progressors for any of the *CCL3L* or *CCL4L* genes, however, strong trends (p=0.07 for both) of higher *CCL3La* and *CCL4L* copy numbers in the VCs were found when compared to progressors (Figure 3.7). As mentioned, *CCL3La* and *CCL4L* copy numbers were identical with the exception of one individual. Therefore, for all subsequent analyses the associations will be shown as representing either gene (i.e as *CCL3La*/*CCL4L*).

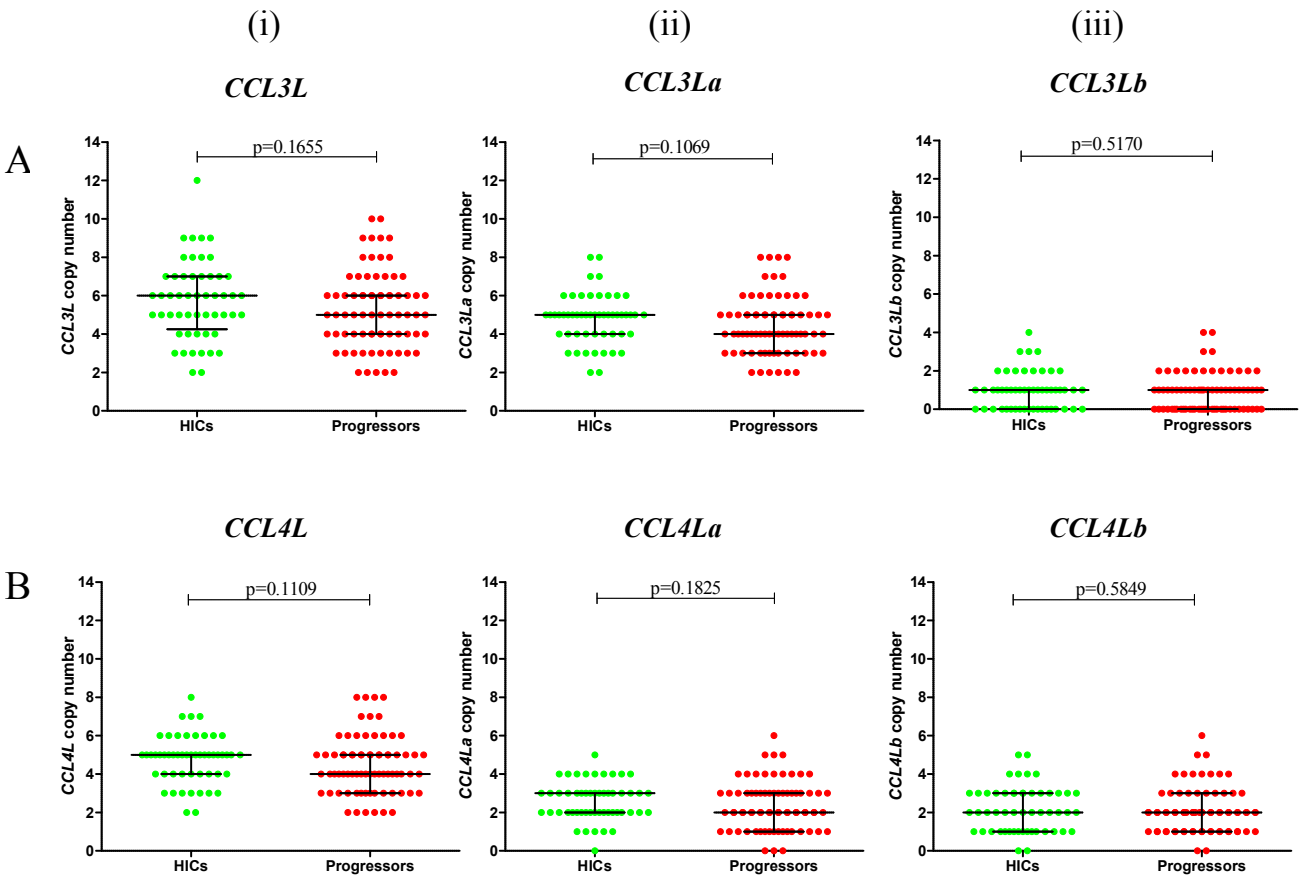


Figure 3.6: Comparisons (Mann–Whitney U test) of the distribution of the (A) *CCL3L*, *CCL3La* and *CCL3Lb* and (B) *CCL4L*, *CCL4La* and *CCL4Lb* copy numbers between HICs (N=52) and Progressors (N=74), showing no significant differences between the groups. Black horizontal lines represent the median (middle) and the interquartile ranges (outer lines).

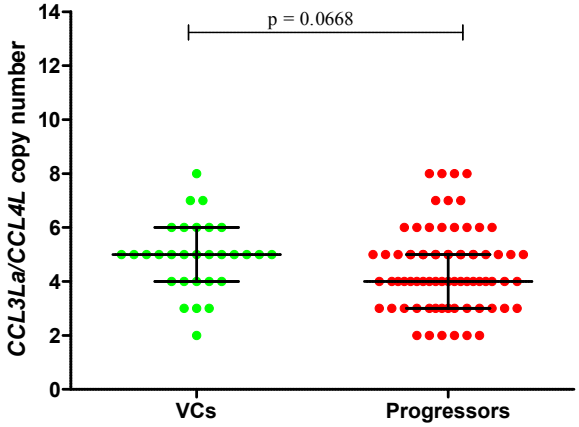


Figure 3.7: Comparisons (Mann–Whitney U test) of the *CCL3La/CCL4L* copy numbers between VCs (N=30) and progressors (N=74), showing strong trends of higher *CCL3La*/*CCL4L* copy numbers in VCs. Black horizontal lines represent the median (median) and the interquartile ranges (outer lines).

3.1.6 Comparisons of stratified *CCL3L* and *CCL4L* copy numbers between HICs and progressors

Previous studies found that when stratified around the population specific median, *CCL3L* copy numbers were associated with different HIV-1 related outcomes including susceptibility to HIV-1 infection, viral loads and rate of progression to AIDS (Gonzalez et al., 2005, Shostakovich-Koretskaya et al., 2009). We therefore also compared the copy numbers between the groups by stratifying the *CCL3L*, *CCL3La* and *CCL3Lb* and *CCL4L*, *CCL4La* and *CCL4Lb* copy numbers around their medians. For comparisons, the *CCL3L* and *CCL4L* copy number medians of the background population (healthy black South Africans) sourced from Picton et al. (2013) and Bharuthram et al. (2014), respectively were used (shown in Table 3.1). We used two stratification strategies, firstly we grouped and compared individuals according to copy number ≥ population median and < population median, and secondly we grouped individuals according to copy numbers ≤ population median and > population median. When comparing individuals with copy numbers ≥ population median and < population median, no significant differences were seen between HICs and progressors as well as HIC subgroups and progressors for any of the *CCL3L* or *CCL4L* genes. However, strong trends of higher representation of individuals with total *CCL3L* copy numbers greater or equal to the population median (≥5) were seen in VCs (p=0.07, OR=0.39, CI=0.14-1.06) and in HICs with viral load <400 RNA copies/ml (<400 HICs; p=0.06, OR=0.27, CI=0.07-1.02) compared to progressors. The distribution of individuals with *CCL3L* copy numbers ≥ population median and < population median is shown in Figure 3.8A.

When comparing individuals with copy numbers ≤ population median and > population median, a strong trend of overrepresentation of individuals with *CCL3La*/*CCL4L* copy numbers > population median was seen in HICs (p=0.05, OR=0.45, CI=0.22-0.94) compared to progressors. A similar relationship was seen when we compared VCs (p=0.03, OR=0.36, CI=0.15-0.9) to progressors, with VCs having a significantly higher representation of individuals with high *CCL3La/CCL4L* copy numbers (> population median) compared to progressors. The most significant result was seen when comparing <400 HICs (p=0.01, OR=0.21, CI=0.07-0.97) and progressors for *CCL3La/CCL4L* copy numbers, with the <400 HICs having more individuals with *CCL3La/CCL4L* copy numbers greater than the median (>4 copies). When comparing the *CCL3L* and *CCL3Lb* gene copy numbers in the same manner, no significant differences were observed between HICs and progressors and HICs subgroups and progressors. The distributions of individuals with copy numbers ≤ population

median and > population median are shown in Figure 3.8B. No significant differences were also seen when comparing *CCL4La* and *CCL4Lb* copy numbers in a similar way.

3.1.7 Analysis of the combined effect of *CCL3L* and *CCL4L* copy numbers

Next we tried to determine if the *CCL3L* and *CCL4L* genes had an additive effect on HIV-1 control by investigating if having either a high *CCL3L* copy number (> population median) and/or high *CCL4L* copy number (> population median) or having both (high *CCL3L* and high *CCL4L*) was associated with HIV-1 control. When comparing HICs and progressors for having either a high *CCL3L* copy number (>6) and/or high *CCL4L* copy number (>5) or both (high *CCL3L* and high *CCL4L*), no significant differences were evident, but when HICs subgroups were compared to progressors, a significant difference was again seen between <400 HICs (p=0.04, OR=0.28, CI=0.08-0.91) and progressors, with the <400 HICs having a higher representation of individuals with high *CCL3L* and/or high *CCL4L* copy number.

Furthermore, when comparing the individual genes, a significant difference was seen when comparing HICs (p=0.01, OR=0.38, CI=0.18–0.80) and progressors with high *CCL3La* and/or high *CCL4La* copy numbers, with HICs having a higher representation of individuals with high *CCL3La* copy number and/or high *CCL4La* copy number (Figure 3.9). A significant difference was also seen in VCs (p=0.02, OR=0.31, CI=0.12–0.78) and <400 HICs (p=0.01, OR=0.21, CI=0.06–0.70) when compared to progressors in the same manner, with VCs and <400 HICs also having a higher representation of individuals with high *CCL3La* and/or high *CCL4La* copy numbers (Figure 3.9). No significant differences were seen when comparing individuals with both high *CCL3La* and high *CCL4La* copy numbers, and we also did not see any significant differences when HICs and progressors and HICs subgroups and progressors were compared for high *CCL3Lb* and/or high *CCL4Lb* copy numbers or both.

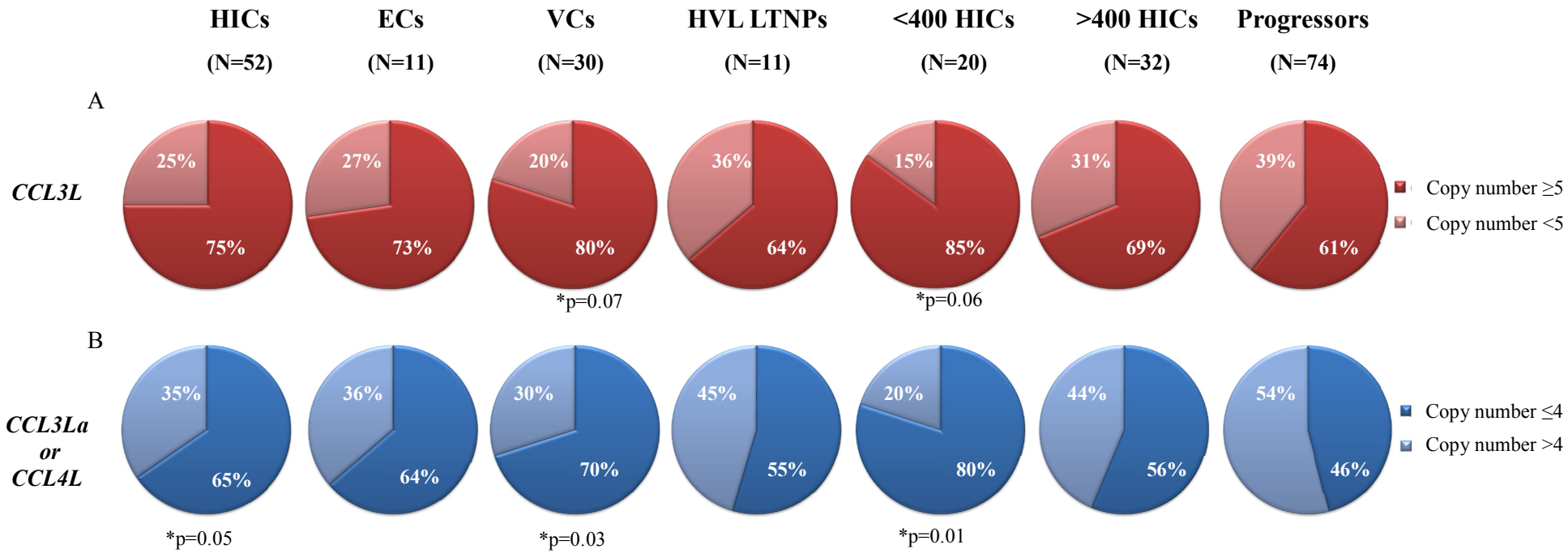


Figure 3.8: Copy number representations in HICs, HIC subgroups and progressors stratified around the population median. (A) *CCL3L* copy numbers stratified by ≥ population median and < population median, (B) *CCL3La/CCL4L* copy numbers stratified by ≤ population median and > population median. *Indicates a significant difference ($p<0.05$) or a trend ($0.05>p<0.1$) (as shown in the Figure) between respective groups and progressors. Fisher's exact test was used for statistical analysis.

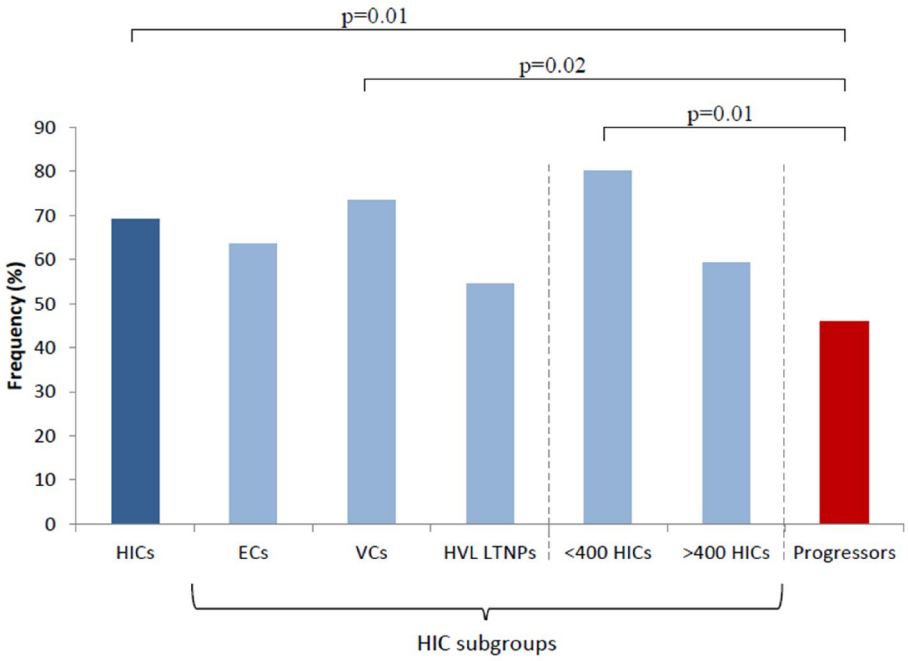


Figure 3.9: A bar graph showing the representation of individuals with high *CCL3La* and/or high *CCL4La* copy numbers in HICs, HIC subgroups and progressors. (HICs, N=52, ECs, N=11; VCs, N=30; HVL LTNPs, N=11, <400 HICs, N=20; >400 HICs, N=32, progressors, N=74). Significant differences are shown ($p<0.05$). Fisher's exact test was used for statistical analysis.

3.2 *CCL3L* and *CCL4L* copy number pattern representation and HIV-1 control

CCL3L and the *CCL4L* gene copy numbers vary in different individuals and represent unique patterns of these gene combinations which may have different functional implications. These combinatory profiles need to be considered, given that effects of some genes may be additive while others (eg. *CCL3Lb*) may be counteractive/subtractive. We therefore described the patterns of *CCL3L* and *CCL4L* gene copy number possession, separately and in combination. We assigned a gene copy number pattern identifier (gPID) for each individual based on how many copies of each gene of *CCL3L* or *CCL4L* the individual has, for example a *CCL3L* gPID of 321 for instance, represents 3 copies of total *CCL3L*, 2 copies of *CCL3La* and 1 copy of *CCL3Lb*, respectively. For *CCL4L* patterns, a gPID of 321 will also represent 3, 2 and 1 copies of total *CCL4L*, *CCL4La* and *CCL4Lb* respectively. Consequently, for the combined *CCL3L* and *CCL4L* patterns, a gPID of 321202 for example represents 3 copies of *CCL3L*, 2 copies of *CCL3La*, 1 copy of *CCL3Lb*, 2 copies of *CCL4L*, 0 copies of *CCL4La* and 2 copies

of *CCL4Lb*, respectively. This approach allows for the interrogation of which combinatory profiles might associate with HIV-1 control.

3.2.1 *CCL3L* gene copy number pattern representation

We found a total of 25 different *CCL3L* gPIDs in HICs and progressors combined, with HICs having 16 different gPIDs of *CCL3L* and the 651 gPID being the most prevalent (21.2%). Figure 3.10 shows the schematic representation of the *CCL3L* gPIDs with their frequencies in HICs, HIC subgroups and progressors. In progressors, 23 gPIDs were seen with two frequently occurring gPIDs, namely 651 and 440 (both 12.2%).

We compared the frequencies of each of the 25 gPIDs between HICs, HIC subgroups and progressors. No *CCL3L* gPIDs were significantly over- or underrepresented when HICs were compared to progressors, but a strong trend of underrepresentation of the *CCL3L* gPID 642 in HICs (p=0.08, OR=8.31, CI=0.45–153.7) was seen when compared to progressors. In fact, this combination was absent in HICs but present in 6.8% of progressors. An overrepresentation of the *CCL3L* gPID 651 was also seen in ECs (p=0.02, OR=0.17, CI=0.04–0.66) and <400 HICs (p=0.04, OR=0.26, CI=0.08–0.82) when compared to progressors. Interestingly, these two gPIDs that show "opposing" effects, i.e 651 and 642, only differ by one *CCL3La* copy and one *CCL3Lb* copy.

3.2.2 *CCL4L* gene copy number pattern representation

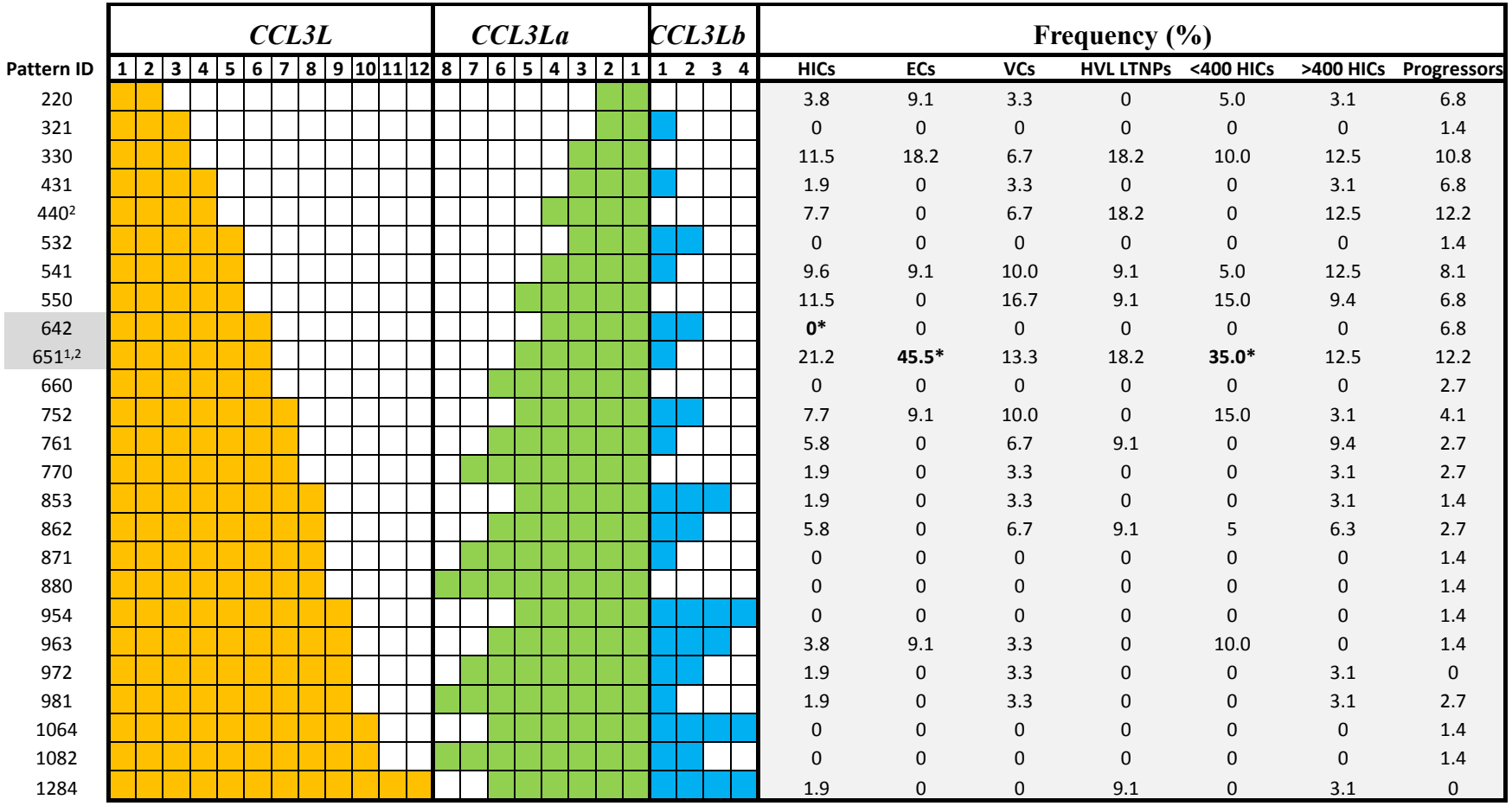
Analysis of the *CCL4L* gPIDs in the same manner, resulted in detection of 24 different *CCL4L* gPIDs in the total study group (HICs and progressors), with 19 different gPIDs in HICs and 23 different gPIDs in progressors. A schematic representation of the *CCL4L* gPIDs and their frequencies are shown in Figure 3.11. In HICs, the *CCL4L* gPID 523, was the most prevalent (15.3%), while in progressors the gPIDs, 422 and 523 were the most prevalent both with frequencies of 10.8%. When comparing the frequencies of *CCL4L* gPIDs in HICs, HIC subgroups and progressors, we found that *CCL4L* gPID 633 was overrepresented in HICs (p=0.03, OR=0.07, CI=0.004–1.38) when compared to progressors. A significant overrepresentation of the *CCL4L* gPID, 633 was also seen in >400 HICs (p=0.03, OR=0.06, CI=0.003–1.130) when compared to progressors. In VCs, a trend of overrepresentation of the *CCL4L* gPID 633 was also seen (p=0.08, OR=0.08, CI=0.004–1.64) compared to progressors.

Additionally, a trend of higher representation of a *CCL4L* gPID 523 was also seen in <400 HICs (p=0.06, OR=0.23, CI=0.05–1.01) when compared to progressors. Interestingly, there were no *CCL4L* gPIDs that were prevalent (>5%) in progressors and completely absent in HICs as was seen with the *CCL3L* 642 gPID.

3.2.3 Combined *CCL3L* and *CCL4L* gene copy number representation

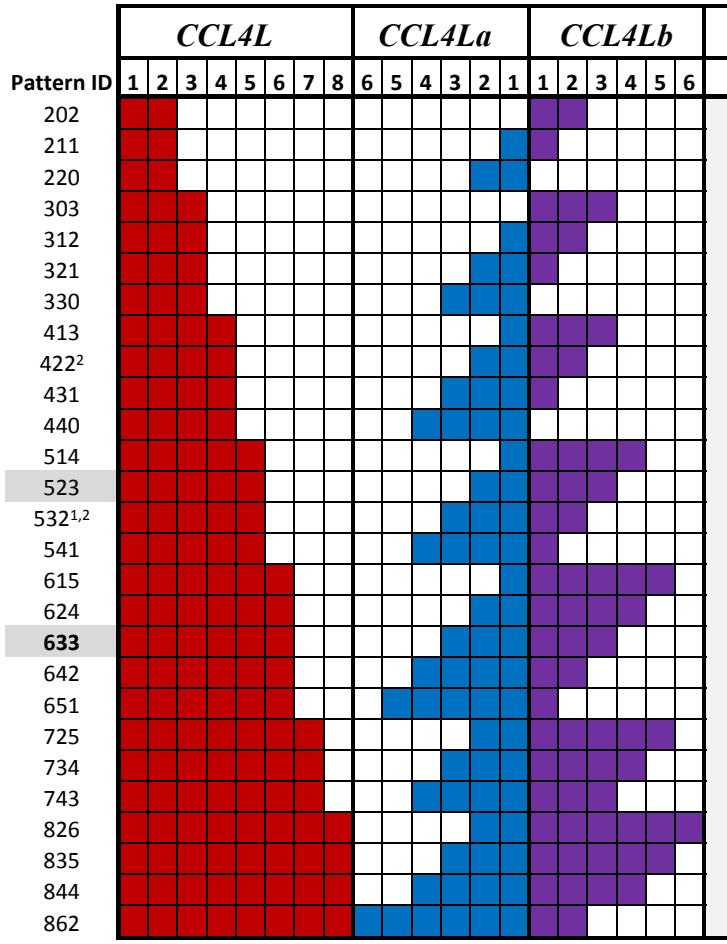
The combined gPIDs of *CCL3L* and *CCL4L* were next analysed. We found a total of 55 different gPIDs in both HICs and progressors, 33 of these gPIDs in HICs and 41 gPIDs in progressors. The schematic representation of the combined *CCL3L* and *CCL4L* gPIDs is shown in Figure 3.12. As expected, all the combined gPIDs were found in relatively low frequencies, and thus only subgroups where significant differences were noted have been listed. In HICs, the 651523 gPID was the most prevalent (9.6%) and in progressors, 5 most prevalent gPIDs were seen, each with a frequency of 5.4%.

When comparing the frequencies of the combined gPIDs between HICs and progressors, no significant differences were seen. However, a significant difference was found when VCs (p=0.02, OR=0.05, CI=0.003-1.05) were compared to progressors with an overrepresentation of individuals with the 550541 gPID in VCs. The 550541 gPID also showed a strong trend of overrepresentation in HICs (p=0.07, OR=0.09, CI=0.005–1.88) when compared to progressors. In addition, a trend of overrepresentation of the 761633 gPIDs was seen in the >400 HICs (p=0.09, OR=0,08, CI=0.004-1.767) compared to progressors. With the exception of the *CCL4L* gPIDs 633, which in combination with the *CCL3L* gPIDs 761 shows a trend of overrepresentation in >400. Significant differences or trends seen when *CCL3L* or *CCL4L* patterns were compared individually between the groups, were lost when tested in combination.



Pattern ID	CCL3L 1	2	3	4	5	6	7	8	9	10	11	12	CCL3La 8	7	6	5	4	3	2	1	CCL3Lb 1	2	3	4	HICs	ECs	VCs	HVL LTNPs	<400 HICs	>400 HICs	Progressors
220																									3.8	9.1	3.3	0	5.0	3.1	6.8
321																									0	0	0	0	0	0	1.4
330																									11.5	18.2	6.7	18.2	10.0	12.5	10.8
431																									1.9	0	3.3	0	0	3.1	6.8
440[2]																									7.7	0	6.7	18.2	0	12.5	12.2
532																									0	0	0	0	0	0	1.4
541																									9.6	9.1	10.0	9.1	5.0	12.5	8.1
550																									11.5	0	16.7	9.1	15.0	9.4	6.8
642																									**0***	0	0	0	0	0	6.8
651[1,2]																									21.2	**45.5***	13.3	18.2	**35.0***	12.5	12.2
660																									0	0	0	0	0	0	2.7
752																									7.7	9.1	10.0	0	15.0	3.1	4.1
761																									5.8	0	6.7	9.1	0	9.4	2.7
770																									1.9	0	3.3	0	0	3.1	2.7
853																									1.9	0	3.3	0	0	3.1	1.4
862																									5.8	0	6.7	9.1	5	6.3	2.7
871																									0	0	0	0	0	0	1.4
880																									0	0	0	0	0	0	1.4
954																									0	0	0	0	0	0	1.4
963																									3.8	9.1	3.3	0	10.0	0	1.4
972																									1.9	0	3.3	0	0	3.1	0
981																									1.9	0	3.3	0	0	3.1	2.7
1064																									0	0	0	0	0	0	1.4
1082																									0	0	0	0	0	0	1.4
1284																									1.9	0	0	9.1	0	3.1	0

Figure 3.10: Schematic representation of *CCL3L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies. Each coloured block represents a single copy. The pattern IDs which are highlighted in grey are those showing trends or significant differences. (ECs, N=11; VCs, N=30; HVL LTNPs, N=11; <400 HICs, N=20; >400 HICs, N=32).* shows where there were significant differences or trends when specific controller subgroups were compared to progressors. Fishers's exact test was used for statistical analysis. [1]Most frequent pattern in HICs, [2]Most frequent patterns in and progressors.



Pattern ID	*CCL4L* 1	2	3	4	5	6	7	8	*CCL4La* 6	5	4	3	2	1	*CCL4Lb* 1	2	3	4	5	6	Frequency (%) HICs	ECs	VCs	HVL LTNPs	<400 HICs	>400 HICs	Progressors
202																					0	0	0	0	0	0	4.1
211																					3.8	9.1	3.3	0	5	5	2.7
220																					0	0	0	0	0	0	1.4
303																					1.9	9.1	0	0	5	0	0
312																					1.9	9.1	0	0	5	0	8.1
321																					7.7	0	10.0	9.1	0	20	9.5
330																					1.9	0	0	9.1	0	5	1.4
413																					1.9	0	0	9.1	0	5	6.8
422[2]																					7.7	0	13.3	0	0	20	10.8
431																					5.8	9.1	0	18.2	5	10	9.5
440																					1.9	0	3.3	0	0	5	0
514																					1.9	9.1	0	0	5	0	4.1
523																					13.5	9.1	16.7	9.1	**20***	15	5.4
532[1,2]																					15.4	18.2	13.3	18.2	25	15	10.8
541																					11.5	18.2	13.3	0	15	15	5.4
615																					0	0	0	0	0	0	1.4
624																					5.8	0	6.7	9.1	5	10	1.4
633																					**7.7***	9.1	**6.7***	9.1	5	**15***	0
642																					0	0	0	0	0	0	4.1
651																					1.9	0	3.3	0	5	0	4.1
725																					1.9	0	0	9.1	0	5	0.0
734																					0	0	0	0	0	0	2.7
743																					3.8	0	6.7	0	0	10	1.4
826																					0	0	0	0	0	0	1.4
835																					1.9	0	3.3	0	0	5	1.4
844																					0	0	0	0	0	0	1.4
862																					0	0	0	0	0	0	1.4

Figure 3.11: Schematic representation of *CCL4L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies. Each coloured block represents a single gene copy. The pattern IDs which are highlighted in grey are those showing trends or significant differences. (ECs, N=11; VCs, N=30; HVL LTNPs, N=11, <400 HICs, N=20; >400 HICs, N=32).* shows where there were significant differences or trends when specific controller subgroups were compared to progressors. Fishers's exact test was used for statistical analysis. [1]Most frequent pattern in HICs, [2]Most frequent patterns in and progressors.

.

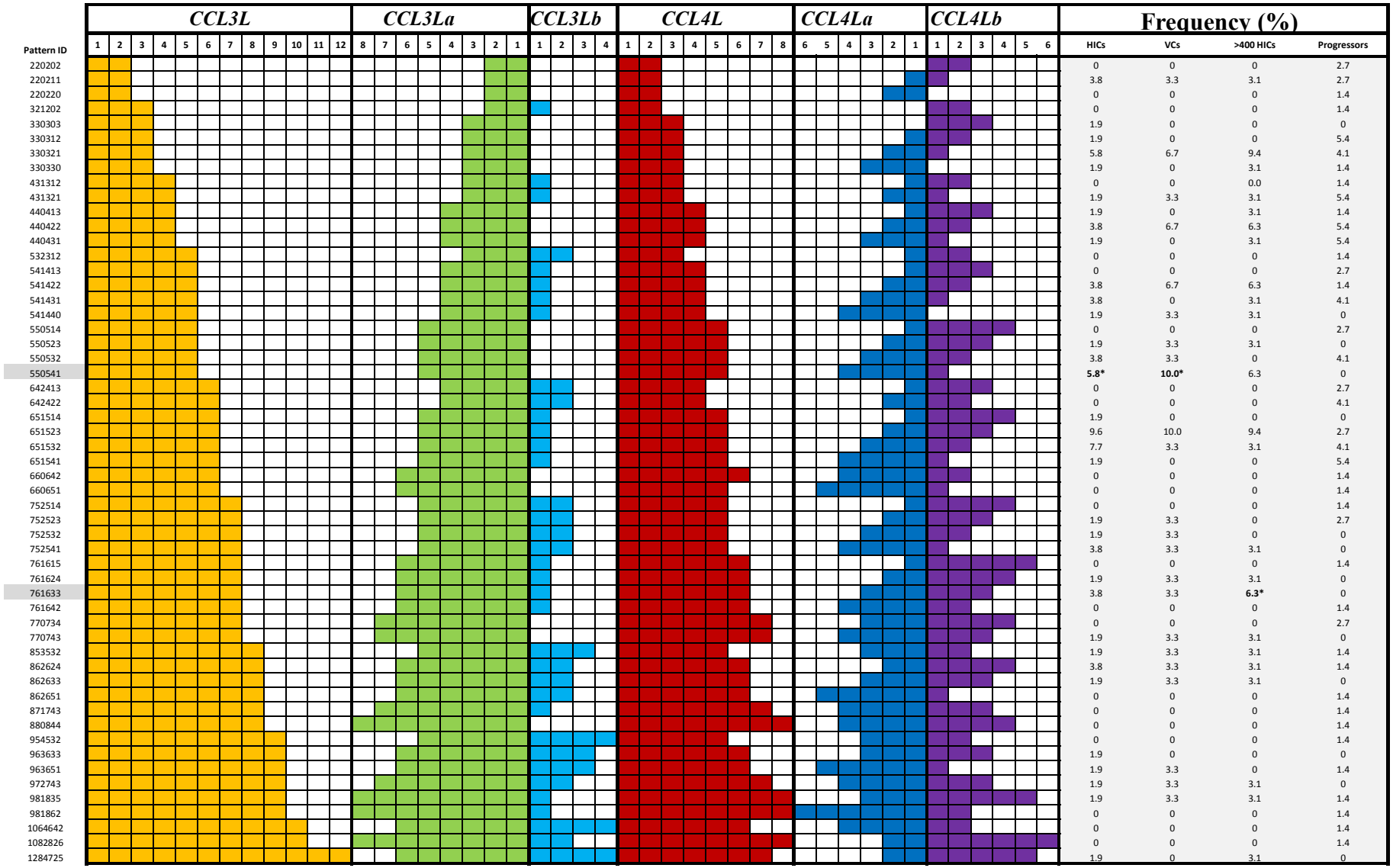


Figure 3.12: Schematic representation of the combined *CCL3L* and *CCL4L* copy number patterns in our study groups, HICs (N=52) and progressors (N=72) and their frequencies. Each coloured block represents a single gene copy. The pattern IDs which are highlighted in grey are those showing trends or significant differences. (VCs , N=30; >400 HICs, N=32).* shows where there were significant differences or trends when specific controller subgroups were compared to progressors. Fishers's exact test was used for statistical analysis.

3.3 Relationship between *CCL3La* and *CCL3Lb* and between *CCL4La* and *CCL4Lb* gene copy numbers

3.3.1 *CCL3La* and *CCL3Lb* copy number difference

Although previously thought to be a pseudogene, and given that Shostakovich-Koretskaya et al. (2009) showed that the *CCL3Lb* gene has 5' exons which result is production of 2 alternatively spliced mRNA transcripts, *CCL3Lb* might possess post-transcriptional regulatory functions. Additionally, given that we saw a positive correlation of *CCL3La* and *CCL3Lb* gene copy numbers (i.e more *CCL3La* copies associate with more *CCL3Lb* copies), this suggests that *CCL3Lb* may be involved in the downregulation of the CCL3La protein. Making an assumption that a single copy of *CCL3Lb* has a counteracting effect (or equal) to a single copy of *CCL3La*, and given that *CCL3La* copy number was always found to be greater than *CCL3Lb* copy number we compared the differences between the *CCL3La* and *CCL3Lb* copy numbers (i.e *CCL3La* – *CCL3Lb*) between HICs, HIC subgroups and progressors.

We subtracted the *CCL3Lb* copies from the *CCL3La* copies for each individual, and obtained the frequencies of each of the *CCL3La* and *CCL3Lb* copy number differences, which ranged from 0 to 8. The distribution of the *CCL3La* and *CCL3Lb* copy number differences in HICs, HIC subgroups and progressors are shown in Figure 3.13.

CCL3La and *CCL3Lb* copy number differences were compared between HICs, HIC subgroups and progressors. When compared as continuous variables, no significant differences were seen between the groups compared. However, when comparing each difference individually (comparisons shown in Table 3.2) we found a *CCL3La* and *CCL3Lb* copy number difference of 2 to be significantly underrepresented in HICs (p=0.03, OR=3.58, CI=1.13-11.36) when compared to progressors. A *CCL3La* and *CCL3Lb* copy number difference of 5 was found to be significantly overrepresented in VCs (p=0.02, OR=0.29, CI=0.09-0.88) compared to progressors. In other HICs subgroups (ECs, HVL LTNPs, <400 HICs and >400 HICs) we did not find any *CCL3La* and *CCL3Lb* copy number differences to be significantly over- or underrepresented, however a strong trend (p=0.08, OR=0.31, CI=0.08-1.16) of overrepresentation of a *CCL3La* and *CCL3Lb* copy number difference of 4 was seen in HVL LTNPs.

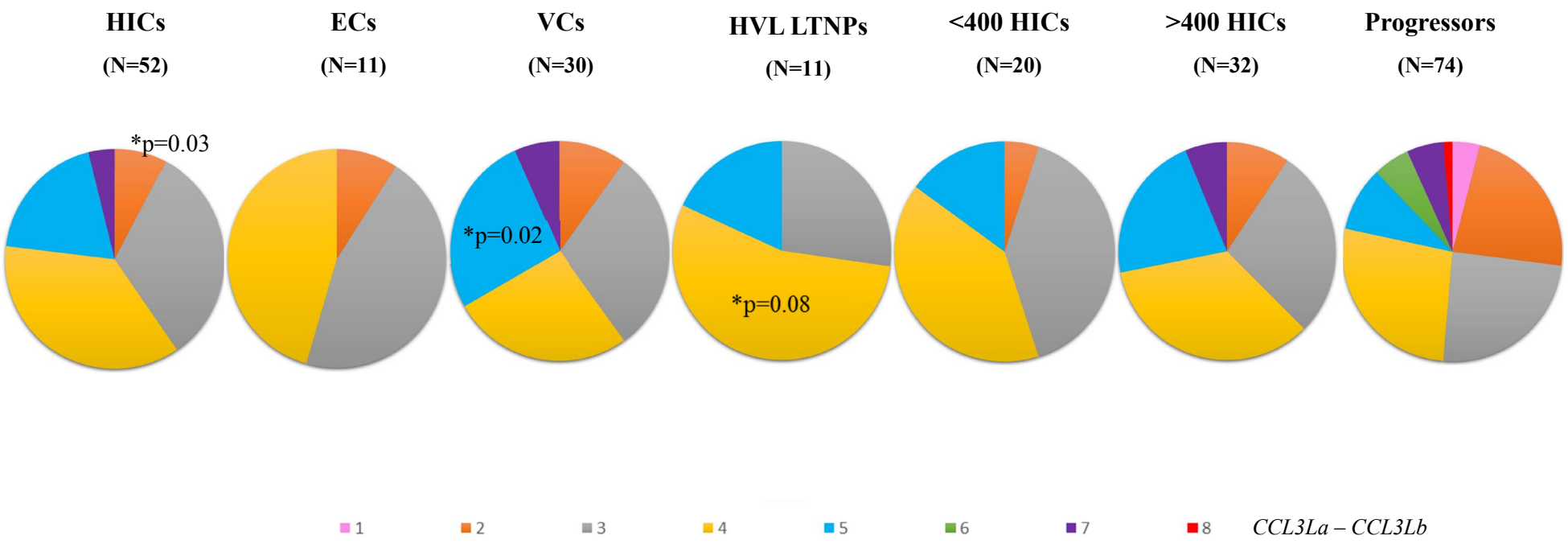


Figure 3.13: Representation of the difference between *CCL3La* and *CCL3Lb* (i.e *CCL3La – CCL3Lb*) copy numbers in HICs, HIC subgroups and progressors. Significant difference ($p<0.05$) or a trend ($0.05>p<0.1$) are shown in the Figure between HIC, HICs subgroups and progressors. Fisher's exact test was used for statistical analysis.

Table 3.2: Table showing comparisons of *CCL3La* and *CCL3Lb* copy number differences in HICs and HIC subgroups compared with progressors.

	HICs vs. P			ECs vs. P			VCs vs. P			HVL LTNPs vs. P			<400 HICs vs. P			>400 HICs vs. P		
Copy number difference	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**
1	0.27	5.14	0.26-101.7	1.00	1.13	0.05-23.27	0.56	2.99	0.15-59.62	1	1.13	0.05-23.27	0.55	2.01	0.1-40.49	0.55	3.18	0.16-63.45
2	0.03	3.58	1.13-11.36	0.44	0.44	0.36-25.00	0.17	2.68	0.72-9.95	0.11	7	0.39-125.0	0.11	5.67	0.71-45.5	0.11	2.88	0.78-10.65
3	0.32	0.66	0.30-1.45	0.16	0.39	0.10-1.42	0.62	0.75	0.29-1.92	1	0.86	0.2-3.58	0.81	0.48	0.17-1.37	0.81	0.82	0.32-2.095
4	0.33	0.64	0.3-1.38	0.29	0.44	0.12-1.62	1.00	1.02	0.39-2.66	0.08	0.31	0.08-1.16	0.49	0.56	0.2-1.56	0.49	0.71	0.29-1.725
5	0.18	0.44	0.16-1.24	0.59	2.56	0.14-47.9	0.03	0.29	0.09-0.88	0.33	0.47	0.08- 2.62	0.12	0.59	0.14-2.53	0.12	0.37	0.12-1.17
6	0.14	6.70	0.35-127.3	1.00	1.47	0.07-29.14	0.32	3.89	0.20-74.63	1	1.47	0.07- 29.14	0.31	2.62	0.14-50.68	0.31	4.15	0.22-79.42
7	1.00	1.43	0.25-8.1	1.00	1.47	0.07-29.14	1.00	0.80	0.14-4.62	1	1.47	0.07-29.14	1.00	2.62	0.14-50.68	1.00	0.86	0.15-4.94
8	1.00	2.14	0.09-53.68	1.00	0.47	0.02-12.24	1.00	1.28	0.05-32.33	1	0.47	0.02-12.24	1.00	0.84	0.03-21.34	1.00	1.33	0.05-33.46

*HICs (HIV-1 controllers, N=52), ECs (elite controllers, N=11), VCs (viraemic controllers, N=30), HVL LTNPs (long term non progressors, N=11) <400 HICs (N=20), >400 HICs (32), P (progressors N=74)

*Shaded boxes indicate relationships that are significant ($p<0.05$) and relationships that show trends ($0.05>p<0.1$). Fisher's exact test was used for statistical analysis.

3.3.2 *CCL4La* and *CCL4Lb* gene copy numbers

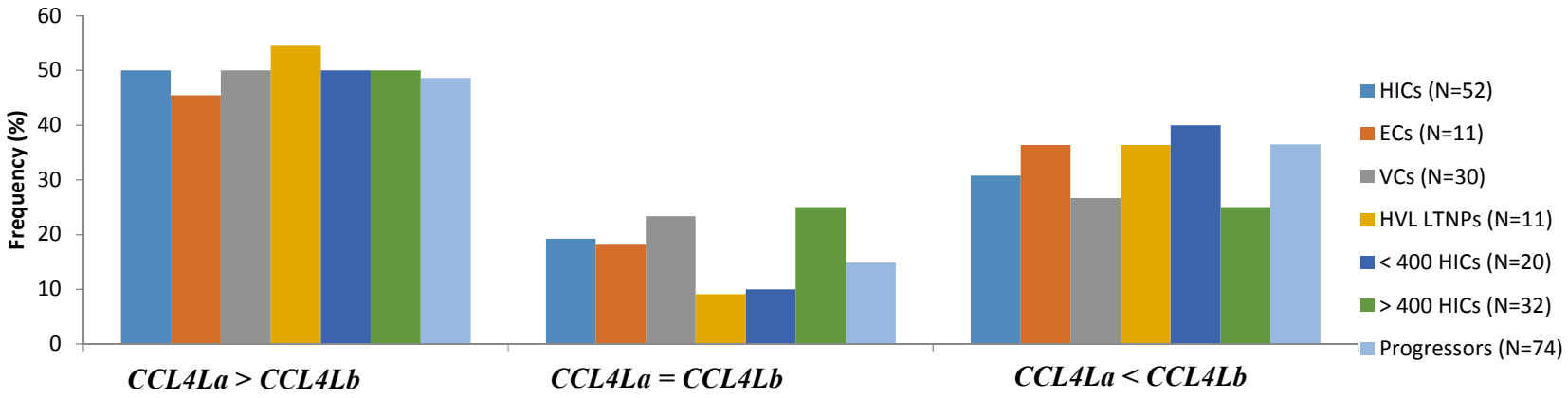
Unlike the *CCL3La* and *CCL3Lb* gene copy numbers which are positively correlated and where an individuals' *CCL3La* copy number always exceeds that of *CCL3Lb*, the *CCL4La* and *CCL4Lb* gene copy numbers are negatively correlated (Figure 3.4B iii). This suggests a possible positive feedback loop and that the balance of *CCL4La* and *CCL4Lb* products may be important. Some individuals have equal numbers of *CCL4La* and *CCL4Lb* copies (*CCL4La* = *CCL4Lb*), while some individuals have either more *CCL4La* than *CCL4Lb* (*CCL4La* > *CCL4Lb)* or less *CCL4La* than *CCL4Lb* (*CCL4La* < *CCL4Lb*). Frequencies of individuals with *CCL4La* > *CCL4Lb*, *CCL4La* = *CCL4Lb* and *CCL4La* < *CCL4Lb* are shown according to HIV-1 groupings in Figure 3.14A. The majority of individuals had more copies of *CCL4La* than *CCL4Lb* followed by individuals with less copies of *CCL4La* than *CCL4Lb* and the least number of individuals with equal numbers of *CCL4La* and *CCL4Lb*.

The comparisons of individuals with *CCL4La* > *CCL4Lb*, *CCL4La* = *CCL4Lb* and *CCL4La* < *CCL4Lb* in controllers and progressors are shown in Figure 3.14B. No significant differences were seen when comparing the distributions between HICs and progressors as well as between HICs subgroups and progressors.

3.4 *CCL3L* and *CCL4L* copy numbers and associations with markers of HIV-1 disease progression

We also investigated the association of *CCL3L* and *CCL4L* gene copy numbers with markers of HIV-1 disease progression (VL and $CD4^+$ T cell counts) in the total group of HIV-1 infected individuals (HICs and progressors combined, N=126). We found a significant but very week negative correlation between *CCL3La*/*CCL4L* gene copy numbers and VL (r = - 0.18, p=0.04). In addition, we also found a negative correlation (not significant) between *CCL3L* gene copy numbers and VL (r = - 0.17, p=0.05). Figure 3.15 shows the correlations of *CCL3L* and *CCL3La*/*CCL4L* copy numbers with viral loads. No significant correlations were seen between *CCL3Lb*, *CCL4La* and *CCL4Lb* copy numbers with viral loads. We also did not find any significant correlations between any *CCL3L* and *CCL4L* gene copy numbers with $CD4^+$ T cell count.

A



B

	HICs vs. P			ECs vs. P			VCs vs. P			HVL LTNPs vs. P			<400 HICs vs. P			>400 HICs vs. P		
Relationship	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**	**P**	**OR**	**CI**
CCL4La > CCL4Lb	1	0.95	0.47-1.93	1	1.14	0.32-4.06	1	0.95	0.40-2.21	0.76	0.79	0.22-2.81	1	0.95	0.35-2.54	1	0.95	0.35-2.55
CCL4La = CCL4Lb	0.63	0.73	0.29-1.88	0.67	0.79	0.15-4.14	0.39	0.57	0.2-1.66	1	1.75	0.2-14.04	0.73	1.57	0.32-7.75	0.27	0.52	0.19-1.46
CCL4La < CCL4Lb	0.57	0.57	0.61-2.75	1	1.01	0.27-3.75	0.37	1.58	0.62-4.04	1	1.01	0.27-3.75	0.8	0.86	0.31-2.37	0.12	2.19	0.85-5.61

Figure 3.14: Distributions (A) and comparisons (B) of the number of individuals with *CCL4La* > *CCL4Lb*, *CCL4La* = *CCL4Lb* and *CCL4La* < CCL4Lb in HICs, HICs subgroups and progressors. Fisher's exact test was used for statistical analysis**.**

However, when individuals were stratified by copy numbers > population median and ≤ population median (i.e 4 copies) we found individuals with *CCL3La*/*CCL4L* copy numbers > population median to have significantly lower viral loads (p=0.009) compared to individuals with copy numbers ≤ population median, however no significant difference were seen in $CD4^+$ T count (Figure 3.16). Additionally, no significant differences were seen in viral load or CD4+ T cell counts when individuals were stratified by *CCL3La*/*CCL4L* copy numbers ≥ population median and < population median. No significant differences were also seen in viral load or $CD4^+$ T cell count when *CCL3L*, *CCL3Lb*, *CCL4La* and *CCL4Lb* copy numbers where stratified by copy numbers > population median and ≤ population median or ≥ population median and < population median.

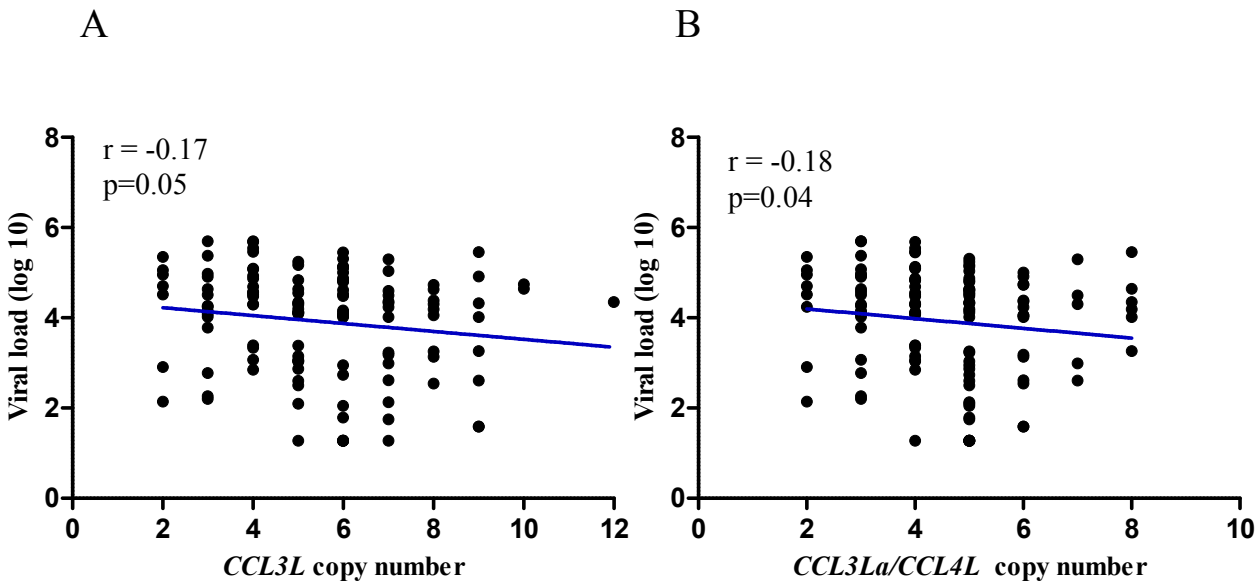


Figure 3.15: Correlations (Spearman's rank correlation) of (A) *CCL3L* and (B) *CCL3La*/*CCL4L* copy numbers with viral loads for the total group of HIV-1 infected individuals (HICs and progressors, N=126).

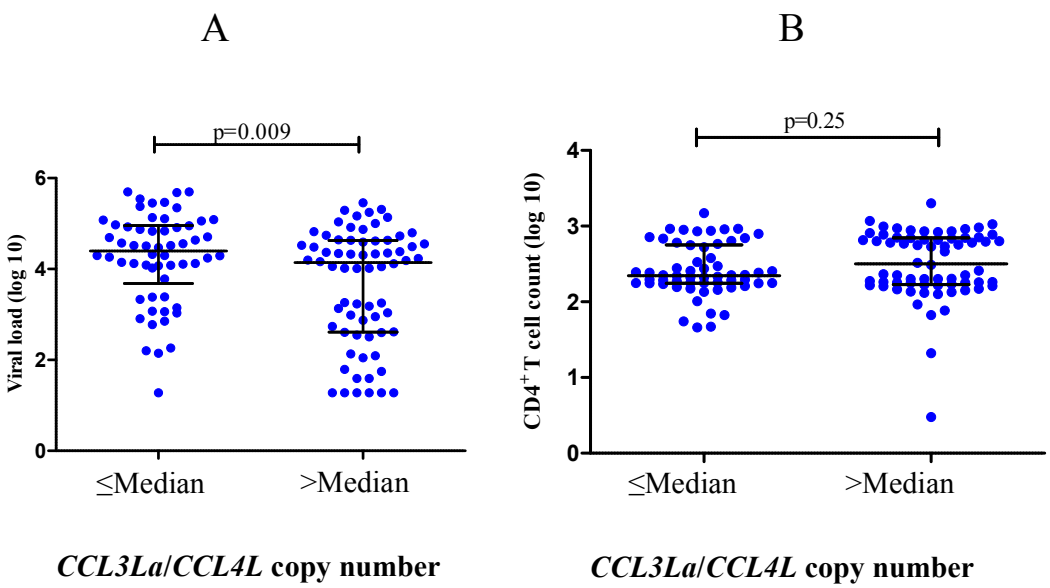


Figure 3.16: Comparison of (A) viral load and (B) $CD4^+$ T cell count (Mann-Whitney's U test) stratified according to *CCL3La*/*CCL4L* copy numbers > population median (4 copies) and ≤ population median (4 copies) for the total group of HIV-1 infected individuals (N=126). Black horizontal lines represent the median (middle) and interquartile ranges (outer lines).

3.5 *CCL3* haplotypes in healthy and HIV-1 infected individuals and control of HIV-1 infection

Using real time PCR, we determined the representation of the two haplotypes, Hap-A1 and Hap-A3 for HICs and progressors as well as the representation of Hap-2SNP haplotype (a 2-SNP haplotype composed of 2 SNPs that are shared by Hap-A1 and Hap-A3).

3.5.1 Representation of *CCL3* haplotypes, Hap-A1, Hap-A3 and Hap-2SNP in the healthy South African black population and other reference population groups

Although certain SNPs within the *CCL3* haplotypes have been investigated for their role in HIV-1 related outcomes (Gonzalez et al., 2001, Modi et al., 2006, Hu et al., 2012), the presence and the prevalence of the three *CCL3* haplotypes, Hap-A1, Hap-A3 and Hap-2SNP in other reference populations have not been described. In order to determine if these 3 haplotypes are present in other populations, we obtained the genotypic data for the SNPs making up the haplotypes from the 1000 genomes database (http://browser.1000genomes.org/index.html) for the total African population (AFR), Yoruba

subpopulation (YRI) and the European population (EUR). We also had access to haplotypic data for healthy controls from the black South African population (Paximadis et al., 2009). Using the haploview software we tested the SNPs making up the haplotypes for linkage disequilibrium (LD) and deviation from the Hardy-Weinberg equilibrium in the different population groups. All the SNPs that form part of Hap-A1 and Hap-A3 did not significantly deviate from the Hardy-Weinberg equilibrium in the black South African population, Yoruba subpopulation and the European population. However, in the total African population from the 1000 genomes database, the SNPs that make up Hap-2SNP, i.e rs6505506 and rs1063340 significantly deviated from the Hardy-Weinberg equilibrium ($p=0.03$ and $p=0.02$, respectively). The linkage disequilibrium plots are given in Figure 3.17.

LD analysis showed that in addition to Hap-A1 and Hap-A3 haplotypes being present in the healthy South African black population (D'=1), they are also present in the Yoruba subpopulation (D'=1) and strong linkage was seen in the total African population. The two SNP haplotype, Hap-2SNP (rs1719130 + rs1063340) is also present in the total African population (D'=1), South African black population (D'=1), Yoruba subpopulation (D'=1) and the European population (D'=0.97).

Having established the presence of Hap-A1 and Hap-A3 in the total African group and the Yoruba subpopulation and the presence of the Hap-2SNP in the total African population, Yoruba subpopulation and the European population, we obtained the allelic and genotypic frequencies of the tagSNPs of the three haplotypes from the 1000 genomes database for the same populations. In addition, we had access to the frequencies of the three haplotypes for healthy controls from the black South African populations (Paximadis et al., 2013). We subsequently compared the allelic and genotypic frequencies of our South African HIV-1 uninfected individuals (healthy controls) to the frequencies of the total African group, Yoruba subpopulation and European population. Representations of allelic and genotypic frequencies of Hap-A1, Hap-A3 and Hap-2SNP (tagSNPs) for the same populations are shown in Figure 3.18.

No significant differences were seen when comparing the Hap-A1 allelic frequencies between the South African black population and the African population as well as the Yoruba subpopulation. However, a trend of overrepresentation of the Hap-A1 allele was seen in the South African black population ($p=0.07$, OR=1.48, CI=0.98-2.23) compared to African population. No significant differences were also seen for Hap-A1 homozygosity when the South African black population was compared to the African population as well as the Yoruba

subpopulation, however, heterozygosity for Hap-A1 was significantly overrepresented in the South African black population (p=0.03, OR=1.68, CI=1.05-2.69) compared to the African population.

Although the Hap-A3 homozygous genotype which is virtually absent in the South African black population was present in relatively low frequencies in the African population as well as the Yoruba subpopulation, no significant differences were seen when Hap-A3 allelic and genotypic frequencies were compared between the South African black population and African population as well as the Yoruba subpopulation. When Hap-2SNP allelic and genotypic frequencies were compared between the South African black population, African population, Yoruba subpopulation, as well as the European population, no significant differences were seen. However, a trend of underrepresentation of the Hap-2SNP allelic frequency (p=0.07, OR=0.71, CI=0.50-1.02) and heterozygosity (p=0.06, OR=0.65, CI=0.42-1.02) was seen in the South African black population compared to the European population.

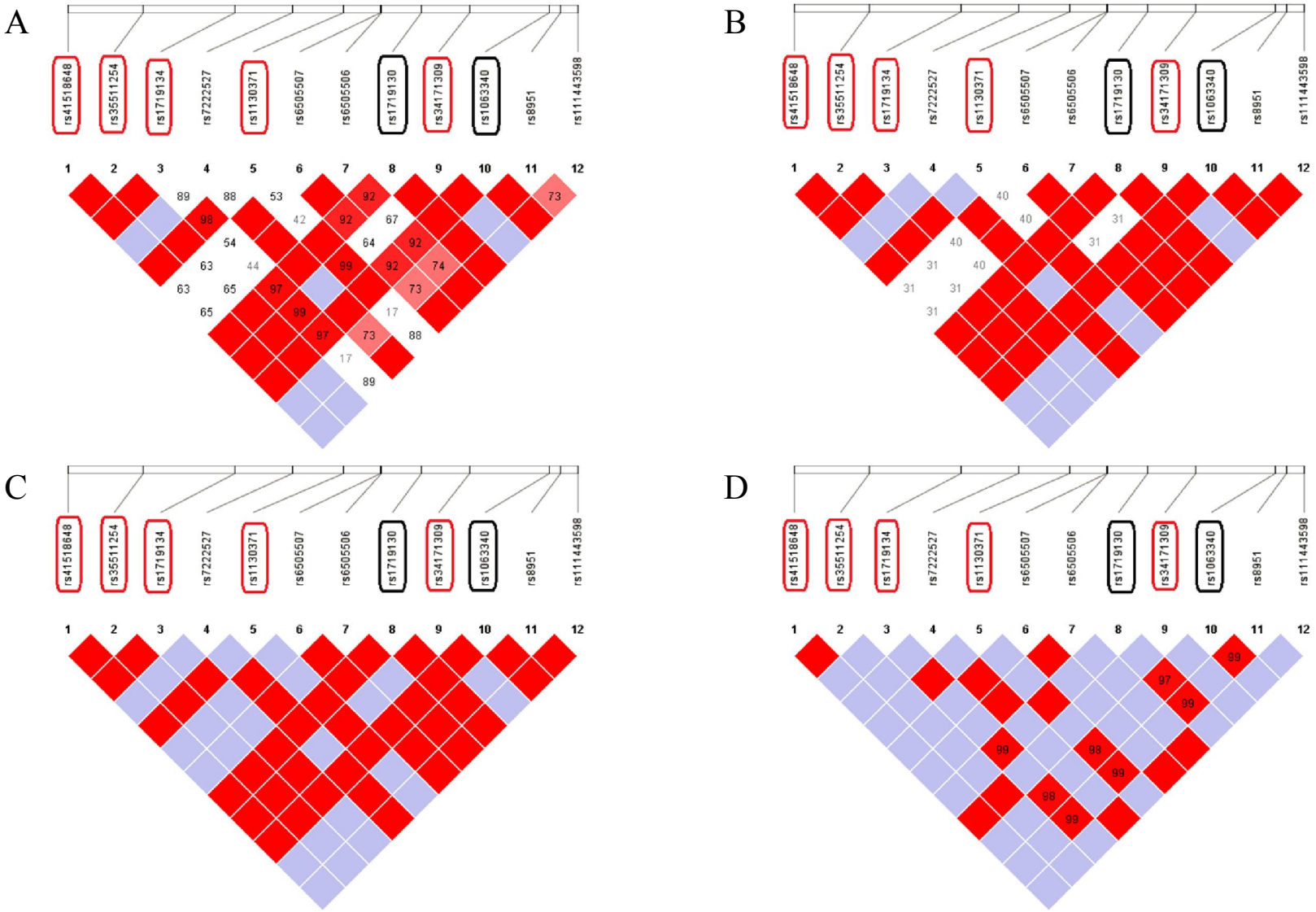


Figure 3.17: Linkage disequilibrium plots of the SNPs making up Hap-A1 and Hap-A3 haplotypes. (A) African population (1000 genomes), (B) Yoruba Sub-population, (C) black South African population and the (D) European population. SNPs enclosed in red rectangles are Hap-A1 SNPs, SNPs not enclosed are Hap-A3 SNPs and SNPs enclosed in black rectangles are shared between Hap-A1 and Hap-A3. Empty red squares indicate complete linkage of D'=1, while numbers inside the squares indicate D' values in percentages. Blue squares indicate that there is no linkage disequilibrium between the SNPs. Data for AFR, YRI and EUR sourced from 1000genomes.org and the data for SAA was sourced from (Paximadis et al., 2009).

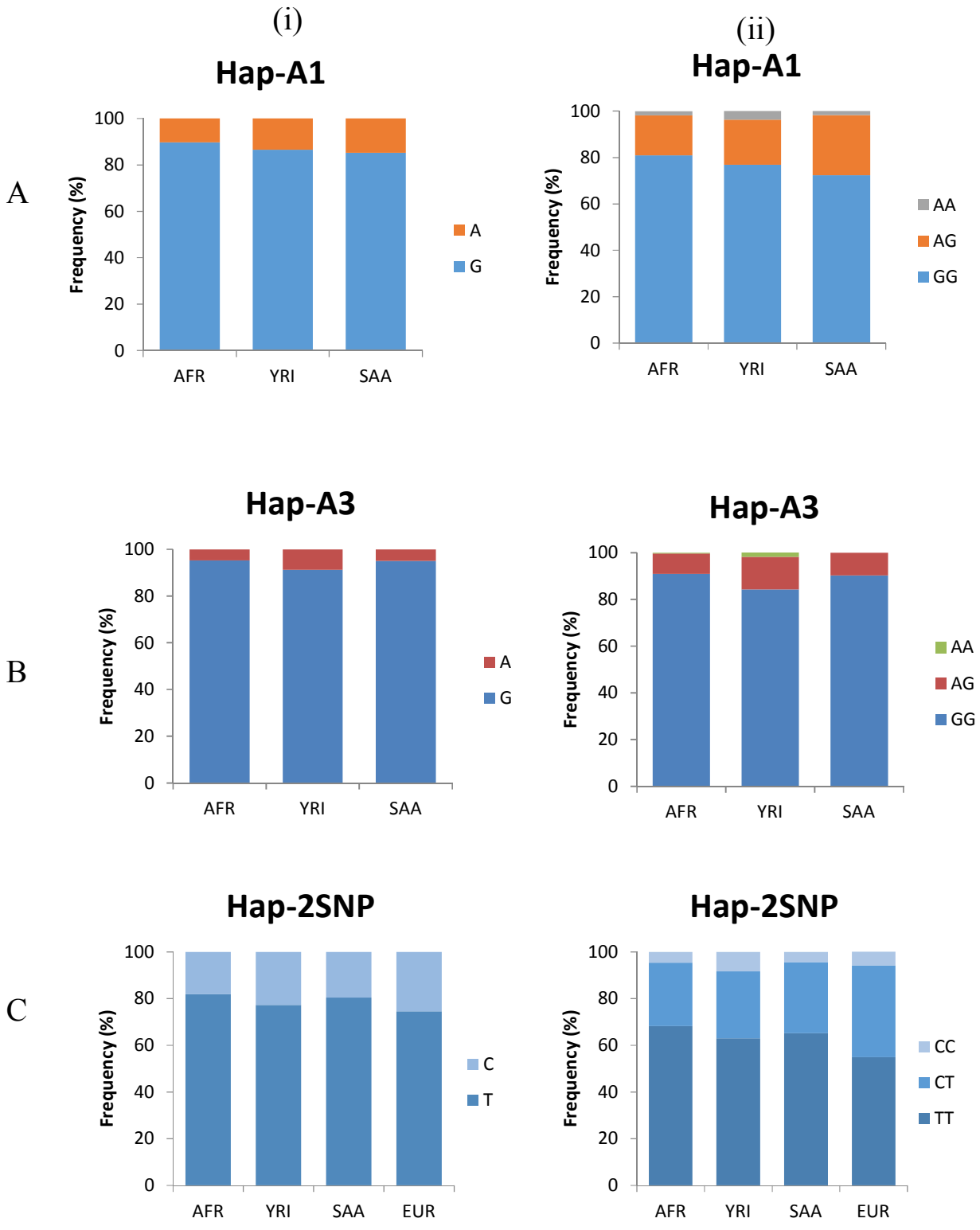


Figure 3.18: Allelic (i) and genotypic (ii) frequencies of (A) Hap-A1, (B) Hap-A3 and (C) Hap-2SNP tagSNPs in the total African population (AFR, N=661), Yoruba subpopulation (YRI, N=108), black South African population (SAA, N=112) and the European population (EUR, N=108). Data for AFR, YRI and EUR sourced from 1000genomes.org and the data for SAA was sourced from(Paximadis et al., 2013). Fisher's exact test was used for statistical analysis.

3.5.2 Comparisons of Hap-A1, Hap-A3 and Hap-2SNP allelic and genotypic frequencies between HIV-1 controllers and healthy controls and progressors

In mother-to-child transmission of HIV-1, Paximadis et al. (2013), found Hap-A1 and Hap-A3 to be associated with protection from HIV-1 infection *in utero* and increased transmission via the intrapartum route, respectively. However, a possible role of the *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2SNP) has not been investigated in the context of HIV-1 control, hence we investigated the role of these haplotypes in natural control of HIV-1.

The minor and major allele frequencies and genotype frequencies of the Hap-A1, Hap-A3 and Hap-2SNP tagSNPs in all study groups are given in Table 3.3. In addition, bar graphs showing the representations of the allelic and genomic distributions of the 3 haplotypes are shown in Figure 3.19. To determine whether the *CCL3* haplotypes, Hap-A1, Hap-A3 and Hap-2SNP were associated with natural control of HIV-1, we compared the allelic and genotypic frequencies of the haplotype tagSNPs between HICs and HIC subgroups with progressors, these results are shown in Tables 3.4 and 3.5. When comparing Hap-A1 allelic and genotypic frequencies between HICs, HIC subgroups and progressors, no significantly over- or underrepresented alleles or genotypes were seen. Even though homozygosity for Hap-A1 was absent in both the ECs and <400 HICs, both groups were not significantly different from progressors. Hap-A3 and Hap-2SNP allelic and genotypic frequencies did not significantly differ between HICs and progressors as well as HIC subgroups and progressors. Again, although Hap-A3 was not present in both ECs and HVL LTNPs, no significant differences were seen when comparing to progressors. The absence of the Hap-A3 haplotype in ECs and HVL LTNPs could have been due to Hap-A3 being rare in black South Africans and to their small sample size (N=11 for each). Homozygosity for the Hap-2SNP was also absent in both ECs and <400 HICs, with no significant differences noted when compared to progressors.

Having access to allelic and genotypic frequencies for HIV-1 uninfected healthy controls from the black South African population (HCs; N=112) from previous work (Paximadis et al., 2013), allowed us to compare representation of these haplotypes between HCs and HICs (and HICs subgroups) and HCs and progressors. No significant shifts in representation were seen for any comparison of the three haplotypes (shown in Tables 3.4 and 3.5). Interestingly, although no significant differences were seen for any of the haplotypes when comparing to both HCs and progressors, homozygosity for the Hap-A1 haplotype showed trends of overrepresentation in the >400 HICs compared to both HCs (p=0.09, OR=0.20, CI=0.01-1.25) and progressors (p=0.09, OR=0.14, CI=0.01-1.37).

Table 3.3: Total number and frequency (%) of Hap-A1, Hap-A2 and Hap-2SNP tagSNP alleles and genotypes in healthy controls, HIV-1 controllers, progressors and HIV-1 controller subgroups.

Groups *CCL3 haplotypes*	Healthy controls* (HCs); N=112	HIV-1 controllers (HICs); N=52	Progressors (P) N=74	HIV-1 Controller subgroups				
				Elite controllers (EC); N=11	Viraemic controllers; N=30	HVL LTNP N=11	<400 HICS; N=20	>400 HICS N=32
Hap-A1								
Allele								
C	191(85.27)	90 (86.54)	132(89.19)	20 (90.91)	52 (86.67)	18 (81.82)	37 (92.50)	53 (81.81)
T	33 (14.73)	14 (13.46)	16 (10.81)	2 (9.09)	8 (13.33)	4(18.18)	3 (7.50)	11 (17.19)
Genotype								
CC	81 (72.32)	41 (78.85)	59(79.73)	9 (81.82)	24 (80.00)	8 (72.73)	17 (85.00)	24 (75.00)
CT	29 (25.89)	8 (15.38)	14 (18.92)	2 (18.18)	4 (13.33)	2 (18.18)	3 (15.00)	5 (15.63)
TT	2 (1.79)	3 (5.77)	1 (1.35)	0 (0.00)	2 (6.67)	1 (9.09)	0 (0.00)	3 (9.38)
Hap-A3								
Allele								
C	213 (95.09)	99(95.19)	140 (94.59)	22 (100.00)	55 (91.67)	22 (100.00)	36 (90.00)	63 (97.62)
T	11 (4.91)	5 (4.81)	8 (5.41)	0 (0.00)	5 (8.33)	0 (0.00)	4 (10.00)	1 (1.56)
Genotype								
CC	101 (90.18)	47 (90.38)	66 (89.19)	11 (100.00)	25 (83.33)	11 (100.00)	16 (80.00)	31 (96.88)
CT	11 (9.82)	5 (9.62)	8 (10.81)	0 (0.00)	5 (16.67)	0 (0.00)	4 (20.00)	1 (3.13)
TT	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Hap-2-SNP								
Allele								
A	180 (80.36)	85 (81.73)	124 (83.78)	20 (90.91)	47 (78.33)	18 (81.82)	33 (82.50)	52 (81.25)
G	44 (19.64)	19 (18.27)	24 (16.22)	2 (9.09)	13 (21.67)	4(18.18)	7 (17.50)	12 (18.75)
Genotype								
AA	73 (65.18)	36 (69.23)	53 (71.62)	9 (81.82)	19 (63.33)	8 (72.73)	13 (65.00)	23 (71.88)
AG	34 (30.36)	13 (25.00)	18 (24.32)	2 (18.18)	9 (30.00)	2 (18.18)	7 (35.00)	6 (18.75)
GG	5 (4.46)	3 (5.77)	3 (4.05)	0 (0.00)	2 (6.67)	1 (9.09)	0 (0.00)	3(9.38)

*Data for healthy controls taken from: Paximadis et al. (2013).

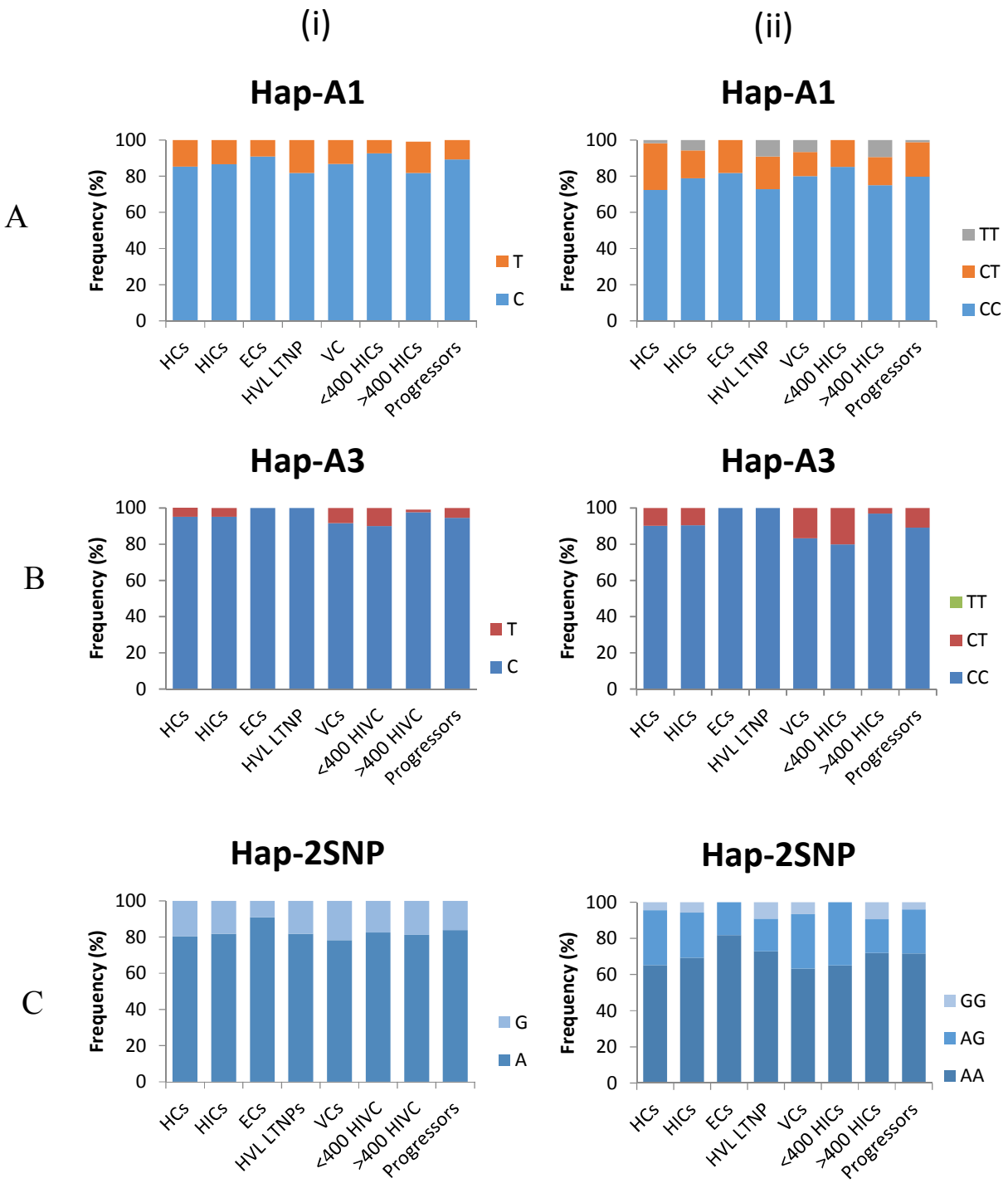


Figure 3.19: Bar graphs showing representation of (A) Hap-A1, (B) Hap-A3 and (C) Hap-2SNP allelic (i) and genotypic (ii) frequencies in healthy controls (HCs, N=112), HIV-1 controllers (HICs, N=52), elite controllers (ECs, N=11), high viral load long term non progressors (HVL LTNPs, N=11), viraemic controllers (VCs, N=32), <400 HICs (N=20), >400 HICs (N=32) and progressors (N=74). Fisher's exact test was used for statistical analysis.

Table 3.4: Comparisons of Hap-A1, Hap-A3 and Hap-2SNP allelic frequencies between healthy controls and HICs, and HIC subgroups and between progressors and HICs and HIC subgroups.

Healthy controls vs. HIV-1 controllers

Hap-A1	HCs vs. HICs			HCs vs. ECs			HCs vs. VCs			HCs vs. HVL LTNPs			HCs vs. <400 HICs			HCs vs. >400 HICs		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (C)	0.90	0.46-1.77	0.87	0.58	0.13-2.59	0.75	0.89	0.39-2.05	1	1.29	0.41-4.04	0.75	0.47	0.14-1.61	0.32	1.2	0.57-2.54	0.69
Mt (T)	1.11	0.57-2.18	0.87	1.73	0.39-7.44	0.75	1.12	0.49-2.58	1	0.78	0.25-2.44	0.75	2.13	0.62-7.31	0.32	0.83	0.39-1.76	0.69
Hap-A3	**HCs vs. HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. <400 HICs**			**HCs vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (C)	0.98	0.33-2.89	1	0.41	0.02-7.24	0.61	1.76	0.59-5.28	0.34	0.41	0.02-7.24	0.61	2.15	0.65-7.13	0.26	0.31	0.04-2.43	0.48
Mt (T)	1.02	0.35-3.02	1	2.42	0.14-42.55	0.61	0.57	0.19-1.70	0.34	2.42	0.14-42.55	0.61	0.46	0.14-1.54	0.26	3.25	0.41-25.70	0.48
Hap-2-SNP	**HCs vs. HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. <400 HICs**			**HCs vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (A)	0.91	0.50-1.66	0.88	0.41	0.09-1.82	0.39	1.13	0.56-2.27	0.72	0.91	0.29-2.82	1	0.87	0.36-2.09	0.83	0.94	0.46-1.9	1
Mt (G)	1.09	0.60-1.99	0.88	2.44	0.55-10.86	0.39	0.88	0.44-1.78	0.72	1.10	0.35-3.41	1	1.15	0.48-2.78	0.83	1.06	0.52-2.15	1

Progressors vs. HIV-1 controllers

Hap-A1	P vs. HICs			P vs. EC			P vs. VCs			P vs. HVL LTNPs			P vs. <400 HICs			P vs. >400 HICs		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (C)	1.28	0.60-2.76	0.56	0.82	0.18-3.86	1	1.27	0.51-3.15	0.63	1.83	0.55-6.10	0.30	0.67	0.18-2.42	0.77	1.71	0.75-3.93	0.26
Mt (T)	0.78	0.36 - 1.68	0.56	1.21	0.26-5.68	1	0.79	0.32-1.95	0.63	0.55	0.16-1.81	0.30	1.50	0.41-5.41	0.77	0.58	0.25-1.34	0.26
Hap-A3	**P vs. HICs**			**P vs. EC**			**P vs. VCs**			**P vs. HVL LTNPs**			**P vs. <400 HICs**			**P vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (C)	0.88	0.28-2.78	1	0.37	0.02-6.59	0.60	1.59	0.50-5.08	0.53	0.37	0.02-6.59	0.60	1.94	0.55-6.82	0.29	0.28	0.03-2.27	0.28
Mt (T)	1.13	0.36-3.56	1	2.72	0.15-48.86	0.60	0.63	0.20-2.01	0.53	2.27	0.15-48.86	0.60	0.51	0.15-1.80	0.29	3.60	0.44-29.4	0.28
Hap-2-SNP	**P vs. HICs**			**P vs. EC**			**P vs. VCs**			**P vs. HVL LTNPs**			**P vs. <400 HICs**			**P vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
Wt (A)	1.16	0.60-2.24	0.73	0.52	0.11-2.36	0.53	1.43	0.67-3.04	0.42	1.15	0.36-3.69	0.76	1.10	0.43-2.77	0.81	1.19	0.55-2.56	0.69
Mt (G)	0.87	0.45-1.68	0.73	1.94	0.42-8.83	0.53	0.70	0.33-1.49	0.42	0.87	0.27-2.80	0.76	0.91	0.36-2.30	0.81	0.84	0.39-1.80	0.69

*Fisher's exact test was used for statistical analysis

Table 3.5: Comparisons of Hap-A1, Hap-A3 and Hap-2SNP genotypic frequencies between healthy controls and HICs, and HIC subgroups and between progressors and HICs and HIC subgroup.

Healthy Controls vs. HIV-1 controllers																		
Hap-A1	**HCs vs. HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. <400 HICs**			**HCs vs. >400 HICs**		
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.84	0.77-4.37	0.22	1.61	0.33-7.90	0.73	2.15	0.69-6.72	0.22	1.43	0.29-7.14	1.00	2.03	0.55-7.43	0.40	1.72	0.60-4.92	0.47
TT	0.34	0.05-2.10	0.34	0.58	0.03-13.08	1.00	0.30	0.04-2.12	0.24	0.20	0.02-2.43	0.27	1.07	0.05-23.38	1.00	0.20	0.03-1.25	0.09
Hap-A3	**HCs vs. HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. <400 HICs**			**HCs vs. >400 HICs**		
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.02	0.34-3.11	1.00	2.61	0.14-47.22	0.60	0.54	0.17-0.71	0.33	2.61	0.14-47.22	0.60	0.44	0.12-1.54	0.24	3.38	0.42-27.21	0.30
Hap-2-SNP	**HCs vs. HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. <400 HICs**			**HCs vs. >400 HICs**		
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
AG	1.29	0.61-2.74	0.58	2.10	0.43-10.23	0.50	0.98	0.40-2.40	1.00	1.86	0.38-9.25	0.72	0.87	0.32-2.36	0.80	1.79	0.67-4.79	0.36
GG	0.82	0.19-3.63	1.00	1.42	0.07-27.871	1.00	0.65	0.12-3.62	0.64	0.55	0.06-5.29	0.49	2.02	0.11-38.73	1.00	0.53	0.12-2.37	0.41
Progressors vs. HIV-1 controllers																		
Hap-A1	**P vs. HICs**			**P vs. ECs**			**P vs. VCs**			**P vs. HVL LTNPs**			**P vs. <400 HICs**			**P vs. >400 HICs**		
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.22	0.47-3.16	0.81	1.07	0.21-5.50	1.00	1.42	0.43-4.77	0.77	0.95	0.18-4.97	1.00	1.35	0.35-5.23	1.00	1.14	0.37-3.51	1.00
TT	0.23	0.02-2.31	0.31	0.48	0.02-12.65	1.00	0.20	0.02-2.35	0.22	0.14	0.01-2.39	0.25	0.88	0.03-22.65	1.00	0.14	0.01-1.37	0.09
Hap-A3	**P vs. HICs**			**P vs. ECs**			**P vs. VCs**			**P vs. HVL LTNPs**			**P vs. <400 HICs**			**P vs. >400 HICs**		
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.14	0.35-3.70	1.00	2.94	0.12-42.72	0.59	0.61	0.81-2.03	0.51	2.94	0.16-54.55	0.59	0.48	0.13-1.81	0.28	3.76	0.45-31.39	0.27
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
AG	0.94	0.41-2.16	1.00	1.52	0.30-7.75	1.00	0.76	0.29-1.96	0.62	1.36	0.26-7.00	1.00	0.63	0.22-1.83	0.41	1.30	0.46-3.70	0.80
GG	0.68	0.13-3.56	0.69	1.25	0.06-22.07	1.00	0.54	0.08-3.47	0.61	0.46	0.04-4.91	0.46	1.77	0.09-36.32	1.00	0.43	0.08-2.31	0.38

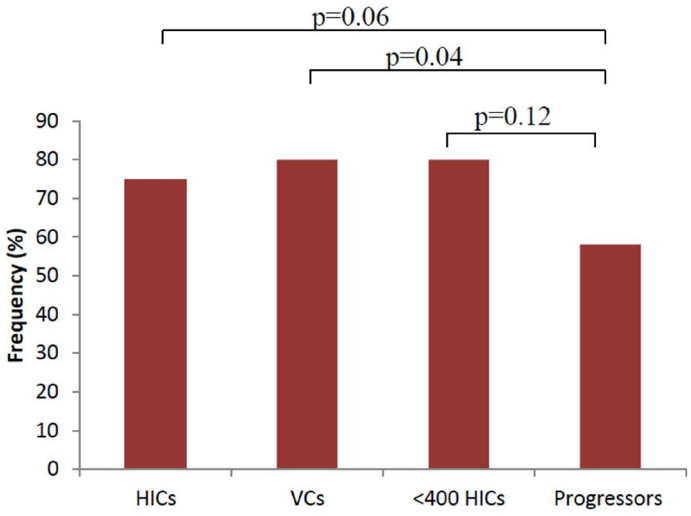
HCs, N=112; HICs, N=52; ECs, N=11; VCs, N=30; HVL LTNPs, N=11; <400 HICs, N=20; >400 HICs, N=32, progresssors (P, N=74)
Grey shaded values indicate relationships that show trends of over representation ($0.05<p<0.1$) Fisher's exact test was used for statistical analysis.

3.6 *CCL3* haplotypes and *CCL3La* copy number

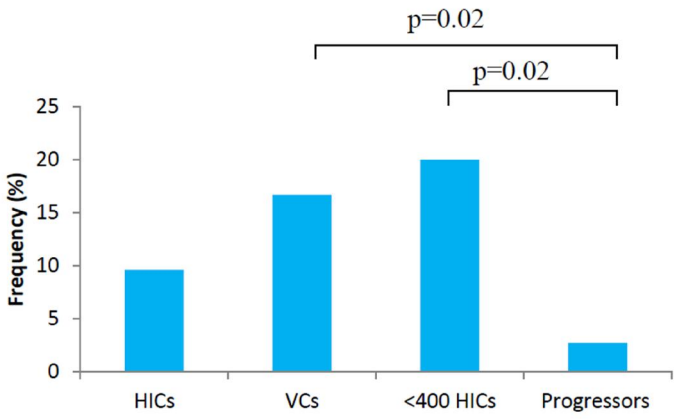
In mother to child transmission of HIV-1, Paximadis et al. (2013) found that having either a high *CCL3L* copy number and/or Hap-A1 was more significantly associated with protection from HIV-1 infection *in utero* compared to having high *CCL3L* copy number alone or Hap-A1 alone. Hence in our current study, although all the *CCL3* haplotypes did not show any significant associations with HIV-1 control, we also investigated whether or not *CCL3La/CCL4L* copy numbers in combination with *CCL3* haplotypes are more significantly associated with control of HIV-1. When comparing HICs, HIC subgroups and progressors for having either a high *CCL3La/CCL4L* copy number (copy number > population median, i.e > 4 copies) and/or the presence of Hap-A1 (heterozygous or homozygous), although trends and significant differences were seen, they were not more statistically significant than high copy number alone. Additionally, having high *CCL3La/CCL4L* copy number and/or Hap-2SNP was not more significantly associated with HIV-1 control than high copy number alone.

Given that Hap-A3 and low *CCL3La* copy numbers have been previously shown to be associated with increased intrapartum transmission of HIV-1 and increased susceptibility to HIV-1, respectively (Gonzalez et al., 2005, Paximadis et al., 2013), we also looked at Hap-A3 in combination with *CCL3La/CCL4L* copy number. In VCs (p=0.02, OR=0.14, CI=0.03-0.76) compared to progressors having a high *CCL3La/CCL4L* copy number (>4 copies) and Hap-A3 was more significantly associated with control of HIV-1 than high copy number alone. Figure 3.20 shows the distribution of individuals with high *CCL3La/CCL4L* copy numbers and/or Hap-A1, high *CCL3La/CCL4L* copy numbers and Hap-A3 and high *CCL3La/CCL4L* copy number and Hap-2SNP in HICs, VCs, <400 HICs and progressors.

A



B



C

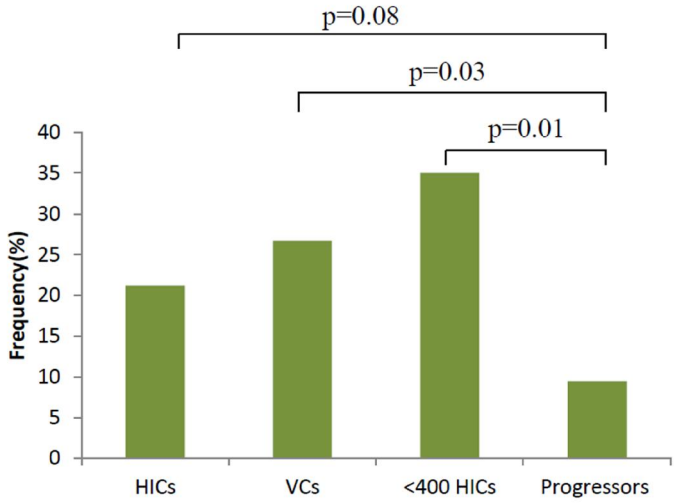


Figure 3.20: Graphs showing the distribution of individuals with (A) high *CCL3La* copy numbers (> population median) and/or Hap-A1 (B) high *CCL3La* copy numbers and Hap-A3 and (C) high *CCL3La* copy number and Hap-2SNP in HICs (N=52), HIC subgroups (VCs, N=30; <400 HICs, N=20) and progressors (N=74). Fisher's exact test was used for statistical analysis.

3.7 Summary of significant findings

Table 3.6 shows the summary of all the significant findings with regards to natural control of HIV-1. All the findings point to the *CCL3La*/*CCL4L* copy number > population median being protective and *CCL3La*/*CCL4L* ≤ population median being deleterious.

Table 3.6: Summary of all significant findings when HICs and HIC subgroups were compared to progressors.

Association	Variant description	P	OR	CI
↑VCs vs. ↓P	*CCL3La*/*CCL4L* (> population median)	0.03	0.36	0.15-0.9
↑<400 HICs vs. ↓P		0.01	0.21	0.07-0.97
↑HICs vs. ↓P	*CCL3La*^{high} and/or *CCL4La*^{high}	0.01	0.38	0.18-0.80
↑VCs vs. ↓P		0.02	0.31	0.12-0.78
↑<400 HICs vs. ↓P		0.01	0.21	0.06-0.70
↑ECs vs. ↓P	651 *CCL3L* gPID	0.02	0.17	0.04-0.66
↑<400 HICs vs. ↓P		0.04	0.26	0.08-0.82
↑HICs vs. ↓P	633 *CCL4L* gPID	0.03	0.07	0.004-1.38
↑>400 HICs vs. ↓P		0.03	0.06	0.003-1.13
↑VCs vs. ↓P	550541 *CCL3L* and *CCL4L* gPID	0.02	0.05	0.003-1.05
↓HICs vs. ↑P	*CCL3La* – *CCL3Lb* = 2	0.03	3.58	1.13-11.36
↑VCs vs. ↓P	*CCL3La* – *CCL3Lb* = 5	0.02	0.29	0.08-1.16
↑VCs vs. ↓P	*CCL3La* ^{high} and/or Hap-A1	0.04	0.35	0.13-0.95
↑VCs vs. ↓P	*CCL3La*^{high} and Hap-A3	0.02	0.14	0.03-0.76
↑<400 HICs vs. ↓P		0.02	0.11	0.02-0.66
↑VCs vs. ↓P	*CCL3La*^{high} and Hap-2SNP	0.03	0.29	0.09-0.88
↑<400 HICs vs. ↓P		0.01	0.19	0.06-0.65

↑: Overrepresented , ↓: underrepresented

CHAPTER 4: Discussion and conclusion

Sub-Saharan Africa has the highest burden of HIV-1/AIDS in the world, accounting for over two thirds of people living with HIV-1 globally, with South Africa having the highest number of people living with HIV-1/AIDS (UNAIDS, 2016). Although the use of ARVs has lessened the burden of HIV-1/AIDS, some regions, mostly in sub-Saharan Africa, still do not have access ARVs. Additionally, the use ARVs does not result in complete restoration of the immune system and interruption of the treatment often results in rapid rebound of viral replication, illustrating the importance of finding a cure for HIV-1 (Currier et al., 2008, Choi et al., 2009, Piot and Quinn, 2013).

In the absence of ARV treatment, while some individuals progress rapidly to AIDS (rapid progressors), HIV-1 controllers are able to suppress viral load and/or maintain $CD4^+$ T cell count >500 cells/µl without the use of ARVs (Okulicz et al., 2009). HIV-1 infected controllers, more specifically elite controllers, characterised by their ability to control viral replication, represent a model for a HIV-1 functional cure (O'Connell et al., 2009). HIV-1 functional cure is defined as long term, host-mediated suppression of viral replication and restoration of the immune system with remission of symptoms and reduced risk of transmission without the use of ARVs (Katlama et al., 2013, Cockerham and Hatano, 2015). Compared to a sterilizing cure which involves complete eradication of replication competent viruses, HIV-1 functional cure is a more achievable goal. The study of HIV-1 controllers can help identify targets which can be potentially manipulated in the development of HIV-1 vaccines and novel therapeutics to achieve functional cure.

Multiple factors including both viral and host factors have been implicated in natural control of HIV-1 infection. However, the isolation of replication competent viruses in some HIV-1 controllers suggests that most of the control is at the level of the host (Alexander et al., 2000, Lamine et al., 2007), with variation in host genes that are involved in HIV-1 cellular entry and immune response being associated with different HIV-1 related outcomes (Singh et al., 2008). Although some HIV-1 controllers have been found to be in possession of known protective alleles, some controllers lack any known protective factors, whereas some HIV-1 infected individuals in possession of known protective alleles have normal progression to AIDS, suggesting that there are other protective host factors which are yet to be identified (Cockerham and Hatano, 2015).

The CC chemokines, CCL3 and CCL4, which are natural ligands of the CCR5 receptor (a major coreceptor for HIV-1 entry) and inhibitors of HIV-1 cellular entry *in vitro*, are among the host factors which have been found to play a role in both HIV-1 susceptibility and the rate of progression to AIDS. In this study, we therefore investigated the role of the CCL3 and

CCL4 chemokines in natural control of HIV-1 infection in the black South African population, by comparing the *CCL3L* and *CCL4L* gene copy numbers as well as previously identified *CCL3* haplotypes, between HIV-1 infected ARV-naive controllers and progressors.

It is interesting that CCL3 and CCL4 are the only chemokines encoded by genes that occur as two copies per diploid genome (*CCL3* and *CCL4*) as well as non-allelic variants (*CCL3L* and *CCL4L*) found in variable copy numbers. In recent years, copy number variation in a number of genes has been found to be an important factor in various diseases, mostly infectious and autoimmune diseases (Zarrei et al., 2015). Low Copy number of the *human beta-defensin 2* gene has been associated with higher risk of developing colonic Crohn disease (Fellermann et al., 2006), while an increased copy number of the *human beta-defensin* genes has been associated with psoriasis (Hollox et al., 2008). Copy number variation in the *CCL3L* and *CCL4L* genes has been associated with various autoimmune diseases (Burns et al., 2005, Mamtani et al., 2007, McKinney et al., 2008) and infectious diseases including HIV-1 infection (Gonzalez et al., 2005, Meddows-Taylor et al., 2006, Kuhn et al., 2007, Shalekoff et al., 2008, Shostakovich-Koretskaya et al., 2009).

Gonzalez et al. (2005) were the first to report an association between *CCL3L1* copy number lower than the population specific median with a number of HIV-1 related outcomes including increased susceptibility to HIV-1 infection, higher rate of progression to AIDS, high viral load set point and higher rate of $CD4^+$ T cell decline. They demonstrated that it was the copy number relative to the population median rather than the absolute copy number that was associated with HIV-1 related outcomes, implicating population specific levels of control. Subsequent studies, including studies which investigated *CCL3L* copy number individually (Meddows-Taylor et al., 2006, Dolan et al., 2007, Kuhn et al., 2007, Ahuja et al., 2008, Shalekoff et al., 2008) or in combination with *CCL4L* copy numbers (Shostakovich-Koretskaya et al., 2009) also found associations with HIV-1 related outcomes. However, some studies failed to see associations between *CCL3L* and *CCL4L* copy numbers and HIV-1 related outcomes (Shao et al., 2007, Bhattacharya et al., 2009, Urban et al., 2009, Larsen et al., 2012), making the association of *CCL3L* and *CCL4L* copy number variations with HIV-1 infection controversial. The controversy has been in part attributed to technical challenges of accurate *CCL3L* and *CCL4L* copy number determination and the differences in the study cohorts (Shrestha et al., 2010).

Interestingly, in a meta-analysis of nine studies with a total of 2434 HIV-1 infected individuals and 4029 healthy controls, Liu et al. (2010) found low *CCL3L1* copy numbers to be associated with susceptibility to HIV-1. The nine studies in the meta-analyses included six

studies which found association of *CCL3L1* copy numbers with susceptibility to HIV-1 infection and three studies which found the contrary.

Subsequent to the Gonzalez et al. (2005) study, our research group reported that reduced production of CCL3 was associated with higher susceptibility to intrapartum HIV-1 infection, leading to the investigation of the role of *CCL3L1* copy numbers in mother-to-child transmission of HIV-1 in the black South African population (Meddows-Taylor et al., 2006). Meddows-Taylor et al. (2006) found an overrepresentation of low *CCL3L1* copy numbers in HIV-1 intrapartum infected infants compared to those who remained uninfected. Subsequently Kuhn et al. (2007) found that having *CCL3L* copy number greater than the population specific median was associated with reduced risk of perinatal HIV-1 transmission (both intrapartum and *in utero* transmission). Additionally, in a cohort of HIV-1 infected mothers, Shalekoff et al. (2008) found a negative correlation between *CCL3L1* copy numbers and HIV-1 viral load.

Real-time PCR and the paralogue ratio test (PRT) are the main techniques which have been used in *CCL3L* and *CCL4L* copy number determination. Field et al. (2009) compared the performance of real-time PCR to PRT in *CCL3L* copy number determination, and found the PRT assay to be more accurate than real-time PCR in determining high *CCL3L* copy numbers. Droplet digital PCR, a relatively new technology, has also been shown to have higher accuracy and sensitivity compared to real-time PCR in various applications including environmental DNA detection (Doi et al., 2015) microRNA quantification (Hindson et al., 2013), bacterial load test (Kim et al., 2014) and host gene copy number determination (Bharuthram et al., 2014). In the South African population, Bharuthram et al. (2014) demonstrated that ddPCR was a more accurate method for determining high *CCL4L* copy numbers compared to real-time PCR. Therefore, in this study, given that black South Africans have high *CCL3L* and *CCL4L* gene copy numbers, we determined the *CCL3L* and *CCL4L* gene copy numbers using ddPCR.

We successfully adapted a *CCL3L* real-time PCR assay to a ddPCR assay format. For *CCL4L*, we used previously developed *CCL4L* ddPCR assays (Bharuthram et al., 2014). Although the *CCL4L* assays had been previously validated, we validated both the assays by correlating the total *CCL3L* and total *CCL4L* copy numbers with the sum of the *CCL3La* and *CCL3Lb* in addition to the sum of *CCL4La* and *CCL4Lb* copy numbers, respectively. Both assays performed very well with exceptionally strong correlations between *CCL3L*/*CCL4L* copy numbers and the sum of the individual genes (i.e *CCL3La* + *CCL3Lb* or *CCL4La* + *CCL4Lb*) without any discordant results. Thus, the high accuracy demonstrated with the *CCL3L* ddPCR

assays, in combination to the already reported *CCL4L* performance, further confirm the superiority of ddPCR for accurate copy number determination in populations where copy numbers are high.

Prior to comparing the *CCL3L* and *CCL4L* copy numbers between HIV-1 controllers and progressors, we investigated the relationship between the individual *CCL3L* and *CCL4L* gene copy numbers as well as how the two genes relate to each other. As previously seen in Shostakovich-Koretskaya et al. (2009) there was a positive correlation between total *CCL3L* gene copy numbers and *CCL3La* as well as *CCL3Lb* gene copy numbers, and the *CCL3La* and *CCL3Lb* copy numbers also positively correlated to each other. The positive correlation between the *CCL3La* and *CCL3Lb* gene copy numbers suggests a possible negative feedback mechanism between the products of the two genes. Given that *CCL3Lb*, previously thought to be a pseudogene, has 5' exons which are transcribed (Shostakovich-Koretskaya et al., 2009), the mRNA transcripts might have a post-transcriptional regulatory function that controls the quantity of the functional CCL3L chemokine produced.

For *CCL4L*, only total *CCL4L* was positively correlated to *CCL4La* and *CCL4Lb*, whereas *CCL4La* and *CCL4Lb*, unlike *CCL3La* and *CCL3Lb*, showed a negative correlation to each other. This is also in agreement with the findings of Shostakovich-Koretskaya et al. (2009) and Bharuthram (2015). The negative correlation between *CCL4La* and *CCL4Lb* gene copy number is interesting and may be suggestive of a positive feedback mechanism between the products of these two genes. Compared to *CCL3La* and *CCL3Lb*, where individuals seem to always have more copies of *CCL3La* than *CCL3Lb*, with *CCL4La* and *CCL4Lb,* individuals can either have *CCL4La* copy numbers greater than *CCL4Lb* (*CCL4La* > *CCL4Lb*), or less than *CCL4Lb* (*CCL4La* < *CCL4Lb*), or equal to *CCL4Lb* (*CCL4La* = *CCL4Lb*). In our study population, more individuals had *CCL4La* > *CCL4Lb* copies with the least number of individuals having the same number of *CCL4La* and *CCL4Lb* copies.

We were also interested in how the *CCL3L* and *CCL4L* genes related to each other, and as previously seen (Shostakovich-Koretskaya et al., 2009), the *CCL3L* and the *CCL3L* genes were positively correlated to each other. However, interestingly each individual had the same number of *CCL3La* copies as the *CCL4L* copies (i.e total of *CCL4La* and *CCL4Lb*), with the exception of one individual (repeated three times with the same result). This relationship between the *CCL3La* and *CCL4L* copies has been previously demonstrated in Caucasians and East Africans, in studies using the PRT designed to detect both the *CCL3L* and *CCL4L* gene copy numbers (Walker et al., 2009, Nordang et al., 2012). What is confusing is that the studies using PRT method refer to the total *CCL4L* genes (i.e *CCL4L1* + *CCL4L2*) as *CCL4L1*

(which is normally used as the acronym for *CCL4La*), hence they describe this relationship as being between *CCL3La* (*CCL3L1*) and *CCL4La* (*CCL4L1*), instead of *CCL3La* and *CCL4L,* further highlighting the confusion created in the literature with the use of variable nomenclature. The relationship between the *CCL3La* and *CCL4L* copy numbers suggests that the *CCL4L* genes were co-duplicated with the functional *CCL3L* gene (*CCL3La*), instead of the total *CCL3L* region. Therefore, in the South African black population, the *CCL3L1* (*CCL3La*) copy numbers cannot be used as a proxy for *CCL4L1* (*CCL4La*), because even though *CCL3La* is equal to *CCL4L*, it does not give any information about the individual *CCL4L* genes (i.e *CCL4La* and *CCL4Lb*). In addition, it is interesting that despite the *CCL3La* copy numbers being equivalent to *CCL4L* copy numbers, in Caucasians over 60% of individuals have the same number of *CCL3La* and *CCL4La* copies (Townson et al., 2002), likely due to low copy numbers in Caucasians. Additionally, given that the *CCL3L* and *CCL4L* genes are found in a region of segmental duplications, it is likely that they have been duplicated together as a block through non-allelic homologous recombination.

Having determined the copy numbers for the *CCL3L* and *CCL4L* genes using ddPCR and having validated that the copy numbers for our HICs and progressors were accurate, we then sought to investigate the role of the CCL3 and CCL4 chemokines in natural control of HIV-1 by comparing the copy numbers of the *CCL3L* and *CCL4L* genes between HICs and progressors as well as between HIC subgroups and progressors. We used a number of different strategies to compare the copy numbers, including looking at the genes individually as well as in combination. Copy numbers were analysed as both continuous variables and when stratified around the population specific median. Two stratification strategies were used, firstly we stratified the individuals according to copy numbers ≥ population median and < population median as done in previous studies (Gonzalez et al., 2005, Shostakovich-Koretskaya et al., 2009, Liu et al., 2010), and we also stratified the individuals according to copy numbers ≤ population median and > population median as previously done by Liu et al. (2010) and Nakajima et al. (2007). Since the copy numbers for *CCL3La* and *CCL4L* were identical, henceforth the *CCL3La* and *CCL4L* copy number will be referred to as *CCL3La*/*CCL4L*.

Although comparison of the *CCL3L* and *CCL4L* copy numbers as continuous variables did not show any significant differences between HICs, HIC subgroups and progressors, with only VCs having a trend of higher *CCL3La*/*CCL4L* copy numbers compared to progressors, stratification (around the median) revealed significant differences. Interestingly, stratification according to ≥ population median and < population median, as has been mostly done in

previous studies (Gonzalez et al., 2005, Shostakovich-Koretskaya et al., 2009, Liu et al., 2010), did not show significant differences and only strong trends of VCs and <400 HICs having higher representation of *CCL3L* ≥ population median were seen when compared to progressors. However, stratification according to ≤ population median and > population median yielded significant differences again in VCs and <400 HICs compared to progressors, with both having higher representation of *CCL3La*/*CCL4L* > population median. There was also a strong trend of higher representation of *CCL3La*/*CCL4L* > population median in HICs compared to progressors.

These findings suggest that having *CCL3La*/*CCL4L* copy numbers greater than the population specific median is beneficial in terms of HIV-1 control in HICs, showing more strongly in VCs and <400 HICs which are in fact overlapping groups. Nevertheless, due to *CCL3La* copy numbers being equivalent to *CCL4L* copy numbers, the protective effect cannot be specifically attributed to CCL3La or to CCL4L. Hence, it would be interesting to determine whether either or both of the two chemokines (CCL3 or CCL4) are key players. Although both CCL3 and CCL4 are natural ligands of the chemokine receptor CCR5, immunologically they have quite distinct and different roles (Lillard et al., 2003). The fact that only total *CCL4L* copy number (equivalent to *CCL3La*) was associated with HIV-1 control while *CCL4La* and *CCL4Lb* copy numbers greater than their respective population specific medians did not associate with HIV-1 control, suggests that the association is more likely attributed to *CCL3La* possibly without any functional implications for CCL4. Hence, CCL3La and CCL4 expression studies may be helpful in determining the importance of these chemokines in the context of HIV-1 control.

To try and get a better understanding of the role of the CCL3 and CCL4 associated copy numbers, we investigated the combined effect of the *CCL3L* and *CCL4L* gene copy numbers, stratified by copy numbers > population median or ≤ population median. We found that when compared to progressors, <400 HICs had a significant overrepresentation of individuals with *CCL3L* copy number > population medians and/or *CCL4L* > population median but the significance was weaker than that of *CCL3La*/*CCL4L* copy number alone, suggesting a negative effect of *CCL3Lb*. Additionally, a significant overrepresentation of individuals with *CCL3La* copy numbers > population median and/or *CCL4La* copy numbers > population median was seen in HICs, VCs and <400 HICs when compared to progressors. Although the significance was not altered in the <400 HICs vs. progressors comparison, in HICs and VCs compared to progressors the significance was stronger than that of *CCL3La*/*CCL4L* copy numbers alone, suggesting that having a high *CCL3La* and/or high *CCL4La* is more protective

compared to *CCL3La*/*CCL4L* copy numbers alone. This result suggests that the CCL4La isoform may in addition to the CCL3La be contributing to the protective effect. What is also important to note is that both *CCL3La* and *CCL4La* are the classical "functional" genes for the *CCL3L* and *CCL4L* gene complex, respectively, and that no significant associations were seen when comparing the *CCL3Lb* and *CCL4Lb* genes individually or in combination. However, given their potential roles in regulation of the functional genes, we attempted to determine their contributions by interrogating and analysing the data in a number of novel ways.

Firstly, since *CCL3L* and *CCL4L* copy numbers occur in different combinations in different individuals, the combination of *CCL3La* and *CCL3Lb* copy numbers as well as *CCL4La* and *CCL4Lb* copy numbers show inter-individual variability. Thus, particular gene copy number combinations may be informative with respect to the contributions of these genes to the protein produced and ultimately their role in HIV-1 control. To our knowledge, analyses of different *CCL3L* and *CCL4L* copy number patterns in natural control of HIV-1 infection has not been considered. Although some *CCL3L* and *CCL4L* copy number patterns, individually or in combination, showed trends of higher representation in certain controller groups, the only significant findings included an overrepresentation of a *CCL3L* 651 pattern (gPID) in ECs and <400 HICs when compared to progressors. Although not significant, an underrepresentation of a *CCL3L* 642 gPID was seen in HICs when compared to progressors. With respect to *CCL4L*, the 633 gPID was also overrepresented in HICs and >400 HICs compared to progressors. Additionally, a significant overrepresentation of a combined *CCL3L* and *CCL4L* gPID 550541 was seen in VCs compared to progressors.

The *CCL3L* 651 gPID has a *CCL3La* copy number of 5 (i.e greater than the population median). It is therefore interesting that although there is a significant overrepresentation of the 651 gPID in both <400 HICs and ECs, in <400 HICs but not ECs as previously mentioned, there is an overrepresentation of individuals with *CCL3La* copy numbers > population median, therefore the significant association of the 651 gPID seen in ECs may be a consequence of small sample number (N=11) or may be indicative of this particular combination of 5 copies of *CCL3La* and 1 copy of *CCL3Lb* being associated with elite control. An interesting observation is that a combination of a *CCL3La* copy number > population median and a *CCL3Lb* copy number equal to the population median (i.e 651 gPID) is protective, while a combination of a *CCL3La* copy number equal to the population median and a *CCL3Lb* copy number greater than the population median (i.e 642 gPID) seems to be potentially deleterious. Although the associations seen with respect to *CCL3L* and *CCL4L*

patterns individually or in combination can be attributed to these patterns having *CCL3La*/*CCL4L* copy numbers greater than the median as already shown, it should be noted that some patterns that also exhibit *CCL3La*/*CCL4L* copies > population median did not show similar associations. It would thus be interesting to determine how the predominant gene combinations (patterns) affect protein production or immunological function. Additionally, there is a possibility that some of the *CCL3L* and/or *CCL4L* copy number patterns are in linkage disequilibrium with other protective variants yet to be identified.

Secondly, we attempted to analyse the data in a manner that may take into account the "effect" of the *CCL3Lb* and *CCL4Lb* genes on the classical functional genes (i.e *CCL3La* and *CCL4La*) by looking at the relationship between *CCL3La* and *CCL3Lb* or *CCL4La* and *CCL4Lb* genes, and comparing between HICs, HIC subgroups and progressors. Having shown that there is a positive correlation between the *CCL3La* and *CCL3Lb* gene copy numbers, we made the assumption that a single copy of *CCL3Lb* has a counteracting effect on a single copy of *CCL3La*, and thus subtracted *CCL3Lb* from *CCL3La* copies for each individual and compared the differences between the relevant groups. However, it is important to note that not all copies of *CCL3La* or *CCL3Lb* may be functional, as variation within the individual copies can result in some copies being pseudogenes or incomplete genes. When looking at the differences, we found that a *CCL3La* and *CCL3Lb* copy number difference of 2 was significantly underrepresented in HICs compared to progressors and a difference of 5 was significantly overrepresented in VCs compared to progressors. A trend of overrepresentation of a *CCL3La* and *CCL3Lb* difference of 4 was also noted in HVL LTNPs compared to progressors. These results suggest that a low *CCL3La* and *CCL3Lb* copy number difference is deleterious, whereas a higher difference is protective. If CCL3Lb does negatively regulate the amount of CCL3La protein produced, then this result suggests that more CCL3La (and consequently less CCL3Lb) is advantageous in the context of viral control. The relationship between *CCL4La* and *CCL4Lb* seems to be rather different from that of *CCL3La* and *CCL3Lb,* with *CCL4La* and *CCL4Lb* copies being negatively correlated. The distribution of individuals with *CCL4La* > *CCL4Lb*, *CCL4La* = *CCL4Lb* and *CCL4La* < *CCL4Lb* did not show any significant differences between the groups compared, possibly suggesting that the relationship between these two genes is not impacting on HIV-1 control in this study, or than the effect of CCL4L is less than that of CCL3L and that the sample numbers may be too small to detect the effect using this analysis.

With regards to the effect of copy number on HIV-1 markers of disease progression (i.e viral loads and $CD4^+$ T cell count), our results are consistent with the studies which found an

association between low viral loads with high *CCL3L1* copy numbers (Gonzalez et al., 2005, Shalekoff et al., 2008, Shostakovich-Koretskaya et al., 2009). Additionally, our results are consistent with studies which did not find association between *CCL3L* copies numbers and $CD4^+$ T counts (Shao et al., 2007, Shalekoff et al., 2008, Bhattacharya et al., 2009, Urban et al., 2009, Aklillu et al., 2013). Overall, these results support the role of *CCL3La*/*CCL4L* copy numbers > population median in HIV-1 control in individuals with low vireamia, as it has been shown that high *CCL3La*/*CCL4L* copy numbers are associated with low viral loads but not with CD4+ T cell count. This may also be why we did not find any significant differences in *CCL3L* and *CCL4L* copy numbers when HVL LTNPs were compared to progressors, however, this needs to be tested in larger sample numbers. Although some studies (Nakajima et al., 2007, Urban et al., 2009) did not find significant differences between *CCL3L* and *CCL4L* copy numbers when LTNPs were compared to progressors, these results cannot be directly compared to our results, as our group of LTNPs were high viral load LTNPs which are not necessarily the same as the LTNPs described in these studies.

The exact mechanism by which *CCL3L* and *CCL4L* copy number variations affect different HIV-1 related outcomes, including natural control of HIV-1 is relatively unknown. It has however been suggested that it is through its effect on protein production, with more chemokine being associated with protection. A number of studies have found a protective role of increased CCL3 and/or CCL4 production with different HIV-1 related outcomes including natural control of HIV-1 (Cocchi et al., 2000, Walker et al., 2015) and reduced susceptibility to HIV-1 infection (Meddows-Taylor et al., 2006).

Copy number variation has been shown to have a dosage effect on protein expression in most genes with variable copy numbers, however in the CCL3 and CCL4 encoding genes, the relationship seems complex. Importantly, immunological assays currently used to quantify CCL3 and CCL4, cannot distinguish between the CCL3 and CCL3L isoform as well as between CCL4 and CCL4L isoforms. As such, the exact contributions of each of the genes on the total CCL3 or CCL4 chemokines produced are largely unknown and the *CCL3L* or *CCL4L* copy numbers have not been directly correlated to the *CCL3L* or *CCL4L* production, respectively.

Townson et al. (2002) however, did demonstrate that high *CCL3L* copy numbers are correlated with increased CCL3 production in LPS-stimulated monocytes. Some subsequent studies (Gonzalez et al., 2005, Meddows-Taylor et al., 2006) also found an association between high *CCL3L* copy numbers with increased protein production, however other studies have not been able to see similar associations (Carpenter et al., 2012, Larsen et al., 2012). In

our research group, Picton et al. (2013) did not find a correlation between *CCL3L* and *CCL3L1* (*CCL3La*) copy numbers and CCL3 production in both healthy HIV-1 uninfected black and Caucasians South Africans. Given that the *CCL3L*, *CCL3La* and *CCL3Lb* copy numbers are significantly different between black and Caucasians South Africans, this study did not find significant differences between the two population groups in phytohaemagglutinin (PHA) stimulated CCL3 production in PBMCs. These results may suggest that each population irrespective of the copy number, produces on average the same amount of chemokine/s and that the retention of the "extra" copies of functional *CCL3L* or *CCL4L* in certain population groups may be compensating for their corresponding allelic variants (2 copies pdg) (i.e *CCL3*/*CCL4*) producing less protein. To test this hypothesis, it would thus be interesting to see if ethnically distinct population groups, for example Caucasians and African individuals differ in their levels of *CCL3* or *CCL4* specific mRNA, and if the 2 copies pdg genes thus differentially contribute to the total chemokine/s produced in the two population groups.

Although the direct relationship of *CCL3L* copy numbers with protein production is unclear, Townson et al. (2002) demonstrated that in LPS-stimulated monocytes, high *CCL3L1* (*CCL3La*) copy number significantly correlated with an increased ratio of *CCL3L1* to *CCL3* mRNA. Subsequent studies also found a correlation between *CCL3L1* copy numbers with *CCL3L1* mRNA expression levels, with mRNA levels increasing with higher copy number (Urban et al., 2009, Carpenter et al., 2012). This relationship suggests a possible post-transcriptional regulation of CCL3 since the effect of the copy number does not seem to be carried through to the protein level (Carpenter et al., 2012), alternatively, the ratio of the CCL3L isoform to the CCL3 isoform may be higher in individuals with high *CCL3L1* copy numbers compared to individuals with low copy numbers. Additionally, Carpenter et al. (2012) demonstrated that there was higher *CCL3* mRNA expression compared to *CCL3L* mRNA expression, suggesting that the CCL3 isoform is likely the major contributor of the CCL3 product, hence the higher production of the CCL3 isoform may lead to the assumption that high *CCL3L* copy numbers do not result in increased CCL3 expression if the CCL3L isoform is only a minor contributor to the pool of total protein. Compared to CCL3, CCL4 and CCL5, CCL3L is the most potent ligand of the CCR5 receptor and has the highest antiviral activity, making it the most potent chemokine in inhibiting replication of R5 strains of HIV-1 in macrophages *in vitro* (Menten et al., 1999, Nibbs et al., 1999, Aquaro et al., 2001). In fact, CCL3L has been shown to be 30-fold more efficient in inhibiting HIV-1 compared with CCL3, and lower concentrations of CCL3L than CCL3 are required to initiate antiviral activity (Menten et al., 1999, Nibbs et al., 1999, Aquaro et al., 2001), suggesting that

even though CCL3L might be a minor contributor to the total CCL3 produced, small changes in its concentrations might have big functional consequences, in terms of inhibiting HIV-1 replication.

Additionally, *CCL3L1* copy numbers have been inversely correlated with CCR5 expression levels (Gonzalez et al., 2005), with high *CCL3L1* copy numbers being associated with low CCR5 expression levels. *In vitro*, low CCR5 expression levels have been associated with reduced replication of R5 tropic viruses (Heredia et al., 2007) and in HIV-1 infected individuals, it has been shown that CCR5 expression is positively correlated with viral loads thus, low CCR5 expression is associated with low viral loads (Reynes et al., 2000). Although *CCL3L1* copy numbers are not the sole determinants of CCR5 expression, these results suggest that the products corresponding to *CCL3L1* copies may be exerting an effect on HIV-1 control through their effect on CCR5 expression.

Whereas the effect of *CCL3L* copy numbers on CCL3 production has been extensively investigated, the effect of *CCL4L* copy numbers on CCL4 production has not been investigated to the same extent. Townson et al. (2002) found no direct correlation between *CCL4L* copy number with protein production or the ratio of *CCL4L* to *CCL4* mRNA. Interestingly though, Colobran et al. (2005) found that individuals in possession of only *CCL4L1* (*CCL4La*) had higher CCL4 protein production and mRNA expression compared to individuals in possession of only *CCL4L2* (*CCL4Lb*). Additionally, having a single copy of *CCL4L* has also been associated with consistently lower expression of CCL4L compared to having more copies of *CCL4L* (Townson et al., 2002).

Although CCL3 and CCL4 are ligands of the CCR5 receptor which serves as the major coreceptor for HIV-1 entry into permissive cells, they mediate chemotactic functions and are involved in immune responses via CCR5 as well as other chemokine receptors. The *CCL3L1* copy numbers in combination with *CCR5* genotypes have been found to have an effect on cell mediated immunity in HIV-1 infected and uninfected individuals (Dolan et al., 2007). Dolan et al. (2007) found that in HIV-1 infected individuals with high genetic risk factors (low *CCL3L1* copy number and detrimental *CCR5* genotypes), there was lower cell mediated immunity as assessed by delayed-type hypersensitivity responses, indicating that in addition to *CCL3L1* and *CCR5* genotypes affecting HIV-1 pathogenesis through entry dependent mechanisms, they also affect pathogenesis through cell mediated immunity. In our research group, Shalekoff et al. (2008) found a significant association between *CCL3L1* copy numbers $\geq$ population specific median and increased magnitude of Gag specific $CD4^+$ T cell responses,

suggesting that *CCL3L1* copy numbers may be affecting HIV-1 related outcomes through HIV-1-specific T cell responses.

Isoforms encoded by the allelic (*CCL3*) and the non-allelic genes (*CCL3L*) cannot be differentiated using currently available immunological assays, however it is important to consider the contribution of the variation within *CCL3* to HIV-1 control. A previous study in our research group looking at mother-to-child transmission of HIV-1, found two *CCL3* haplotypes (each comprised of seven SNPs), namely Hap-A1 and Hap-A3 to be associated with protection to HIV-1 infection *in utero* and increased intrapartum transmission of HIV-1, respectively (Paximadis et al., 2013). Hap-A1 and Hap-A3 have two SNPs in common, which are in linkage disequilibrium making up a 2 SNP haplotype (Hap-2SNP). Paximadis et al. (2013) also found that having a high *CCL3L* copy number and/or Hap-A1 was more strongly associated with protection against HIV-1 infection acquired *in utero* than high copy number alone or Hap-A1 alone, thereby suggesting an additive effect.

Given that Hap-A1, Hap-A2 and Hap-2SNP have been shown to play a role in susceptibility to HIV-1 infection in the context of mother-to-child transmission, we therefore investigated the role of these *CCL3* haplotypes in natural control of HIV-1, individually and in combination with *CCL3La* copy numbers (already shown to be protective). Having access to the data available on the 1000 genomes project allowed us to extend our investigation regarding the genetic structure of the *CCL3* gene further, prior to comparisons between HICs, HIC subgroups and progressors. We investigated the presence and prevalence of the three *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2SNP) in the Yoruba subpopulation, European population and the total African population from the 1000 genomes project database. Linkage disequilibrium analysis showed that in addition to being present in the South African back population, Hap-A1 and Hap-A3 were present in the Yoruba subpopulation and African reference population, but not in Europeans. In addition, Hap-2SNP was also shown to be present in the Yoruba subpopulation, total African population and in the European population. Although no significant differences were seen when Hap-A3 and Hap-2SNP frequencies were compared between the South African black populations, total African population, Yoruba subpopulation and European population (for Hap-2SNP), a significant overrepresentation of Hap-A1 heterozygosity was seen in the South African black population compared to the African population.

Comparison of Hap-A1, Hap-A3 and Hap-2SNP allelic and genotypic frequencies between HICs, HIC subgroups and progressors, revealed no significantly over or underrepresented alleles or genotypes. We also had access to the allelic and genotypic frequencies of the three

CCL3 haplotypes for healthy HIV-1 uninfected controls from the black South African population. When comparing the allelic and genotypic frequencies of Hap-A1, Hap-A3 and Hap-2SNP between HICs, HIC subgroups and progressors to healthy controls, again no significantly over or underrepresented genotypes or alleles were seen. However, a strong trend of higher representation of Hap-A1 heterozygosity was seen in >400 HICs when compared to both healthy controls and progressors, suggesting a possible role of Hap-A1 heterozygosity in natural control of HIV-1 infection in controllers with higher viraemia.

When looking at the combined role of the *CCL3* haplotypes (Hap-A1, Hap-A3 and Hap-2SNP) and *CCL3La* copy numbers, although some significant differences were seen for having high (> population median) *CCL3La* and/or Hap-A1, or high *CCL3La* and Hap-A3 or high *CCL3La* and Hap-2SNP, the associations for *CCL3La* and/or Hap-A1 and high *CCL3La* and Hap-2SNP were not more significant than that of high *CCL3La*/*CCL4L* copy number alone, suggesting that it is the copy number that is responsible for the associations seen. However, the significance seen for having high *CCL3La* and Hap-A3 was slightly higher than that of high *CCL3La*/*CCL4L* copy number alone when VCs were compared to progressors. However, given that Hap-A3 alone did not show significant associations in VCs, it is likely that this result is predominantly due to the effect of *CCL3La*/*CCL4L*. These findings suggest that the combined role of the *CCL3* haplotypes and *CCL3La* copy numbers in the context of HIV-1 control is different from their role in HIV-1 susceptibility.

Although the role of these three haplotypes in HIV-1 related outcomes has not been investigated in other studies besides the study by Paximadis et al. (2013), some of the SNPs making up these haplotypes have been investigated in terms of their role in susceptibility to HIV-1 infection and the rate of progression to AIDS (Gonzalez et al., 2001, Modi et al., 2006, da Silva et al., 2016). However, the combined effect of these SNPs and *CCL3L* copy numbers has not been investigated in other studies. Hap-A1 and Hap-A3 are comprised of SNPs which are found in both coding and noncoding regions and mostly likely have an effect on *CCL3* gene expression.

In conclusion, we have shown that for the majority of individuals, *CCL3La* copy numbers are identical to the *CCL4L* copy numbers in the black South African population. In agreement with earlier studies which found high *CCL3L* copy numbers to be protective against disease progression, our results demonstrated that having a *CCL3La*/*CCL4L* copy number greater than the population median or having high *CCL3La* and/or high *CCL4La* is associated with natural control of HIV-1. However, it would be beneficial for future work to be conducted with larger sample sizes, especially of ECs and HVL LTNPs to confirm these findings. We

have also demonstrated that the different combinations of *CCL3L* and *CCL4L* gene copy numbers (copy number patterns), individually or in combination display distinct patterns that associate with HIV-1 control. Nonetheless, given the large numbers of identified copy number patterns for both *CCL3L* and *CCL4L*, this investigative approach needs to be tested in much larger numbers of HIV-1 infected individuals, as the importance of some low prevalent patterns may be overlooked. Additionally, the difference between *CCL3La* and *CCL3Lb* copy numbers also seems to be important in the context of HIV-1 control. With regards to *CCL3* haplotypes, although Hap-A1 showed a moderate effect on natural control in >400 HICs, the role of these haplotypes in the context of HIV-1 control seems to be different from their role in HIV-1 susceptibility.

In addition to small sample sizes of ECs and HVL LTNPs being a limitation of this study, it is important to note that copy number variation only represents one component contributing to variation in the *CCL3L* and *CCL4L* genes and genetic variations within the individual copies (SNPs and indels) also likely affect protein production and chemokine function. Hence studying the additional variations in the *CCL3L* and *CCL4L* genes can help further elucidate the associations of these genes with natural control of HIV-1. Furthermore, studies that incorporate investigations of the post translational modification of the CCL3 and CCL4 chemokines and their binding to the CCR5 receptor in the context of HIV-1 control can also be beneficial in trying to elucidate these complex interactions. For further characterization of the CCL3 and CCL4 chemokine encoding gene cluster, studying the different isoforms of CCL3 and CCL4 and quantification of the transcripts can also be beneficial. Overall, our findings suggest that CCL3 and CCL4 chemokines play a role in natural control of HIV-1 infection in the black South African population and can be studied in combination with other protective *CCR5* genotypes and gene targets when investigating the development of new vaccines and anti-HIV therapies.

References

Ahuja, S. K., Kulkarni, H., Catano, G., Agan, B. K., Camargo, J. F., He, W., O'Connell, R. J., Marconi, V. C., Delmar, J., Eron, J., Clark, R. A., Frost, S., Martin, J., Ahuja, S. S., Deeks, S. G., Little, S., Richman, D., Hecht, F. M. & Dolan, M. J. 2008. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. *Nature Medicine,* 14**,** 413-420.

Aklillu, E., Odenthal-Hesse, L., Bowdrey, J., Habtewold, A., Ngaimisi, E., Yimer, G., Amogne, W., Mugusi, S., Minzi, O., Makonnen, E., Janabi, M., Mugusi, F., Aderaye, G., Hardwick, R., Fu, B., Viskaduraki, M., Yang, F. & Hollox, E. J. 2013. CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *BMC Infectious Diseases,* 13**,** 536.

An, P. & Winkler, C. A. 2010. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends in Genetics* 26**,** 119-131.

Andrade, A., Bailey, J. R., Xu, J., Philp, F. H., Quinn, T. C., Williams, T. M., Ray, S. C., Thomas, D. L. & Blankson, J. N. 2008. CD4+ T cell depletion in an untreated HIV type 1-infected human leukocyte antigen-B*5801-positive patient with an undetectable viral load. *Clinical Infectious Diseases,* 46**,** e78-82.

Aquaro, S., Menten, P., Struyf, S., Proost, P., Van Damme, J., De Clercq, E. & Schols, D. 2001. The LD78beta isoform of MIP-1alpha is the most potent CC-chemokine in inhibiting CCR5-dependent human immunodeficiency virus type 1 replication in human macrophages. *Journal of Virology,* 75**,** 4402-4406.

Bailey, J. R., Williams, T. M., Siliciano, R. F. & Blankson, J. N. 2006. Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. *Journal of Experimental Medicine,* 203**,** 1357-1369.

Barre-Sinoussi, F., Ross, A. L. & Delfraissy, J. F. 2013. Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology,* 11**,** 877-883.

Berger, E. A., Doms, R. W., Fenyo, E. M., Korber, B. T., Littman, D. R., Moore, J. P., Sattentau, Q. J., Schuitemaker, H., Sodroski, J. & Weiss, R. A. 1998. A new classification for HIV-1. *Nature,* 391**,** 240.

Berkman, N., John, M., Roesems, G., Jose, P. J., Barnes, P. J. & Chung, K. F. 1995a. Inhibition of macrophage inflammatory protein-1 alpha expression by IL-10. Differential sensitivities in human blood monocytes and alveolar macrophages. *Journal of Immunology,* 155**,** 4412-4418.

Berkman, N., Jose, P. J., Williams, T. J., Schall, T. J., Barnes, P. J. & Chung, K. F. 1995b. Corticosteroid inhibition of macrophage inflammatory protein-1 alpha in human monocytes and alveolar macrophages. *American Journal of Physiology,* 269**,** L443-452.

Bharuthram, A., Paximadis, M., Picton, A. C. & Tiemessen, C. T. 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.

Bharuthram, A. (2015). The Role of CCL4/Mip1-β in HIV infection. MSc dissertation.
University of the Witwatersrand, Johannesburg

Bhattacharya, T., Stanton, J., Kim, E. Y., Kunstman, K. J., Phair, J. P., Jacobson, L. P. & Wolinsky, S. M. 2009. CCL3L1 and HIV/AIDS susceptibility. *Nature Medicine,* 15**,** 1112-1115.

Burns, J. C., Shimizu, C., Gonzalez, E., Kulkarni, H., Patel, S., Shike, H., Sundel, R. S., Newburger, J. W. & Ahuja, S. K. 2005. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. *Journal of Infectious Diseases,* 192**,** 344-349.

Buseyne, F., Janvier, G., Teglas, J., Ivanoff, S., Burgard, M., Bui, E., Mayaux, M.-J., Blanche, S., Rouzioux, C. & Riviere, Y. 1998. Impact of Heterozygosity for the Chemokine Receptor CCR5 32-bp—Deleted Allele on Plasma Virus Load and CD4 T

Lymphocytes in Perinatally Human Immunodeficiency Virus—Infected Children at 8 Years of Age. *Journal of Infectious Diseases,* 178**,** 1019-1023.

Carpenter, D., McIntosh, R. S., Pleass, R. J. & Armour, J. A. 2012. Functional effects of CCL3L1 copy number. *Genes & Immunity,* 13**,** 374-379.

Carrington, M. & Alter, G. 2012. Innate immune control of HIV. *Cold Spring Harbor Perspectives in Medicine,* 2**,** a007070.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K. & O'Brien, S. J. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science,* 283**,** 1748-1752.

Chan, D. C. & Kim, P. S. 1998. HIV entry and its inhibition. *Cell,* 93**,** 681-684.

Choi, A. I., Shlipak, M. G., Hunt, P. W., Martin, J. N. & Deeks, S. G. 2009. HIV-infected persons continue to lose kidney function despite successful antiretroviral therapy. *AIDS,* 23**,** 2143-2149.

Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C. & 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**,** 1811-1815.

Cocchi, F., DeVico, A. L., Yarchoan, R., Redfield, R., Cleghorn, F., Blattner, W. A., Garzino-Demo, A., Colombini-Hatch, S., Margolis, D. & Gallo, R. C. 2000. Higher macrophage inflammatory protein (MIP)-1alpha and MIP-1beta levels from CD8+ T cells are associated with asymptomatic HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 13812-13817.

Cockerham, L. R. & Hatano, H. 2015. Elite control of HIV: is this the right model for a functional cure? *Trends in Microbiology,* 23**,** 71-75.

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

Colobran, R., Adreani, P., Ashhab, Y., Llano, A., Este, J. A., 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**,** 5655-5664.

Colobran, R., Casamitjana, N., Roman, A., Faner, R., Pedrosa, E., Arostegui, J. I., Pujol-Borrell, R., Juan, M. & Palou, E. 2009. Copy number variation in the CCL4L gene is associated with susceptibility to acute rejection in lung transplantation. *Genes & Immunity,* 10**,** 254-259.

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 & Immunity,* 9**,** 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. *Clinical & Experimental Immunology,* 162**,** 41-52.

Currier, J. S., Lundgren, J. D., Carr, A., Klein, D., Sabin, C. A., Sax, P. E., Schouten, J. T., Smieja, M. & Working, G. 2008. Epidemiological evidence for cardiovascular disease in HIV-infected patients and relationship to highly active antiretroviral therapy. *Circulation,* 118**,** e29-35.

da Silva, R., Coelho, A., Arraes, L., Brandão, L., Guimarães, R. & Crovella, S. 2016. Chemokines SNPs in HIV-1+ Patients and Healthy Controls from Northeast Brazil: Association with Protection against HIV-1 Infection. *Current HIV research,* 14**,** 340-345.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R. & O'Brien, S. J. 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**,** 1856-1862.

Deeks, S. G. & Walker, B. D. 2007. Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy. *Immunity,* 27**,** 406-416.

Degenhardt, J. D., de Candia, P., Chabot, A., Schwartz, S., Henderson, L., Ling, B., Hunter, M., Jiang, Z., Palermo, R. E., Katze, M., Eichler, E. E., Ventura, M., Rogers, J., Marx, P., Gilad, Y. & Bustamante, C. D. 2009. Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). *PLoS Genetics,* 5**,** e1000346.

Doi, H., Takahara, T., Minamoto, T., Matsuhashi, S., Uchii, K. & Yamanaka, H. 2015. Droplet digital polymerase chain reaction (PCR) outperforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environmental Science and Technology,* 49**,** 5601-5608.

Dolan, M. J., Kulkarni, H., Camargo, J. F., He, W., Smith, A., Anaya, J. M., Miura, T., Hecht, F. M., Mamtani, M., Pereyra, F., Marconi, V., Mangano, A., Sen, L., Bologna, R., Clark, R. A., Anderson, S. A., Delmar, J., O'Connell, R. J., Lloyd, A., Martin, J., Ahuja, S. S., Agan, B. K., Walker, B. D., Deeks, S. G. & Ahuja, S. K. 2007. CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. *Nature Immunology,* 8**,** 1324-1336.

Douek, D. C., Brenchley, J. M., Betts, M. R., Ambrozak, D. R., Hill, B. J., Okamoto, Y., Casazza, J. P., Kuruppu, J., Kunstman, K., Wolinsky, S., Grossman, Z., Dybul, M., Oxenius, A., Price, D. A., Connors, M. & Koup, R. A. 2002. HIV preferentially infects HIV-specific CD4+ T cells. *Nature,* 417**,** 95-98.

Eichler, E. E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends in Genetics,* 17**,** 661-669.

Fellermann, K., Stange, D. E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C. L., Reinisch, W., Teml, A., Schwab, M., Lichter, P., Radlwimmer, B. & Stange, E. F. 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *American Journal of Human Genetics,* 79**,** 439-448.

Field, S. F., Howson, J. M., Maier, L. M., Walker, S., Walker, N. M., Smyth, D. J., Armour, J. A., Clayton, D. G. & Todd, J. A. 2009. Experimental aspects of copy number variant assays at CCL3L1. *Nature Medicine,* 15**,** 1115-1117.

Frost, S. D. W., Martinez-Picado, J., Ruiz, L., Clotet, B. & Leigh Brown, A. J. 2002. Viral Dynamics during Structured Treatment Interruptions of Chronic Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology,* 76**,** 968-979.

Gilbert, P. B., McKeague, I. W., Eisen, G., Mullins, C., Guéye-NDiaye, A., Mboup, S. & Kanki, P. J. 2003. Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. *Statistics in medicine,* 22**,** 573-593.

Gonzalez, E., Bamshad, M., Sato, N., Mummidi, S., Dhanda, R., Catano, G., Cabrera, S., McBride, M., Cao, X. H., Merrill, G., O'Connell, P., Bowden, D. W., Freedman, B. I., Anderson, S. A., Walter, E. A., Evans, J. S., Stephan, K. T., Clark, R. A., Tyagi, S., Ahuja, S. S., Dolan, M. J. & Ahuja, S. K. 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**,** 12004-12009.

Gonzalez, E., Dhanda, R., Bamshad, M., Mummidi, S., Geevarghese, R., Catano, G., Anderson, S. A., Walter, E. A., Stephan, K. T., Hammer, M. F., Mangano, A., Sen, L., Clark, R. A., Ahuja, S. S., Dolan, M. J. & Ahuja, S. K. 2001. Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 5199-5204.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R. J., Freedman, B. I., Quinones, M. P., Bamshad, M. J., Murthy, K. K., Rovin, B. H., Bradley, W., Clark, R. A., Anderson, S. A., O'Connell R, J., Agan, B. K., Ahuja, S. S., Bologna, R., Sen, L., Dolan, M. J. & Ahuja, S. K. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science,* 307**,** 1434-1440.

Gornalusse, G., Mummidi, S., He, W., Silvestri, G., Bamshad, M. & Ahuja, S. K. 2009. CCL3L Copy number variation and the co-evolution of primate and viral genomes. *PLoS Genetics,* 5**,** e1000359.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A. & 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**,** 1425-1431.

Grunhage, F., Nattermann, J., Gressner, O. A., Wasmuth, H. E., Hellerbrand, C., Sauerbruch, T., Spengler, U. & Lammert, F. 2010. Lower copy numbers of the chemokine CCL3L1 gene in patients with chronic hepatitis C. *Journal of Hepatology,* 52**,** 153-159.

Guan, E., Wang, J., Roderiquez, G. & Norcross, M. A. 2002. Natural truncation of the chemokine MIP-1 beta /CCL4 affects receptor specificity but not anti-HIV-1 activity. *Journal of Biological Chemistry,* 277**,** 32348-32352.

Gurdasani, D., Iles, L., Dillon, D. G., Young, E. H., Olson, A. D., Naranbhai, V., Fidler, S., Gkrania-Klotsas, E., Post, F. A., Kellam, P., Porter, K. & Sandhu, M. S. 2014. A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS,* 28**,** 149-162.

Heredia, A., Gilliam, B., DeVico, A., Le, N., Bamba, D., Flinko, R., Lewis, G., Gallo, R. C. & Redfield, R. R. 2007. CCR5 density levels on primary CD4 T cells impact the replication and Enfuvirtide susceptibility of R5 HIV-1. *AIDS,* 21**,** 1317-1322.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L. & Tewari, M. 2013. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nature methods,* 10**,** 1003-1005.

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

Hollox, E. J., Huffmeier, U., Zeeuwen, P. L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P. C., Traupe, H., de Jongh, G., den Heijer, M., Reis, A., Armour, J. A. & Schalkwijk, J. 2008. Psoriasis is associated with increased beta-defensin genomic copy number. *Nature Genetics,* 40**,** 23-25.

Howard, O. M., Turpin, J. A., Goldman, R. & Modi, W. S. 2004. Functional redundancy of the human CCL4 and CCL4L1 chemokine genes. *Biochemical and Biophysical Research Communications,* 320**,** 927-931.

Hu, L., Song, W., Brill, I., Mulenga, J., Allen, S., Hunter, E., Shrestha, S., Tang, J. & Kaslow, R. A. 2012. Genetic variations and heterosexual HIV-1 infection: analysis of clustered genes encoding CC-motif chemokine ligands. *Genes & Immunity,* 13**,** 202-205.

Kahn, J. O. & Walker, B. D. 1998. Acute human immunodeficiency virus type 1 infection. *The New England Journal of Medicine,* 339**,** 33-39.

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

Kaur, G. & Mehra, N. 2009. Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection. *Tissue antigens,* 73**,** 289-301.

Kiepiela, P. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature,* 432**,** 769-775.

Kiepiela, P., Ngumbela, K., Thobakgale, C., Ramduth, D., Honeyborne, I., Moodley, E., Reddy, S., de Pierres, C., Mncube, Z., Mkhwanazi, N., Bishop, K., van der Stok, M., Nair, K., Khan, N., Crawford, H., Payne, R., Leslie, A., Prado, J., Prendergast, A., Frater, J., McCarthy, N., Brander, C., Learn, G. H., Nickle, D., Rousseau, C., Coovadia, H., Mullins, J. I., Heckerman, D., Walker, B. D. & Goulder, P. 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Medicine,* 13**,** 46-53.

Kim, T. G., Jeong, S.-Y. & Cho, K.-S. 2014. Comparison of droplet digital PCR and quantitative real-time PCR for examining population dynamics of bacteria in soil. *Applied Microbiology and Biotechnology,* 98**,** 6105-6113.

Koopmann, W., Ediriwickrema, C. & Krangel, M. S. 1999. Structure and function of the glycosaminoglycan binding site of chemokine macrophage-inflammatory protein-1 beta. *Journal of Immunology,* 163**,** 2120-2127.

Kostrikis, L. G., Huang, Y., Moore, J. P., Wolinsky, S. M., Zhang, L., Guo, Y., Deutsch, L., Phair, J., Neumann, A. U. & Ho, D. D. 1998. A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation. *Nature Medicine,* 4**,** 350-353.

Kuhn, L., Schramm, D. B., Donninger, S., Meddows-Taylor, S., Coovadia, A. H., Sherman, G. G., Gray, G. E. & Tiemessen, C. T. 2007. African infants' CCL3 gene copies influence perinatal HIV transmission in the absence of maternal nevirapine. *AIDS,* 21**,** 1753-1761.

Larsen, M. H., Thorner, L. W., Zinyama, R., Amstrup, J., Kallestrup, P., Gerstoft, J., Gomo, E., Erikstrup, C. & Ullum, H. 2012. CCL3L gene copy number and survival in an HIV-1 infected Zimbabwean population. *Infection, Genetics and Evolution* 12**,** 1087-1093.

Lillard, J. W., Jr., Singh, U. P., Boyaka, P. N., Singh, S., Taub, D. D. & McGhee, J. R. 2003. MIP-1alpha and MIP-1beta differentially mediate mucosal and systemic adaptive immunity. *Blood,* 101**,** 807-814.

Liu, S., Yao, L., Ding, D. & Zhu, H. 2010. CCL3L1 copy number variation and susceptibility to HIV-1 infection: a meta-analysis. *PLoS One,* 5**,** e15778.

Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. M. & Clore, G. M. 1994. High-resolution solution structure of the beta chemokine hMIP-1 beta by multidimensional NMR. *Science,* 263**,** 1762-1767.

Louisirirotchanakul, S., Liu, H., Roongpisuthipong, A., Nakayama, E. E., Takebe, Y., Shioda, T. & Wasi, C. 2002. Genetic analysis of HIV-1 discordant couples in Thailand: association of CCR2 64I homozygosity with HIV-1-negative status. *Journal of acquired immune deficiency syndromes,* 29**,** 314-315.

Mamtani, M., Rovin, B., Brey, R., Camargo, J. F., Kulkarni, H., Herrera, M., Correa, P., Holliday, S., Anaya, J.-M. & Ahuja, S. K. 2007. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythematosus. *Annals of the rheumatic diseases*.

Martin, M. P., Qi, Y., Gao, X., Yamada, E., Martin, J. N., Pereyra, F., Colombo, S., Brown, E. E., Shupert, W. L., Phair, J., Goedert, J. J., Buchbinder, S., Kirk, G. D., Telenti, A., Connors, M., O'Brien, S. J., Walker, B. D., Parham, P., Deeks, S. G., McVicar, D. W. & Carrington, M. 2007. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature Genetics,* 39**,** 733-740.

Martinson, J. J., Chapman, N. H., Rees, D. C., Liu, Y.-T. & Clegg, J. B. 1997. Global distribution of the CCR5 gene 32-basepair deletion. *Nature genetics,* 16**,** 100-103.

McKinney, C., Merriman, M. E., Chapman, P. T., Gow, P. J., Harrison, A. A., Highton, J., Jones, P. B., McLean, L., O'Donnell, J. L. & Pokorny, V. 2008. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Annals of the rheumatic diseases,* 67**,** 409-413.

Meddows-Taylor, S., Donninger, S. L., Paximadis, M., Schramm, D. B., Anthony, F. S., Gray, G. E., Kuhn, L. & Tiemessen, C. T. 2006. Reduced ability of newborns to produce CCL3 is associated with increased susceptibility to perinatal human immunodeficiency virus 1 transmission. *Journal of General Virology,* 87**,** 2055-2065.

Mellors, J. W., Munoz, A., Giorgi, J. V. & et al. 1997. PLasma viral load and cd4+ lymphocytes as prognostic markers of hiv-1 infection. *Annals of Internal Medicine,* 126**,** 946-954.

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

Menten, P., Struyf, S., Schutyser, E., Wuyts, A., De Clercq, E., Schols, D., Proost, P. & Van Damme, J. 1999. The LD78beta isoform of MIP-1alpha is the most potent CCR5 agonist and HIV-1-inhibiting chemokine. *Journal of Clinical Investigation* 104**,** R1-5.

Menten, P., Wuyts, A. & Van Damme, J. 2002. Macrophage inflammatory protein-1. *Cytokine & Growth Factor Reviews,* 13**,** 455-481.

Middleton, D. & Gonzelez, F. 2010. The extensive polymorphism of KIR genes. *Immunology,* 129**,** 8-19.

Modi, W. S., Bergeron, J. & Sanford, M. 2001. The human MIP-1β chemokine is encoded by two paralogous genes, ACT-2 and LAG-1. *Immunogenetics,* 53**,** 543-549.

Modi, W. S., Lautenberger, J., An, P., Scott, K., Goedert, J. J., Kirk, G. D., Buchbinder, S., Phair, J., Donfield, S., O'Brien, S. J. & Winkler, C. 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**,** 120-128.

Moser, B. & Loetscher, P. 2001. Lymphocyte traffic control by chemokines. *Nature Immunology,* 2**,** 123-128.

Moss, J. A. 2013. HIV/AIDS Review. *Radiologic Technology,* 84**,** 247-267.

Nakajima, T., Ohtani, H., Naruse, T., Shibata, H., Mimaya, J. I., Terunuma, H. & Kimura, A. 2007. Copy number variations of CCL3L1 and long-term prognosis of HIV-1 infection in asymptomatic HIV-infected Japanese with hemophilia. *Immunogenetics,* 59**,** 793-798.

Nibbs, R. J., Yang, J., Landau, N. R., Mao, J. H. & Graham, G. J. 1999. LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity. *Journal of Biological Chemistry,* 274**,** 17478-17483.

Nordang, G. B., Carpenter, D., Viken, M. K., Kvien, T. K., Armour, J. A. & Lie, B. A. 2012. Association analysis of the CCL3L1 copy number locus by paralogue ratio test in Norwegian rheumatoid arthritis patients and healthy controls. *Genes & Immunity,* 13**,** 579-582.

Novembre, J., Galvani, A. P. & Slatkin, M. 2005. The geographic spread of the CCR5 Delta32 HIV-resistance allele. *PLoS Biology,* 3**,** e339.

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

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

Okulicz, J. F., Marconi, V. C., Landrum, M. L., Wegner, S., Weintrob, A., Ganesan, A., Hale, B., Crum-Cianflone, N., Delmar, J., Barthel, V., Quinnan, G., Agan, B. K., Dolan, M. J. & Infectious Disease Clinical Research Program, H. I. V. W. G. 2009. Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study. *The Journal of Infectious Diseases* 200**,** 1714-1723.

Palella, F. J., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., Aschman, D. J. & Holmberg, S. D. 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine,* 338**,** 853-860.

Paximadis, M., Minevich, G., Winchester, R., Schramm, D. B., Gray, G. E., Sherman, G. G., Coovadia, A. H., Kuhn, L. & Tiemessen, C. T. 2011. KIR-HLA and maternal-infant HIV-1 transmission in sub-Saharan Africa. *PLoS One,* 6**,** e16541.

Paximadis, M., Mohanlal, N., Gray, G. E., Kuhn, L. & Tiemessen, C. T. 2009. 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,* 36**,** 21-32.

Paximadis, M., Schramm, D. B., Gray, G. E., Sherman, G., Coovadia, A., Kuhn, L. & Tiemessen, C. T. 2013. Influence of intragenic CCL3 haplotypes and CCL3L copy number in HIV-1 infection in a sub-Saharan African population. *Genes & Immunity,* 14**,** 42-51.

Pelak, K., Need, A. C., Fellay, J., Shianna, K. V., Feng, S., Urban, T. J., Ge, D., De Luca, A., Martinez-Picado, J., Wolinsky, S. M., Martinson, J. J., Jamieson, B. D., Bream, J. H., Martin, M. P., Borrow, P., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A., Carrington, M., Goldstein, D. B., Alter, G. & Immunology, N. C. f. H. A. V. 2011. Copy number variation of KIR genes influences HIV-1 control. *PLoS Biology,* 9**,** e1001208.

Picton, A. C., Paximadis, M. & Tiemessen, C. T. 2013. Contribution of variable CCL3L copy number to CCL3 protein production in two ethnically divergent South African populations. *Infection, Genetics and Evolution* 14**,** 347-356.

Picton, A. C. P., Paximadis, M. & Tiemessen, C. T. 2010. Genetic variation within the gene encoding the HIV-1 CCR5 coreceptor in two South African populations. *Infection, Genetics and Evolution,* 10**,** 487-494.

Piot, P. & Colebunders, R. 1987. Clinical manifestations and the natural history of HIV infection in adults. *The Western Journal of Medicine,* 147**,** 709-912.

Piot, P. & Quinn, T. C. 2013. Response to the AIDS pandemic--a global health model. *The New England Journal of Medicine,* 368**,** 2210-2218.

Proost, P., Menten, P., Struyf, S., Schutyser, E., De Meester, I. & Van Damme, J. 2000. Cleavage by CD26/dipeptidyl peptidase IV converts the chemokine LD78beta into a most efficient monocyte attractant and CCR1 agonist. *Blood,* 96**,** 1674-1680.

Qi, Y., Martin, M. P., Gao, X., Jacobson, L., Goedert, J. J., Buchbinder, S., Kirk, G. D., O'Brien, S. J., Trowsdale, J. & Carrington, M. 2006. KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathogens,* 2**,** e79.

Qu, W., Robinson, M. & Zhang, F. J. 2008. Factors influencing the natural history of HIV-1 infection. *Chinese Medical Journal (Engl),* 121**,** 2613-2621.

Reynes, J., Portales, P., Segondy, M., Baillat, V., Andre, P., Reant, B., Avinens, O., Couderc, G., Benkirane, M., Clot, J., Eliaou, J. F. & Corbeau, P. 2000. CD4+ T cell surface CCR5 density as a determining factor of virus load in persons infected with human immunodeficiency virus type 1. *The Journal of Infectious Diseases* 181**,** 927-932.

Sabin, C. A. & Lundgren, J. D. 2013. The natural history of HIV infection. *Current Opinion in HIV and AIDS,* 8**,** 311-317.

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

Shalekoff, S., Meddows-Taylor, S., Schramm, D. B., Donninger, S. L., Gray, G. E., Sherman, G. G., Coovadia, A. H., Kuhn, L. & Tiemessen, C. T. 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. *Journal of acquired immune deficiency syndromes,* 48**,** 245-254.

Shao, W., Tang, J., Song, W., Wang, C., Li, Y., Wilson, C. M. & Kaslow, R. A. 2007. CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. *Genes & Immunity,* 8**,** 224-231.

Shostakovich-Koretskaya, L., Catano, G., Chykarenko, Z. A., He, W., Gornalusse, G., Mummidi, S., Sanchez, R., Dolan, M. J., Ahuja, S. S., Clark, R. A., Kulkarni, H. & Ahuja, S. K. 2009. Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. *AIDS,* 23**,** 679-688.

Shrestha, S., Nyaku, M. & Edberg, J. C. 2010. Variations in CCL3L gene cluster sequence and non-specific gene copy numbers. *BMC Research Notes,* 3**,** 74.

Siegal, F. P., Lopez, C., Hammer, G. S., Brown, A. E., Kornfeld, S. J., Gold, J., Hassett, J., Hirschman, S. Z., Cunningham-Rundles, C., Adelsberg, B. R. & et al. 1981. Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions. *The New England Journal of Medicine,* 305**,** 1439-1444.

Singh, P., Kaur, G., Sharma, G. & Mehra, N. K. 2008. Immunogenetic basis of HIV-1 infection, transmission and disease progression. *Vaccine,* 26**,** 2966-2980.

Smith, M. W., Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Lomb, D. A., Goedert, J. J., O'Brien, T. R., Jacobson, L. P. & Kaslow, R. 1997. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. *Science,* 277**,** 959-965.

Sowrirajan, B. & Barker, E. 2011. The Natural Killer Cell Cytotoxic Function Is Modulated by HIV-1 Accessory Proteins. *Viruses,* 3**,** 1091-1111.

Standiford, T. J., Kunkel, S. L., Liebler, J. M., Burdick, M. D., Gilbert, A. R. & Strieter, R. M. 1993. Gene expression of macrophage inflammatory protein-1 alpha from human blood monocytes and alveolar macrophages is inhibited by interleukin-4. *American Journal of Respiratory Cell and Molecular Biology,* 9**,** 192-198.

Stankiewicz, P. & Lupski, J. R. 2002. Genome architecture, rearrangements and genomic disorders. *Trends in Genetics,* 18**,** 74-82.

Taub, D. D., Conlon, K., Lloyd, A. R., Oppenheim, J. J. & Kelvin, D. J. 1993. Preferential migration of activated CD4+ and CD8+ T cells in response to MIP-1 alpha and MIP-1 beta. *Science,* 260**,** 355-358.

Thananchai, H., Gillespie, G., Martin, M. P., Bashirova, A., Yawata, N., Yawata, M., Easterbrook, P., McVicar, D. W., Maenaka, K., Parham, P., Carrington, M., Dong, T. & Rowland-Jones, S. 2007. Cutting Edge: Allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *Journal of Immunology,* 178**,** 33-37.

Touloumi, G. & Hatzakis, A. 2000. Natural history of HIV-1 infection. *Clinics in Dermatology,* 18**,** 389-399.

Townson, J. R., Barcellos, L. F. & Nibbs, R. J. 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. *European journal of immunology,* 32**,** 3016-3026.

UNAIDS. 2016. Aids by the numbers. *UNAIDS Web site* [Online]. Available: http://www.unaids.org

Urban, T. J., Weintrob, A. C., Fellay, J., Colombo, S., Shianna, K. V., Gumbs, C., Rotger, M., Pelak, K., Dang, K. K., Detels, R., Martinson, J. J., O'Brien, S. J., Letvin, N. L., McMichael, A. J., Haynes, B. F., Carrington, M., Telenti, A., Michael, N. L. & Goldstein, D. B. 2009. CCL3L1 and HIV/AIDS susceptibility. *Nature Medicine,* 15**,** 1110-1112.

Vergis, E. N. & Mellors, J. W. 2000. Natural history of HIV-1 infection. *Infectious Disease Clinics of North America,* 14**,** 809-825, v-vi.

Walker, S., Janyakhantikul, S. & Armour, J. A. 2009. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics,* 93**,** 98-103.

Walker, W. E., Kurscheid, S., Joshi, S., Lopez, C. A., Goh, G., Choi, M., Barakat, L., Francis, J., Fisher, A., Kozal, M., Zapata, H., Shaw, A., Lifton, R., Sutton, R. E. & Fikrig, E. 2015. Increased Levels of Macrophage Inflammatory Proteins Result in Resistance to R5-Tropic HIV-1 in a Subset of Elite Controllers. *Journal of Virology,* 89**,** 5502-5514.

Weber, J. 2001. The pathogenesis of HIV-1 infection. *British Medical Bulletin,* 58**,** 61-72.

Wigginton, J. E., Cutler, D. J. & Abecasis, G. R. 2005. A Note on Exact Tests of Hardy-Weinberg Equilibrium. *American Journal of Human Genetics,* 76**,** 887-893.

Wilen, C. B., Tilton, J. C. & Doms, R. W. 2012. HIV: cell binding and entry. *Cold Spring Harbor Perspectives in Medicine,* 2.

Wilson, E. M. P. & Sereti, I. 2013. Immune restoration after antiretroviral therapy: the pitfalls of hasty or incomplete repairs. *Immunological reviews,* 254**,** 343-354.

Wu, L., Paxton, W. A., Kassam, N., Ruffing, N., Rottman, J. B., Sullivan, N., Choe, H., Sodroski, J., Newman, W. & Koup, R. A. 1997. CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. *The Journal of experimental medicine,* 185**,** 1681-1692.

Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. 2015. A copy number variation map of the human genome. *Nature Reviews Genetics,* 16**,** 172-183.

Zlotnik, A. & Yoshie, O. 2000. Chemokines: a new classification system and their role in immunity. *Immunity,* 12**,** 121-127.

Zolopa, A., Andersen, J., Powderly, W., Sanchez, A., Sanne, I., Suckow, C., Hogg, E. & Komarow, L. 2009. Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections: a multicenter randomized strategy trial. *PLoS One,* 4**,** e5575.

APPENDIX A: Supplementary data for Chapter 3

Primers and probes used for *CCL3L* and *CCL4L* copy number determination obtained from Townson et al. (2002) and (Shostakovich-Koretskaya et al., 2009).

Gene	Forward primers (5' – 3')	Reverse primers (5' – 3')	Probes [3] (5' – 3')
CCL3L[1]	TCTCCACAGCTTCCTAACCAAGA	CTGGACCCACTCCTCACTGG	AGGCCGGCAGGTCTGTGCTGA
CCL3La[2]	GGGTATGACTTCTTGAACCGACAAA	GGTTCTCTGTTTCTCTATGTGATCCA	CATGAAGAGAGCTAAGAGAAC
CCL3Lb[2]	CATCCACTCGCTCACACCTGTA	GCGGTCGGCGTGTCA	AGAGTTGGGCTTATTCT
CCL4L[2]	CATGGTCAGGCAGAGGAAGATG	GCTTGCCTCTTTTGGTTTGGAAT	TACCACAGGCAAGGGAT
CCL4La[2]	GGAAGATGCCTACCACAGGC	GCGCAGACTTGCTTGCC	CTTGTTCTACaGATTCC
CCL4Lb[2]	GGAAGATGCCTACCACAGGC	GCGCAGACTTGCTTGCC	CTTGTTCTACgGATTCC
BGB	TCGCTTTCTTGCTGTCCAATTTCTA	ATGCTCAAGGCCCTTCATAATATCC	CCTAAGTCCAACTACTAAACTG

[1](Townson et al., 2002)

[2](Shostakovich-Koretskaya et al., 2009)

[3]Minor grove binding probes [FAM fluorescent label for all probes except for *BGB* (VIC)]

Haplotype	Forward primers (5' – 3')	Reverse primers (5' – 3')
Hap-A1	TAGCATGACAGCATCACTA**Y**[1]	TGGAGACCTGCATGATTCT
Hap-A3	GAAGAGTCAAGGAGAAAGAAGG	GAGCAGTTGAGGAAGGCA**R**[1]

Primers for *CCL3* haplotype determination obtained from (Paximadis et al., 2013)

[1]allele specific primers

APPENDIX B: Ethical clearance

R14/49 Miss Sizanani Mncube

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M1511103

NAME: (Principal Investigator)	Miss Sizanani Mncube
DEPARTMENT:	Virology Centre for HIV and STI's (CHIVSTI), National Institute for Communicable Diseases (NICD) National Health Laboratory Services (NHLS), Sandringham, Johannesburg
PROJECT TITLE:	The Role of CCL3 and CCL4 Encording Genes in Control of HIV-1 Infection
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-Study
SUPERVISOR:	Prof Caroline Tiemessen and Dr Maria Paximadis
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	09/12/2015

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

Date 25/01/2016

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX C: Turnitin report

ORIGINALITY REPORT

%16	%8	%13	%3
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	wiredspace.wits.ac.za Internet Source	%4
2	Paximadis, M, D B Schramm, G E Gray, G Sherman, A Coovadia, L Kuhn, and C T Tiemessen. "Influence of intragenic CCL3 haplotypes and CCL3L copy number in HIV-1 infection in a sub-Saharan African population", Genes and Immunity, 2012. Publication	%1
3	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. Publication	%1
4	Submitted to University of Witwatersrand Student Paper	%1
5	Bharuthram, Avani, Maria Paximadis, Anabela C.P. Picton, and Caroline T. Tiemessen. "Comparison of a quantitative Real-Time PCR assay and droplet digital	%1