Genetic/epigenetic determinants in chemokines and chemokine receptor genes that influence HIV susceptibility in a cohort of high-risk women from South Africa

by

VERON RAMSURAN

Submitted in partial fulfilment of the academic requirements for the degree of

Doctor of Philosophy in the Discipline of Genetics,

School of Biochemistry, Genetics and Microbiology

University of KwaZulu-Natal

Pietermaritzburg

2010

PREFACE

The experimental work described in this thesis was carried out in the following laboratories: HIV Pathogenesis Programme (HPP), Hasso Plattner Research Laboratory (HPRL), Centre for the AIDS Program of Research in South Africa (CAPRISA), located in the Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, 4013, South Africa; Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, San Antonio, Texas 78229; Department of Medicine, University of Texas Health Science Center, San Antonio, Texas; and, finally, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. This research was carried out from January 2006 to December 2010, under the supervision of Professor Thumbi Ndung'u (Durban), Professor Sunil K. Ahuja (Texas) and Professor Emil Kormuth (Pietermaritzburg). These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any University. Where use has been made of the work of others it is duly acknowledged in the text.

_____

Signed: Veron Ramsuran (candidate).

_____

Signed: Professor Sunil K. Ahuja (co-supervisor).

_____

Signed: Professor Thumbi Ndung'u (supervisor).

_____

Signed: Professor Emil Kormuth (co-supervisor).

ABSTRACT

Background: There is extensive heterogeneity in the risk of acquiring HIV-1 and in the rate of disease progression among HIV infected persons. Host genetic and epigenetic factors may account for some of this variation. In this study we investigated the influence of the following parameters on the risk of acquiring HIV infection: (i) the conjoint influence of cellular factors (peripheral blood cell counts) and the null state on erythrocytes of Duffy Antigen Receptor for Chemokines (DARC); and (ii) the combinatorial content of the copy number of *CCL3L* and *CCL4L*, genes that are ligands for the HIV coreceptor CCR5. We also initiated studies to determine the association between the DNA methylation status in the regulatory regions of CCR5 in HIV-uninfected and HIV–infected subjects from the cohort.

Materials and Methods: The study group was a cohort of Black South African women, including commercial sex workers. The assays used were genotyping, genome-wide association studies (GWAS), and pyrosequencing assays to assess DNA methylation at CpG dinucleotides in the CCR5 regulatory regions.

Conclusions:

(i) Pre-seroconversion neutrophil and platelet counts influence risk of HIV infection. The trait of Duffy-null-associated low neutrophil counts i.e. individuals with DARC −46C/C genotype and neutrophil count below 2500 cells/mm3 influences HIV susceptibility. Because of the high prevalence of this trait among persons of African ancestry, it may contribute significantly to the dynamics of the HIV epidemic in Africa.

(ii) The combinatorial content of the genes in the 17q12 region, namely, *CCL3L*- and *CCL4L*-related genes rather than the gene dose of any one of these genes influences both HIV

susceptibility and disease progression. Specifically, our results point to an important role for the balance between the copy numbers of protective (*CCL3La* and *CCL3L*b) versus detrimental (*CCL4Lb*) chemokine genes in influencing HIV susceptibility, viral replication and CD4+ T cell loss. Our findings supports previous data that shows subjects with a low dose of *CCL3L* genes are preferentially infected with HIV.

(iii) We demonstrate a strong negative correlation between the methylation levels from the CCR5 promoter and intron regions with the CCR5 expression levels and further show that the DNA methylation status are lower in HIV+ subjects compared to HIV- subjects.

These studies reinforce the idea that both genetic and epigenetic factors are likely to influence HIV-AIDS pathogenesis.

Dedication

To

Nāradi Gopi Ramsuran

Lord Krishna

Acknowledgements

Firstly I want to acknowledge and convey a huge thank you to CAPRISA for funding, samples and support. I would like to thank Prof. S. A. Karim, Q. A. Karim and all the members of the CAPRISA Acute infection 002 and HEPS 078 team and also a very special thank you to the participants of these studies.

There is one person that I have gained an invaluable amount of knowledge from during the course of this PhD that is Dr. Sunil K Ahuja. I would like to deeply and sincerely thank Dr. Sunil Ahuja for the excellent mentorship as well as bringing out my passion for science and helping me on a personal level and making me a better scientist.

Thank you to my supervisor Prof. Thumbi Ndung'u for all the support and advice also the opportunities that I know would have not being possible without him.

Prof. Julie McElrath from Seattle for hosting me in her laboratory, thank you for being an excellent mentor. Spending time in her laboratory gives me hope that we can beat HIV.

To all the members of the HPRL/HPP laboratory, thank you guys for all the support and help through out my degree. I am deeply greatful to each one of you with a special thank you to Mrs Keshni Hiramen and members of Room 213.

To the members of Dr. Sunil Ahuja's laboratory I want to express my deep gratitude to Dr. Hemant Kulkarni, Dr. Manju Mantani, Dr. Weijing He, Dr. German Gornalusse, Dr. Maria Soledad Valera, Dr. Gabriel Cataño, Robert Maldonado, Andrew Carrilo, Una Aluyen, Erica Hayes and Keith Joseph.

To my parents, I am extremely grateful for everything that you have done for me. I know it's because of you I reached this stage in my life. I can not thank you enough.

To my all my family members, especially my brothers Duran, Nedon and Kishen thank you for everything and bearing with me during this degree.

To my wife Sri Radhika Ramsuran, thank you for all the support, help and understanding. A special thank you to my baby, Nāradi Gopi Ramsuran for bringing me untold joy and happiness.

Finally a huge thank you to Lord Krishna for making all this possible, Hare Krishna.

TABLE OF CONTENTS

List of Tables

List of Figures

partners in the prior 3 month (335). 462 of these were HIV-positive subjects and 68 met exclusion criteria as described previously (e.g. pregnant, declined follow-up for two years) (335). 245 subjects (green box) were enrolled into the Acute Infection study (335). 30 were excluded for the present study because they were either not black or DNA was unavailable. Of the 215 HIV-negative women who were followed prospectively, 169 (78.6%) were self-reported sex workers and at the end of the two year follow up period 28 women had seroconverted and 122 remained HIV seronegative (exposed uninfected). 65 women who had less than 2 years of follow-up were excluded from the current analyses. PBC and GWAS data was available on 27 HIV-seroconverting women and 115 HIV non-seroconverting women. (B) Evaluation of population stratification in the study groups. Red squares and blue diamond's indicate the mean PC score for the first 10 principal components (ordered from left to right) for HIV-infected and -uninfected subjects, respectively. Error bars represent the 95% confidence interval. Numbers at the right side are significant values obtained using Student's T test.

List of abbreviations

7TMD	-	Seven Transmembrane Domain
Ab	-	Anti-body
AIDS	-	Acquired Immune Deficiency Syndrome
AIMS	-	Ancestry-informative markers
APC	-	Allophycocyanin
APOBEC3G	-	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G
ART	-	Antiretroviral therapy
BGB	-	Beta-globin
Bp	-	Base pair
C4	-	Complement component C4
CAPRISA	-	Centre for the Aids Programme of Research
CCL	-	CC chemokine ligands
CCL3L	-	CC chemokine ligand 3-like 1 (MIP-1αP and LD78b)
*CCL4L*1	-	CC chemokine ligand 4-like 1 (MIP-1b-like)
CCR5	-	C-C Chemokine Receptor 5
CN	-	Copy Number
CNV	-	Copy number variation

CNVR	-	Copy number variation regions
CSW	-	Commercial sex workers
CXCR4	-	C-X-C Chemokine Receptor 4
DARC	-	Duffy antigen receptor for chemokines
DDMRI	-	Doris Duke Medical Research Institute
DNMT	-	DNA methyltransferases
EC	-	Elite Controllers
EIA	-	Enzyme immunoassay
EU	-	Exposed Uninfected
FACS	-	Fluorescence-activated cell sorting
FITC	-	Fluorescein isothiocyanate
FSW	-	Female sex workers
GCN	-	Gene copy number
GEE	-	Generalized estimating equations
Gp120	-	Glycoprotein 120
Gp41	-	Glycoprotein 41
GPCRs	-	G protein-coupled receptors
GWAS	-	Genome Wide Association study
HAART	-	Highly Active Anti-Retroviral Therapy

HBV	-	Hepatitis B virus
HCC	-	Haemofiltrate CC chemokine
HDACs	-	Histone deacetylases
HEP	-	Human Epigenome Project
HEPS	-	Highly Exposed but Persistently Seronegative
HGDP-CEPH	-	Human Genome Diversity Cell Line Panel
HHA-HHG	-	Human-Haplogroup A to G
HIV	-	Human Immunodeficiency Virus
HIV-	-	Human Immunodeficiency Virus Negative
HIV+	-	Human Immunodeficiency Virus Positive
HPP	-	HIV pathogenesis programme
HRW	-	High-risk women
HWE	-	Hardy-Weinberg equilibrium
Indels	-	Insertions and deletions
KRITH	-	KwaZulu-Natal Research Institute for TB and HIV
L/D	-	Live/dead
LTNPs	-	Long Term Non-Progressors
LTR	-	Long Terminal Repeat
MACS	-	Multicenter AIDS Cohort Study

MBDs	-	Methyl-binding domain proteins
MCP-1	-	Monocyte chemoattractant protein-1
MGB	-	Minor groove binder
MGBNFQ	-	Molecular-Groove Binding Non-fluorescence Quencher
MZT	-	Monozygotic twins
NK	-	Natural killer
NSI	-	Non-Syncytium Inducing
ORF	-	Open reading frame
PC	-	Principal component
PCA	-	Principle components analysis
Pd	-	Downstream promoter
Pdg	-	Per diploid genome
PE	-	Phycoerythrin
Pr	-	Promoter
PRT	-	Paralogue ratio test
PTM	-	Post-translational histone modifications
Pu	-	Upstream promoter
qPCR	-	Quantitative Polymerase Chain Reaction
QSY7	-	Quencher Series 7

RBC	-	Red blood cell
RFLP	-	Restriction fragment length polymorphism
Rs	-	Reference SNP
SAHA	-	Suberoylanilide hydroxamic acid
SI	-	Syncytium Inducing
SIV	-	Simian Immunodeficiency Virus
SLE	-	Systemic lupus erythematosus
SNP	-	Single Nucleotide Polymorphism
STDs	-	Sexually transmitted diseases
STIs	-	Sexually transmitted infections
TAMRA	-	Tetramethylrhodamine
TRIM5 α	-	Tripartite motif-containing 5 alpha
TSS	-	Transcription start site
UKZN	-	University of KwaZulu-Natal
VL	-	Viral Load
WBC	-	White blood cell

CHAPTER 1 - INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) is caused by Human Immunodeficiency Virus (HIV). HIV/AIDS is currently considered a pandemic, as an estimated 35 million people are infected globally (296). Recent reports by UNAIDS has shown that AIDS has already claimed 21 million lives since the first reported cases in 1981. The epidemic is particularly severe in Sub-Saharan Africa with approximately 21 million people infected, which is about 60% of the HIV infections globally, yet this region accounts for about 10% of the world's population (297).

The transmission of HIV generally occurs when the HIV infected blood and genital secretions come in contact with tissues e.g. vaginal area, anal area, mouth, or eyes (the mucosal membranes), or with a break in the skin. The most common mode of transmission throughout the world include sexual contact, sharing needles, and by transmission from infected mothers to their newborns during pregnancy, labour (the delivery process), or breastfeeding.

HIV can be found in the blood and genital secretions of all individuals acutely infected with HIV, despite whether or not they display clinical symptoms of HIV infection. Typical symptoms during early HIV infection are fever, fatigue and rash. Some other common symptoms reported include headache, swollen lymph nodes, and sore throat, which are very similar to the common cold. HIV is very effective due to its ability to weaken the immune system thus leaving the individual susceptible to opportunistic infections and tumours. HIV infects fundamental cells in the human immune system and these include helper T cells, macrophages and dendritic cells as well as cells of the nervous system.

For the vast majority of infectious diseases there have always been individuals that are naturally resistant to infection, and generally the cure for the disease comes from exploring the defence mechanism in these individuals. The HIV disease is no different; there are individuals that remain HIV negative despite repeated exposure to the virus. They include persons that are intravenous drug users (19), sex workers who engage in unprotected sex (82, 142), health care workers who are accidently exposed (46, 235), infants that are born to infected mothers (260), exposed homosexuals (46, 64) and heterosexuals in contact with HIV-infected patients (167). Individuals that are infected with HIV also display contrasting phenotypes and have differences in their HIV-1 viral set points, the rates of decline of CD4 T cells, emergence of CTL escape mutants, and risk of opportunistic infections resulting in varying rates of disease progression (146). Approximately 5% - 10% (60) of the infected individuals, do not progress to disease for more than 10 years, and have low or undetectable levels of virus. These individuals are termed Long Term Non-Progressors (LTNPs). There are also some instances whereby individuals can control the viral load below the level of detection i.e. <50 copies HIV-RNA/mL, these individuals are termed Elite Controllers (EC) (60).

Multiple sexual partners and frequent coital acts leads to higher risk of acquiring and transmitting HIV than the general population (142) particaularly in areas where heterosexual transmission is the predominant form of transmission, such as Sub-Saharan Africa (243). High-risk sexual behaviour is not the only reason for increased infections, this can also be attributed to poor social conditions, poor knowledge about HIV infection and also the prevalence of other sexually transmitted infections which have been observed in sex worker

cohorts in Kenya (277), Zaire (222), Somalia (54) and South Africa (142) as those sexual transmitted diseases are believed to increased the risk of HIV acquisition.

Of particular interest are individuals that do not acquire the virus despite repeated exposure to the virus. These individuals are termed Highly Exposed but Persistently Seronegative (HEPS) or Exposed Uninfected (EU). The mechanisms by which such individuals resist infection might represent important correlates of protection that may potentially be helpful in the design and development of an effective vaccine against HIV. A considerable amount of effort has been put in to developing cohorts that can help identify the protective mechanisms found in these individuals (301).

HIV progresses to AIDS at variable rates in different individuals, which can be attributed to three factors; viral, host, and environmental factors (Figure 1.1). An example of this is the viral factors that determine the replication properties of the virus or its ability to escape immune responses i.e. if a virus is less fit, then it is unable to replicate and hence results in delayed progression to AIDS. In the environment factors, if an individual does not engage in risky behaviour this may confer protection from various viral co-infections that can influence rates of progression to AIDS. Finally, humans have intrinsic defense mechanisms against HIV-1 that may afford resistance to HIV-AIDS. Host genetic factors have been identified that are associated with resistance/susceptibility to HIV-1 infection and variable responses to therapy (146).

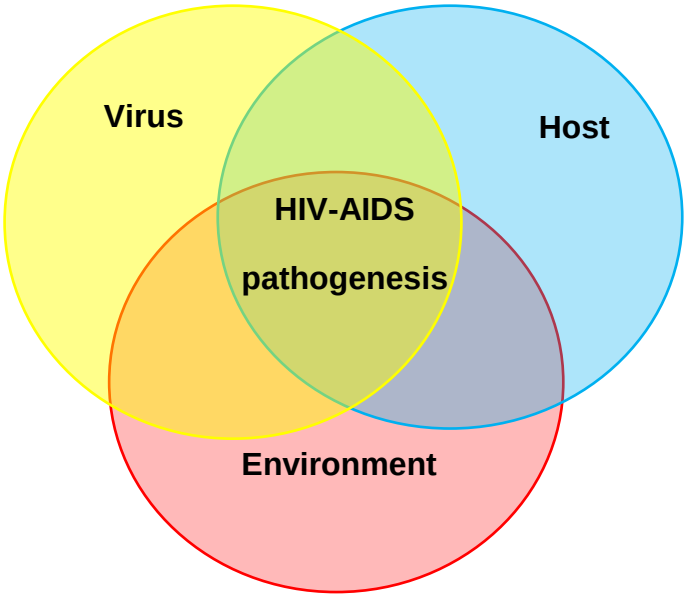


Figure 1.1: Factors influencing HIV-AIDS pathogenesis.

Relevance to this thesis

To date considerable progress has been made in identifying cohorts to study HIV-1 transmission and AIDS progression. However, there have been difficulties in recruiting individuals with high-risk behaviours and documented exposure to HIV for prospective follow-up into cohorts. In this study, my work focuses on a cohort of highly exposed female sex workers. This group of repeatedly HIV-1 exposed individuals some remain negative throughout the duration of the study, therefore implying that they possess or utilize a protective mechanism. The cohort is also comprised completely of individuals of African descent. Hence, the characteristics of the cohort have value for understanding the host factors that influence HIV susceptibility in a cohort of South African women.

CHAPTER 2 - LITERATURE REVIEW

2.1. Genetics

2.1.1. Host Genetic Factors

Over the past few years, HIV-host interaction studies have identified a number of host cellular factors that are involved in the HIV infection process. To put an approximate number to the cellular factors needed for HIV infection, various investigators have performed a range of studies including genome-wide RNA scans. Several studies have identified genes that have shown to interact physically or functionally with HIV-1 (12, 28, 75, 88, 153, 234, 242, 326). When combined there are 1,254 human genes that may be involved in viral replication (12, 28, 33, 75, 88, 153, 234, 242, 326). These human genes are involved in every part of the viral life cycle from HIV entry right up to assembly and release from the cell membrane.

The HIV entry is a promising step for intervention; this is due to interactions on the surface of the cell that can be inhibited and thus leaving the viral pathogen entry restricted. Since this is the first step of the viral replication, inhibiting viral entry further prevents integration of the proviral genome into the host cell therefore preventing latent viral reservoirs.

HIV requires a mechanism to get into the host cells, the CD4 receptor, discovered more than 20 years ago, and has been shown to be the primary receptor for HIV (56). Therefore HIV replication occurs in CD4+ cells, these cells are mainly lymphocytes but also include monocytes and dendritic cells (151, 269). For a long time, it was believed HIV only required

one receptor to get into a cell however an experiment showing human cells expressing a transfected CD4 gene were permissive for viral entry while murine cells expressing human CD4 on the cell surface was not able for viral entry (13, 63, 187, 229). These experiments suggested that a species-specific coreceptor, that is unique to human CD4 cells, is required in addition to CD4 for viral entry. Chemokine receptors were discovered as a secondary receptor required for HIV entry (13, 63, 187, 229).

Although there are many host genes involved in the HIV life cycle (as explained above), this thesis is restricted to focus mainly on members of the chemokine system as some chemokine receptors are the gateway for viral entry and have been shown to be extremely important for HIV resistance (197).

2.1.2. Chemokines

The name chemokines is derived from chemoattractant cytokines. These are small molecules in the cytokine family that promote cellular movement by chemotaxis (210). Chemokines are expressed in a number of different cell types that include T cells, macrophages, natural killer (NK) cells, B cells, fibroblasts, and mast cells. These different cell types are mainly involved in cell trafficking and immunomodulation of inflammation and immune responses (146). This super-family of chemokines contains about 50 related proteins, which range from 68 to 120 amino acids (in the mature form) (240). These structurally related chemokines contain four invariant cysteine residues. The chemokines are classified depending on the arrangement of the first two of these cysteines, such that they are divided into four subfamilies; CXC, CC,

C and CX_3C, these four groups are also less commonly known as α, β, γ and δ respectively (210).

CC chemokines

The CC chemokines (or ß-chemokine) have two adjacent cysteines located close to the amino terminus (29, 130). Four cysteines (C4-CC chemokines), are found in most members in this subgroup, however, there are five members that possess six cysteines (C6-CC chemokines) (163, 309). In mammals, it has been reported that there are 28 members in the CC chemokines subfamily called CC chemokine ligands (CCL), ranging from CCL1 to CCL28 (163). The genes encoding this cluster of chemokines are located mainly on chromosome 17q11.2–q12 region. The CC subfamily can be further divided into various groups based on activities and/or sequence similarity that include; allergenic, pro-inflammatory, haemofiltrate CC chemokine (HCC), developmental and homeostatic groups. The CC chemokines that are classified in the allergenic, pro-inflammatory and HCC subgroups are considered inducible chemokines whereas the CC chemokines classified within the developmental and homeostatic subgroups are generally constitutively expressed, with exceptions to this rule (163). The CC chemokine subgroup have been shown to be potent chemoattractants of eosinophils and basophils (184), and induces chemotaxis of monocytes, NK cells, immature dendritic and B cells (226). Some examples of the CC chemokine subfamily are monocyte chemoattractant protein-1 (MCP-1 or CCL2) and this member has been shown to induce the monocytes to leave the bloodstream and move to surrounding tissue to become macrophages. Another example is the CCL5 (or RANTES) which has been shown to attract T cells, eosinophils and basophils which express the receptor, C-C Chemokine Receptor

5 (CCR5) (163, 226). The chemotactic proteins for macrophages, T-cells, eosinophils and basophils are RANTES, MIP-1α, MIP-1β, MCP-1, MCP-2, MCP-3 and eotaxin (63)

CXC Chemokines

The CXC chemokines (or α-chemokine) are a 16 member subfamily, and these genes are located mainly the chromosome 4q12–21 region. Some members in this family possess a three residue motif Glu-Leu-Arg (ELR) located near the amino terminus immediately adjacent to the CXC motif (163). There are eight members in this subfamily that lack this motif and hence this forms the basis for the subdivision within the CXC chemokines (163). The major difference between the ELR possessing and lacking groups are their ability to attract neutrophils, the ELR lacking group does not have neutrophil attracting ability, but attract lymphocytes and moncytes. The main role of the subgroup is to promote the adherence of neutrophils to endothelial cells and cell surfaces toward inflammatory sites (163, 226).

C Chemokines

The C chemokines (or γ-chemokine) contains two very similar members, the XCL1 and XCL2 (57). These two members differ by two amino acids found in the seventh and eighth position. These chemokines have been shown to induce chemotaxis of lymphocytes (163, 226).

CX_3C Chemokines

The CX_3C chemokine (or δ-chemokine) has just one member in this group that is the CX_3CL1 with three intervening amino acids between the first two cysteines (163). CX_3CL1 has potent chemoattractant activity for T cells and monocytes.

2.1.3. Chemokine Receptors

The chemokine receptors are a family of proteins that bind onto one or more members of the chemokine superfamily (240). It has been reported that there have been about 18 proteins that have followed this definition of being classified as a chemokine receptor. These are as follows; CCR1 to CCR11, CXCR1 to CXCR5, XCR1 and CX_3CR1, this nomenclature was assigned based on their specific chemokine preferences (210). The chemokine receptors are a branch of the rhodopsin family of cell surface, seven-transmembrane domain (7TMD), G protein-coupled receptors (GPCRs). Two other 7TMD chemokine receptors, namely; D6 and Duffy (also known as Duffy antigen receptor for chemokines or DARC) where excluded from systematic nomenclature due to these proteins lack of ability to produce a signal. (131, 218). The chemokine receptors common seven-transmembrane structure is made up of three extracellular loops and 4 intracellular loops separated by a hydrophobic domain (**figure 2.1**) (169).

Figure 2.1: The seven transmembrane structure and amino acid sequence of the CCR5 gene (adapted from McNicholl et al, 1997; (197)).

Intracellular pathogens have exploited these chemokine receptors for cellular entry and disease transmission. This phenomenon may be illustrated by two examples: firstly with malaria caused by the pathogen *Plasmodium vivax* and the Duffy receptor (131) (explained in detail later in literature section 2.1.6), secondly, instance is with HIV and the CCR5 receptor (discussed below). Other chemokine receptors including CXCR4 also function as HIV receptors however their role in disease progression is not fully understood (130, 131, 262).

2.1.4. HIV coreceptors

HIV is similar to all retroviruses and is surrounded by an envelope, consisting of a host cell derived lipid bilayer and virus-encoded envelope glycoprotein. There are two protein subunits on the surface of HIV, glycoprotein 120 (gp120) the exterior envelope glycoprotein, and glycoprotein 41 (gp41) the transmembrane glycoprotein (11). HIV infects CD4+ cells, due to the requirement of cell surface expression of CD4. The CD4 molecule is used as a receptor for the virus (63).

In 1986 it was discovered by Maddon et al that HIV needs a second receptor for infection into target cells, by showing that HIV-1 could not infect CD4+ cells in mice, even though these cells were engineered to express CD4 (187). The authors also showed that HIV was able to bind to the human CD4 molecule expressed on the murine cell, therefore implying that it was intact and the virus was unable to get into the cell (187). In another experiment Landau et al demonstrated that when the human CD4 molecule, expressed on the murine cell, was fused with human cells that do not express the CD4 molecule, HIV infection did occur (166). Therefore implying that the lack of HIV-1 infection on mouse cells was due to the absence of another receptor required for viral entry, this second receptor was found in human cells, thus only infecting those cells (166).

About a decade after the first reports of HIV-1 requiring a second receptor, Berger et al, in 1996, identified the seven trans-membrane receptor belonging to the chemokine superfamily of receptors (78). This receptor was then termed *fusin*, by this group of researchers, due to its

ability to fuse to the HIV-1; fusin was subsequently renamed CXCR4 (265). A further study done by Samson et al in the same year, reported the biology of a chemokine receptor by cloning, sequencing and functional characterisation, this receptor also responded to MIP-1 alpha, MIP-1 beta, and RANTES. Marc Parmenteir's group named this receptor CXCR4 (265).Within a very short space of time five independent groups showed this receptor to be the required co-receptor for HIV-1 infection (5, 43, 63, 67, 68).

The open reading frame for human CCR5 has a polypeptide sequence of 355 amino acids and is located on a single exon (210). The CCR5 protein is expressed on a number of important cells including peripheral blood-derived dendritic cells (102, 261), CD341 hematopoietic progenitor cells (263), and activated/memory CD26high CD45RAlow CD45R01Th1 lymphocytes (25, 180). In addition fresh monocytes have been shown to express low levels of CCR5 which is increased when grown by in vitro culture (5). CCR5 is the chemokine used by most circulating strains of HIV-1 for entry in host target cells.

Co-receptor usage is the main determinant of viral cell tropism. Viral strains that use CCR5 as the co-receptor are called macrophage (M)-tropic, R5 subtypes, non-syncytium inducing (NSI) HIV-1 strains, and they account for more than 95% of HIV-1 infections (189). These strains are found predominantly during the first few years of infection, and are subsequently thought to be responsible for transmission of HIV-1 infection. HIV also uses another coreceptor, CXCR4 that was the first HIV co-receptor discovered.Viruses that use CXCR4 are termed the T-tropic, X4 or syncytium-inducing (SI) strainsand usually appear later in the course of HIV disease. However it has also been seen close to HIV acquisition, and this could be due to direct or indirect transmission from an individual with advanced disease (11, 189).

Besides the R5 and X4 viruses, a further group termed dual tropic viruses require both CCR5 and CXCR4 for viral entry Into host target cells.

Following the discovery that CCR5 was identified as a co-receptor for HIV entry, a deletion of 32 base pairs in the open reading frame (coding region) was identified in the CCR5 gene in a few individuals who were exposed to the virus but remained resistant (179). This truncated form of the CCR5 gene (**Figure 2.2),** termed CCR5 Δ32, introduces a premature stop codon and the protein is not expressed on the cell surface. The protein coded by this mutant allele lacks the three transmembrane segments of the receptor. The third intracellular loop and carboxy-terminal cytoplasmic domains are the two regions missing for the mutant form and these two regions and lack the parts for binding involved in G-protein coupling (265). Individuals that have 2 copies of this mutant form of CCR5 i.e. homozygous for the 32-bp deletion are resistant to the R5 strain of the virus. Individuals that are heterozygous for the mutation i.e. they have one mutant allele of CCR5 Δ32 and one wild type, are at lower risk of acquiring HIV, however when these individuals do acquire infection they progress to disease at a much slower rate (169, 179, 265). The CCR5 Δ32 mutation is mainly found in Caucasian individuals with a allele frequency of 10% and a prevalence of 1% for individuals that are homozygous for this mutation (59, 265). Although these individuals are resistant to the R5 form of HIV they are however, susceptible to infection with other pathogens spread by sex and needles such as gonorrhoea, HCV etc (189). There have been no obvious indications of adverse health effects in persons with the CCR5Δ32/Δ32 genotype. A possible reason for this could be attributed to other chemokine receptors on immune cells that could replace the role of CCR5. However, there is an exception, whereby individuals possessing the CCR5 Δ32/Δ32 genotype tend to be more susceptible to the West Nile Virus (93).

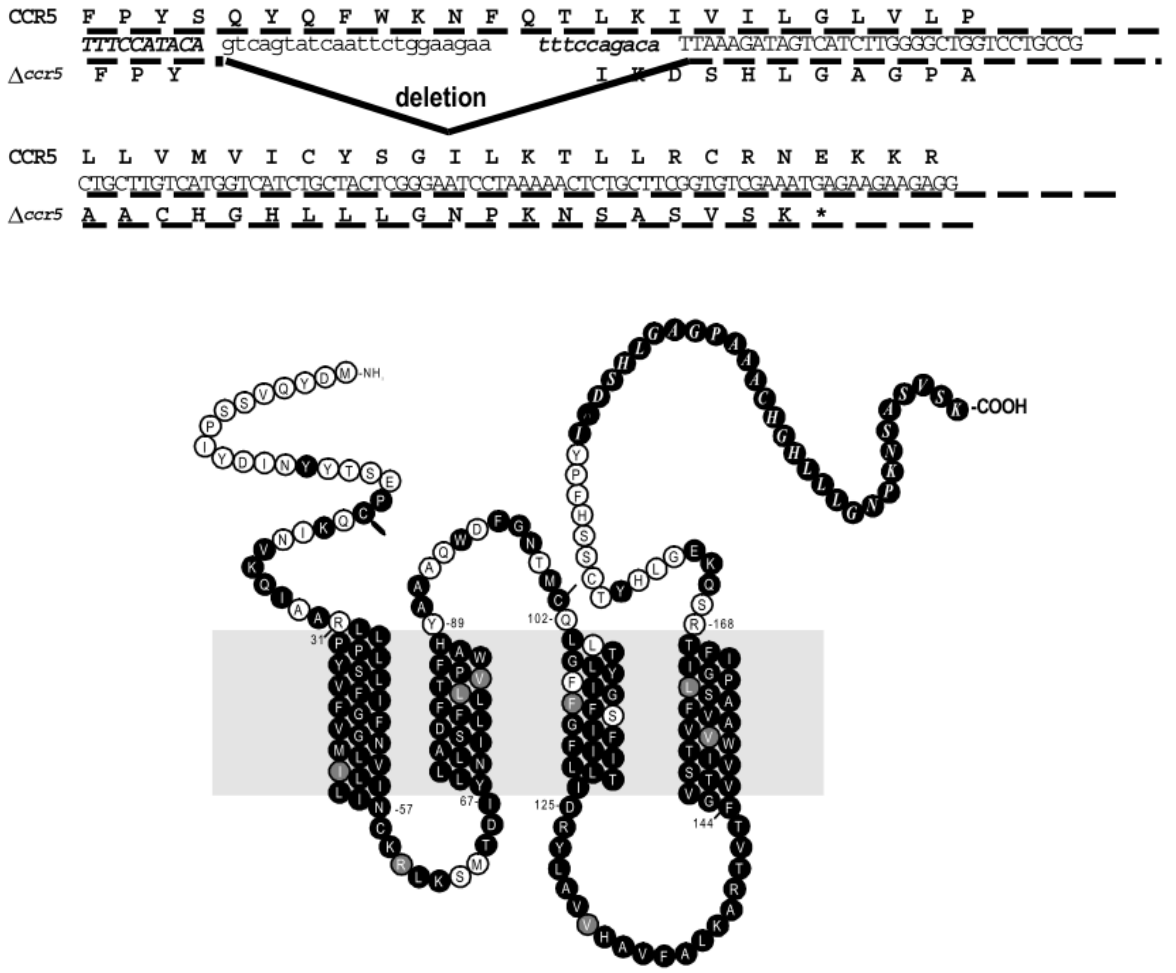


Figure 2.2: Predicted structure and amino acid sequence of the mutant form of human CCR5. The mutant protein lacks the last three transmembrane segments of CCR5, as well as the regions involved in G-protein coupling (adapted from McNicholl et al, 1997; (197)).

The presence of this mutation in predominantly Caucasian individuals indicates that this mutation may have arisen in humans after the populations diverged (169). It has also been suggested that this mutation is at least 2900 years old (134). There are also hypotheses that CCR5 Δ32 may have arisen because of the selection pressure of a historical disease such as black death (281) or smallpox (221). In a survey of 4166 subjects from 38 ethnic populations the CCR5 Δ32 allele frequencies were absent outside Europe. Stephens *et al* also argued that the Black Death of 1348 was the most promising candidate for the supposed epidemic and this then set in motion the 'enormous selective mortality' (281).

The CCR5Δ32/Δ32 genotype can account for HIV resistance in only a select few individuals, however there are many other individuals who are also resistant to the virus and lack the CCR5Δ32 mutation. In another words, due to the low prevalence of this genotype and the population specific distribution nature of this mutation, CCR5 Δ32 can only explain a small proportion of the reported resistance of the disease. This therefore led researchers to look at other mutations that are present in CCR5 in relation to HIV.

There have been several other mutations affecting the coding sequence of CCR5 (8, 41, 42, 232). Many of these mutations are very rare and also very specific to certain race groups. Some of them introduce frame-shifts that result in truncated proteins that, similarly to the Δ32 variant, fails to bring the protein to the cell surface e.g., FS299 mutation which is found with a frequency of 4% of Asians (165). Surface expression of the receptor and its ability to bind to ligands are also affected by changing the formation of the disulfide bridges which are due to the mutations C20S and C269F with a frequency of 0.3% in Caucasians and 1.4% in Asians, respectively (24, 165). The A29S mutation, involving a conservative substitution within the N-terminal region, results in failure of binding to the ligands of CCR5 and occurs with a frequency 1.5% in Africans (42). Some mutations result in undetectable or very low levels of surface receptor C178R (frequency 0.5% in Asians) (24), G106R (frequency 1.4% in Asians) (38), C101X (frequency 1.4% in Africans) (42, 165). The R60S mutation results in poor co-receptor internalization with a frequency of 1.3% in Africans (42, 165). There are also a number of CCR5 promoter polymorphisms including CCR5 59029A, CCR5 59353C, (45) and CCR5 -2459A/G (123). These CCR5 promoter polymorphisms have been reported

to associate with risk of HIV acquisition or disease progression, presumably though their impact on CCR5 expression (96, 191).

There are a few chemokine receptors that are considered as minor co-receptors for HIV-1 cell entry. One of these is the CCR2 gene, which is located on the same chromosome (3q21) (figure 3) as CCR5 has also been shown to play a role in AIDS (123, 170, 280). The V64I (G190A) substitution found in the first transmembrane region of CCR2 is associated with delayed progression to AIDS by 2-4 years when compared to in individuals possessing the homozygous wild type (280). The frequency of the protective 'A' allele in populations ranges from 8%-10% among Caucasians, 15%-17% among Chinese, 12% among North Indians (146, 280). It has also been shown that the CCR2-V64I homozygosity is associated with reduced risk of acquiring HIV in discordant couples in Thailand (146, 182). Polymorphisms in CCR5 and CCR2 are in strong linkage disequilibrium (41) therefore performing combined analyses on these genes have assisted in greatly in creating extended haplotypes.

Mummidi *et al* have created CCR5 human haplogroups, whereby a haplogroup is a collection of several distinct haplotypes that share a common ancestry (208). They have grouped CCR5 haplotypes into seven phylogenetically distinct groups HHA, HHB, HHC, HHD, HHE, HHF, and HHG (Figure 3). The HHA represents the ancestral CCR5 haplogroup, which is shared with chimpanzees. Each haplotype within a haplogroup is characterised by collection distinct signature of invariant CCR5/CCR2 polymorphisms, but may vary from each other by exclusive but rare single nucleotide polymorphisms (SNPs) (96). The authors have shown

that the CCR5 haplotypes are associated with varying rates of risk of HIV-1 acquisition as well as disease progression, and they also noticed differences between race groups (96).

The HHA haplotypes were associated with a delay in progression to AIDS and consistent with its higher prevalence in persons of African ancestry, this association was found mainly in persons of African ancestry (96). Whereas the HHC haplotypes found in Caucasians, Hispanics and Thai race groups was associated with disease retardation, particularly delayed progression to death (96, 217). In contrast the HHC haplotype in African American individuals was related with disease acceleration (96). HHE homozygosity (but not HHE heterozygosity) in Caucasians and Thais was associated with disease acceleration, particularly an accelerated progression to death; however HHE homozygosity was not associated with disease-modifying effects in African Americans. In the Indian population HHE haplotype has been implicated with susceptibility to infection and development to AIDS (147). The HHG*2 and HHF*2 haplotypes that contain the CCR5-d32 mutation and the CCR2-64I polymorphism, respectively, have been shown to be associated with slower progression among Caucasian and African American populations, respectively (96, 146).

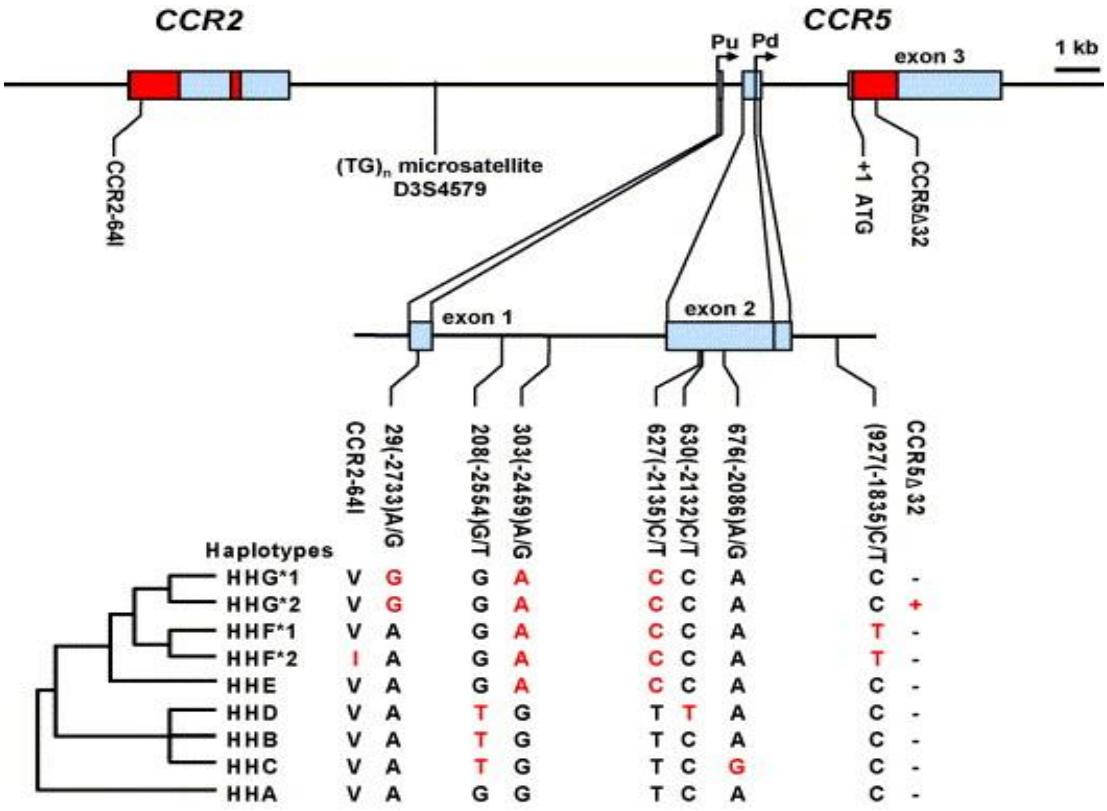


Figure 2.3: Structure of the CCR5 locus, single nucleotide polymorphisms and main haplotypes. The structure of the human 3p21 chromosomal region containing the CCR2 and CCR5 genes is schematically displayed. Blue boxes represent exons, and red boxes the coding regions of CCR2 and CCR5. The position of the upstream promoter (Pu) and the predominant downstream promoter (Pd) of CCR5 is indicated. The promoter region is enlarged and a subset of single sequence polymorphisms described in this region is displayed. Numbering is according to Mummidi et al, (208)while the numbering relative to the translational start site is provided between brackets. (adapted from Arenzana-Seisdedos, 2006).

Relevance to this thesis

Previous data has shown that the HHA haplotype plays a role in delaying progression to AIDS in HIV-positive African-Americans (96, 208). Therefore in this study we are provided with a unique opportunity to examine the effect of this and the other CCR5 haplotypes of

viral infection and disease progression in an exclusive cohort of individuals from Africa. Since not much work has been done on individuals from this ethnic background, this study will open doors to new avenues to explore as well as to help answer some questions related to the resistance of infection noticed in individuals enrolled in the cohort despite repeated exposure to HIV-1.

2.1.5. Copy Number Variation

The human genome is comprised of a number of insertions and deletions (indels) as well as duplications. These duplications can range from just a few base pairs long to entire chromosomal segments including entire gene clusters, such that the number of copies of a gene is replicated a variable number of times. This polymorphism is termed copy number variation (CNV). It has been estimated that there are 57,829 copy number variable regions (search done March 25, 2010), which consists of about 29.74% of the human genome (refer to table 2.1) (323). According to the SNP database there are 105,098.087 entered into the database and 23,652,081 Reference SNP (rs) are included in this release. The size of a SNP is 1 base pair (bp) in length, in comparison, the size of a typical CNV can range from 100bp to 3Mb (323). The presence of a CNV or SNP leads to a change in the expression of the gene or could also have no effect. There is a 17% contribution to expression variation for CNV, while SNPs account for 83.6% of the expression variation (282). When we consider the mutation rate, it is interesting to note that the point mutations have an estimated rate of about 1.8 - 2.5 x 10^{-8} per base pair per generation (152, 211, 323). Calculations of the mutation estimates for CNV are much harder to pin down as the locus-specific mutation rates were calculated using a number of different methods. The CNV mutation estimates range from 1.7 x 10^{-6} to 1.0 x 10^{-4} per locus per generation (183, 303, 323). That is an increase of 100 to 10,000 times more than point mutation rates (323). This therefore indicates that CNV is an important starting point to examine for differences in phenotypes (e.g. disease susceptibility) between individuals.

Table 2.1: Copy number variation (CNV) versus single nucleotide polymorphisms (SNP)

	CNV (Database of Genomic Variants, http://projects.tcag.ca/variation/)	**SNP (dbSNP,http://www.ncbi.nlm.nih.gov./SNP/)**
Total number	57,829 (Mar 25, 2010)	105,098,087 (Build 131)
Size	100 bp to 3 Mb	Mostly 1bp
Type	Deletion, duplication, complex	Transition, transversion, short deletion, short insertion
Effects on genes	Gene dosage, interruption, etc.	Missense, nonsense, frameshift, splice site
Percentage of the reference genome covered	29.74%[a]	<1%
Contribution to expression variation	17%	83.6%
Mutation rate	$1.8–2.5 \times 10^{-8}$ per base pair per generation	1.7×10^{-6} to 1.0×10^{-4} per locus per generation

[a] This value may be overestimated owing to the resolution limitation of the technologies (e.g., BAC cloning vs CGH arrays) used in CNV screening but also could be underestimated because of the inability to resolve smaller CNVs (1- to 20-kb range) and the limitations of the current reference genome (adapted with modifications from Zhang et al, 2009).

Studies have shown that copy number variation is as important as SNPs in determining the differences between individuals humans (17). CNV also has a contributing role in evolution. A study done by Bruder et al shows identical twins can have different CNVs, therefore indicating that CNVs can arise both meiotically and somatically (30). CNV may also be directly related to the cause of diseases when there are genomic rearrangements, this leads to the disruption of vital genes used for development; e.g., both micro-deletions and -duplications located on the chromosome 17p11.2 region. This results in syndromes

characterized by developmental delay and mental retardation (239). In the cancer field a change in copy number is involved in cancer formation and progression (306) and shown to contribute to cancer proneness (85).

The effect of CNV can be particularly complicated, such that the CNV of three separate genes (e.g. FCGR3B, C4 and *CCL3L*) found in different chromosomal locations can be both independently linked to the susceptibility for example in the case of systemic lupus erythematosus (SLE), a complex polygenic autoimmune disease characterized by autoantibody production, immune complex deposition, and inflammatory damage to multiple organ systems (314). The complement component C4 (with isotopes C4A and C4B) are located on chromosome 6 and has been shown to have inter-individual gene CNV and this variation lead to different susceptibilities to SLE (320). The activatory Fc receptor for IgG, FCGR3 (also known as FCGRIII) located on chromosome 1, also shown to have inter-individual gene CNV, which also results in different susceptibilities to SLE (4).

One of the most studied CNV regions is the human region 8p23.1; this region contains the defensin cluster which has two sub-clusters within this segment; namely the alpha defensin and beta defensin. The 2Mb defensin locus has been shown to have variable copies of both alpha and beta defensin genes (126, 190). A study done by Fellermann et al. shows that a variable number of copies of human beta-defensin 2 (*HBD-2*) affects the acquisition of an inflammatory bowel disease, Crohn disease (CD) (76). Individuals that posses lower copies ($\leq$3) of *HBD-2* showed a significantly greater risk of developing colonic CD when compared to individuals with higher copies ($\geq$4) of *HBD-2* (76). The converse shows that an increase

number of copies of the beta defensin genes results in an increase in psoriasis, a chronic inflammatory dermatosis (127).

2.1.5.a. Effect of CNV on HIV

In this study we focus on HIV as the disease model, therefore we reviewed literature on the effect CNV plays on HIV. However to explain how CNV effects the replication of HIV we need to first explain how HIV uses the chemokine receptor to enter the cell. The mechanism of HIV entry into the target cell is a multi-step process. Firstly the CD4 molecule acts as a docking station, the HIV particle more specifically the HIV-1 gp120 protein latches onto the CD4 molecule and this allows the virus while hanging atop the cell to make contact with the large seven-transmembrane cell-surface receptor, CCR5. Thereafter, the CCR5 molecule which is found to be floating around the fluid-like cell surface membrane, floats unto and binds the CD4-snagged HIV (223). The bound HIV-CD4-CCR5 causes the HIV gp41 protein to change shape and thus "drill" into the cell membrane. Once this step is completed the HIV particle then fuses with the cell membrane and hence all the viral components enter into the cell.

The CC chemokine ligand 3-like 1 (*CCL3L*) gene product is the most potent CCR5 agonist, and the most efficient inhibitor of R5 HIV entry. *CCL3L* competes with HIV to bind onto CCR5 (97, 223). Hence the more copies of *CCL3L* available, the greater chance of it binding to CCR5 rather than HIV (Figure 2.4). As indicated in figure 2.4 Mip-1α, Mip-1β or RANTES binds onto CCR5 while SDF1 binds onto CXCR4 (223). The free chemokines

produced are a result of inflammatory responses and thus would sterically inhibit HIV-1 gp120 binding and limit entry through CCR5 (186).

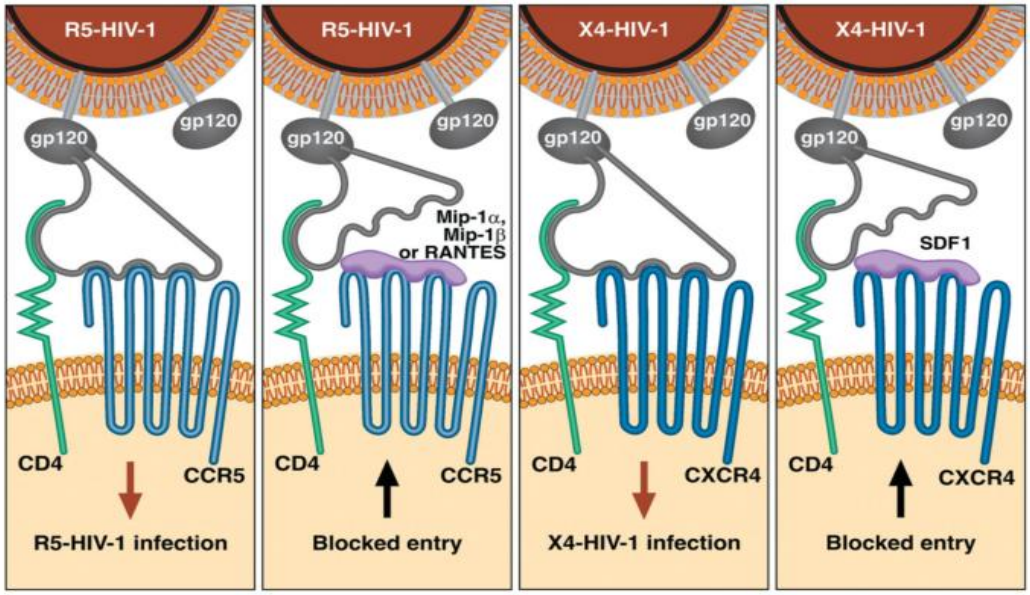


Figure 2.4: Illustration of how HIV-1 enters cells, suggesting candidate ARGs for inspection. (Adapted from **O' Brein and Nelson, 2004**)

Thus having explained how HIV needs the chemokine receptor for viral entry, the chemokine CNV role in HIV-1 life cycle is described. There is a CNV anomaly found in the human chromosome 17q region (Figure 2.5a). The 16 genes from the CC chemokines cluster are localized to a 2.06 Mb interval at 17q11.2–q12 on genomic contig NT_010799 (52). This region has been found to be a hot spot for segmental duplications, and plays a role in driving immunity to infectious diseases like HIV. *CCL3L* other names, MIP-1αP and LD78b also including *CCL4L*1 (MIP-1b-like) acts as a natural ligand for CCR5 (200, 293).

The non-allelic copies (*CCL3L* and *CCL4L*1) represent the duplicated isoforms of the genes encoding CCL3 and CCL4, respectively. The two non-allelic CCL3 isoforms encode highly

related proteins. CCL3 and *CCL3L* display more than 90% identity. Variable numbers of copies per diploid genome are a result of the duplications and hence the copy number of *CCL3L* and *CCL4L*1 varies among individuals (14, 97, 200, 293). Within all the human chemokine genes the noticeable characteristic of *CCL3L* and *CCL4L* is these genes are found to contain variable number of copies in the human genome.

There is also evidence that the human *CCL3L–CCL4L* region contains complex homologous recombination events. This can be demonstrated when a high resolution CNV analysis is performed, the data reveals there is significant architectural complexity in the *CCL3L–CCL4L* region, and this is clearly shown by smaller CNVs embedded within larger ones and interindividual variation in breakpoints (Figure 2.5b) (40, 52, 231). Colobran et al states that 'one of the consequences of this complexity is that individuals may vary not only in the total copy number of *CCL3L* and *CCL4L* genes, but also their individual components' (52).

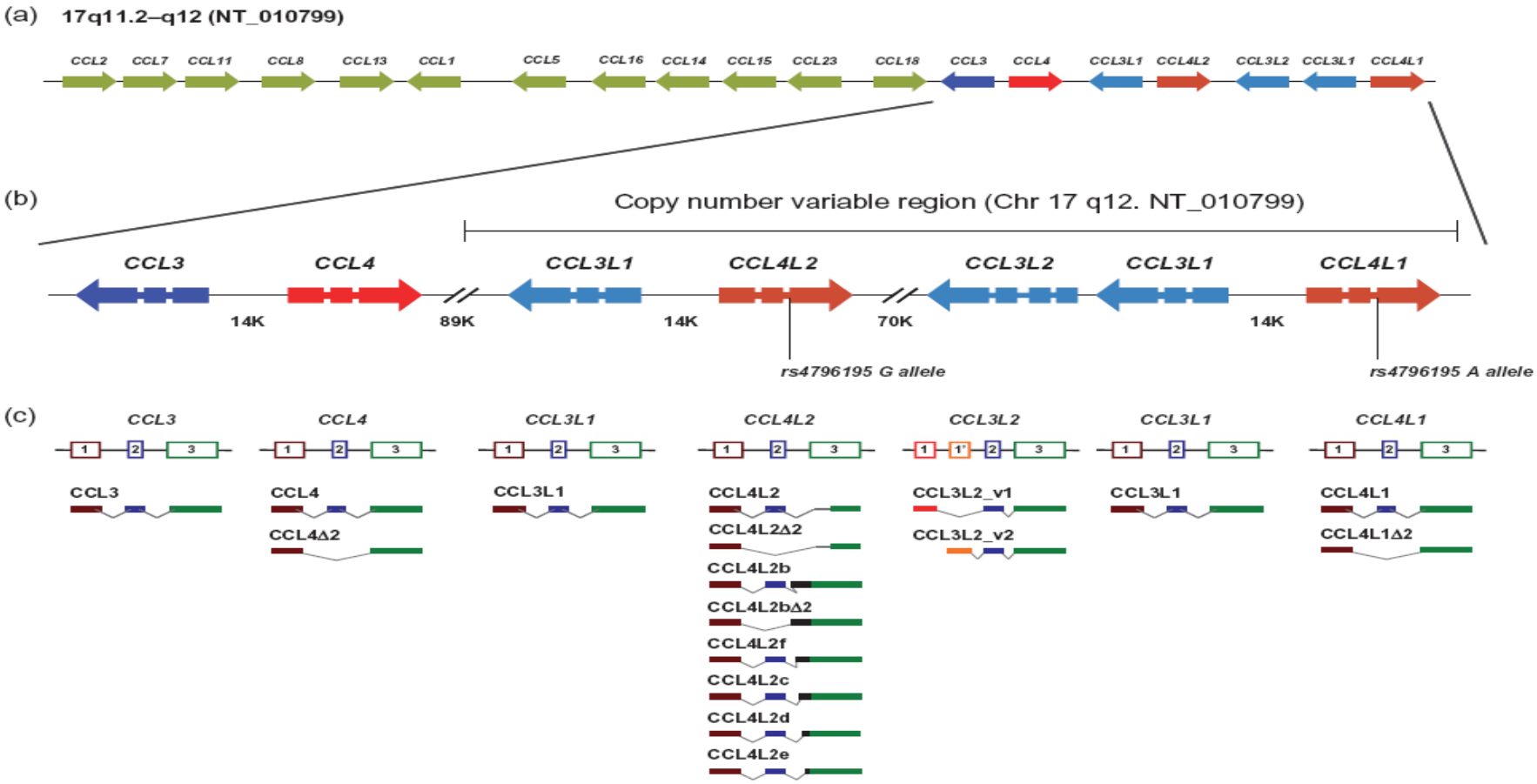


Figure 2.5. Genomic organization and mRNA products of human *CCL3–CCL4* and *CCL3L–CCL4L* genes. (a) Map of the CC chemokine cluster in the 17q11.2–q12 region, based on the genomic sequence NT_010799. The orientation of each gene is shown by an arrow. (b) Genomic organization of human *CCL3–CCL4* and *CCL3L–CCL4L* genes based on the genomic sequence NT_010799. Distances between genes are

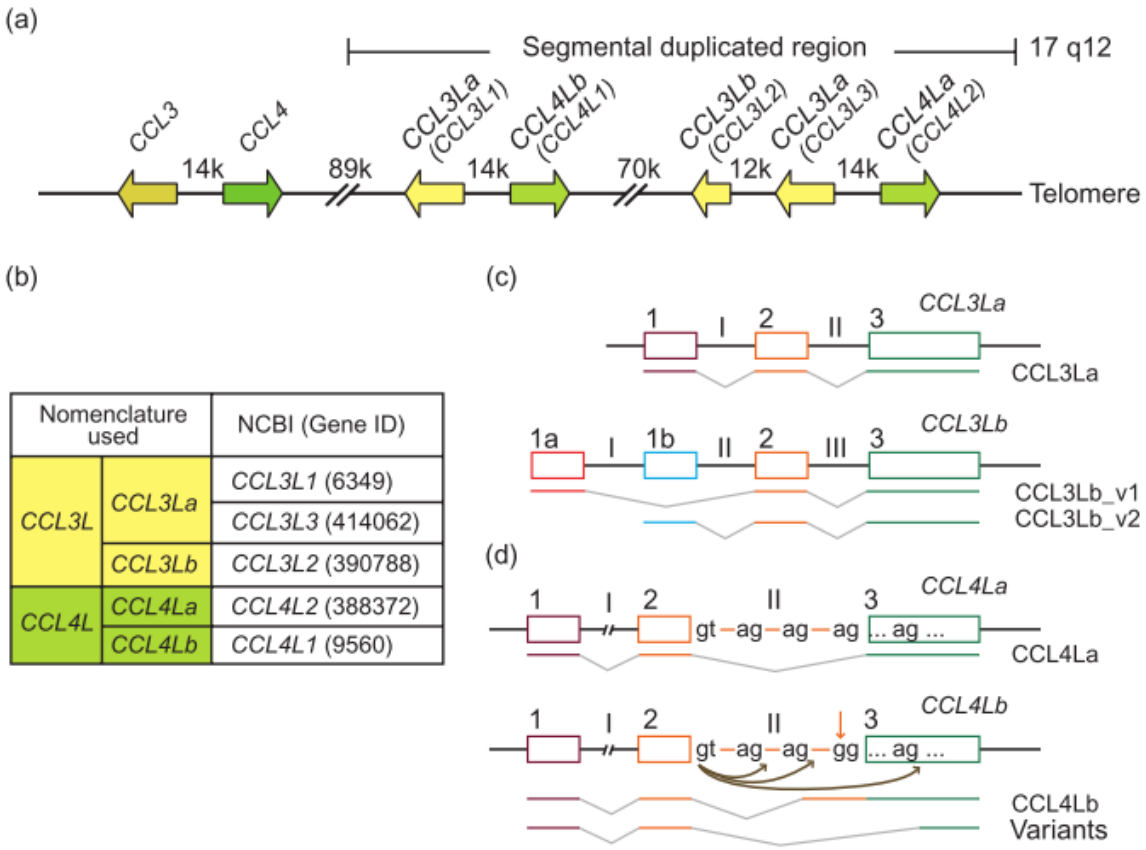
expressed in Kb. The nucleotide change [single nucleotide polymorphism (SNP) rs4796195] that leads to *CCL4L1* (A allele) or *CCL4L2* (G allele) is shown. (c) Transcription pattern of human *CCL3–CCL4* and *CCL3L–CCL4L* genes. mRNAs derived from each individual gene are shown (adapted from Colobran et al, 2010).

There are a number of genes included in the *CCL3L–CCL4L* cluster and the official symbols for these genes are based on the public human genome sequence. The *CCL3L* gene has three copies while the *CCL4L* gene has two copies. Each of these genes has a completely random or by chance occurrence of possessing the mentioned number of copies for each gene. The *CCL3L* and *CCL4L* genes have been allocated their numbering according to their position relative to the centromere and telomere regions (52). The *CCL3L* genes are termed according to NCBI, *CCL3L1* (GeneID: 6349), *CCL3L2* (GeneID: 390788) and *CCL3L3* (GeneID: 414062). There are two *CCL4L* genes and are termed as *CCL4L1* (GeneID: 9560) and *CCL4L2* (GeneID: 388372).

A study done by Shostakovich-Koretskaya et al. recently helped clarify the nomenclature of these genes (figure 2.6) (275). The *CCL3L*3 is a separate gene from the *CCL3L1*, although each gene has three exons that encode the same amino acid (200, 202, 275). Shostakovich-Koretskaya *et al* has denoted *CCL3L*a as the addition of *CCL3L1* and *CCL3L*3 (figure 2.5c and 7). *CCL3L*2 (known previously as LD78g or GOS19-3) contains only two exons which are identical to those found in exons 2 and 3 in *CCL3L1* and *CCL3L*3, but lacks the first exon and hence was thought to be a pseudogene (200, 202, 275). Shostakovich-Koretskaya *et al* has identified 5' novel exons for *CCL3L*2 (which is denoted as *CCL3L*b). The authors have shown that the *CCL3L*b gives rise to alternatively spliced transcripts (*CCL3L*b-v1 and *CCL3L*b-v2), which contains chemokine-like domains but are not predicted to encode classical chemokines (275) (figure 2.5c and 2.6). The two genes *CCL3L*a and *CCL3L*b, as

explained by the authors, produce two distinct kinds of transcripts; *CCL3L*a encodes the classical chemokines while *CCL3L*b encodes the alternatively spliced transcripts with novel 5' exons (275).

There are two *CCL4L* genes: *CCL4L*b (*CCL4L*1 – with GeneID: 388372) and *CCL4L*a (*CCL4L*2 – with GeneID: 9560) (figure 2.5c and 2.6). *CCL4L*a and *CCL4L*b have been denoted by Shostakovich-Koretskaya *et al* (275), and these genes have the same exonic sequences i.e. they share 100% sequence identity in the coding regions (figure 2.6). However, a single A to G mutation in the splice acceptor site in the intron 2 of *CCL4L*b results in an abnormal transcript formed in comparison to *CCL4L*a. The formation of the abnormal transcripts occurs with the retention of parts of the intron 2 or a partial loss of exon 3. Similar to *CCL3L*a and *CCL3L*b the two genes *CCL4L*a and *CCL4L*b produce two distinct kinds of transcripts. *CCL4L*a produces transcripts that encodes a classical chemokine, whereas *CCL4L*b encodes abnormally up to 11 spliced mRNA transcripts (275).



Nomenclature used		NCBI (Gene ID)
CCL3L	*CCL3La*	*CCL3L1* (6349)
		CCL3L3 (414062)
	CCL3Lb	*CCL3L2* (390788)
CCL4L	*CCL4La*	*CCL4L2* (388372)
	CCL4Lb	*CCL4L1* (9560)

Figure 2.6: Chromosome 17q12 segmental duplication, chemokine gene copy number variation nomenclature and correlation between copy numbers of *CCL3L* and *CCL4L*. (a) Schematic representation of CCL3, CCL4, *CCL3L* and*CCL4L* genes. Arrows indicate the orientation of each gene. Shown on top is the distance between the indicated genes. (b) indicates the nomenclature previous literature used. (c) and (d) Schematic representation of the exon–intron structure of (c) *CCL3L* and (d) *CCL4L* genes. In(d), the downward pointing arrow indicates the A→G transition that leads to the generation of aberrantly spliced *CCL4L*b transcripts, and the splicing patterns of these mRNA species are indicated by the curved arrows. Thus, a *CCL4L* copy, designated here as*CCL4L*a, has the intact AG intron–exon splice sequence and is predicted to transcribe an intact full-length *CCL4L* transcript (adapted from Shostakovich-Koretskaya *et al*, 2009).

Of note the difference between CCL3 and *CCL3L* mature proteins is three amino acids. The latter protein has a proline (P) located in position 2 as opposed to the serine (S) located in CCL3 also noticed two changes are reciprocal S/G (glycine) swaps in the position between cysteines 3 and 4 (figure 2.7) (52). When we consider CCL4 and *CCL4L*1 it is noticed that the mature proteins have only one difference in the amino acid sequence, and this is a

conservative Serine to Glycine change at amino acid 47 of the mature protein (figure 8) (52). An interesting observation is that both the *CCL4* and *CCL4L1* genes produce alternatively spliced mRNAs which has been shown to lack the second exon, and hence this creates the CCL4Δ2 and *CCL4L*1Δ2 variants (figures 6 and 8) (49, 52, 203). These proteins of 29 amino acids retain only the first two aa therefore lacking the defining CC's to be considered as chemokines and vital for intramolecular disulphide bonding, and these genes therefore lose their ability to bind onto CCR5 and thought to possess no CCL4/*CCL4L*1 activity (49, 52).

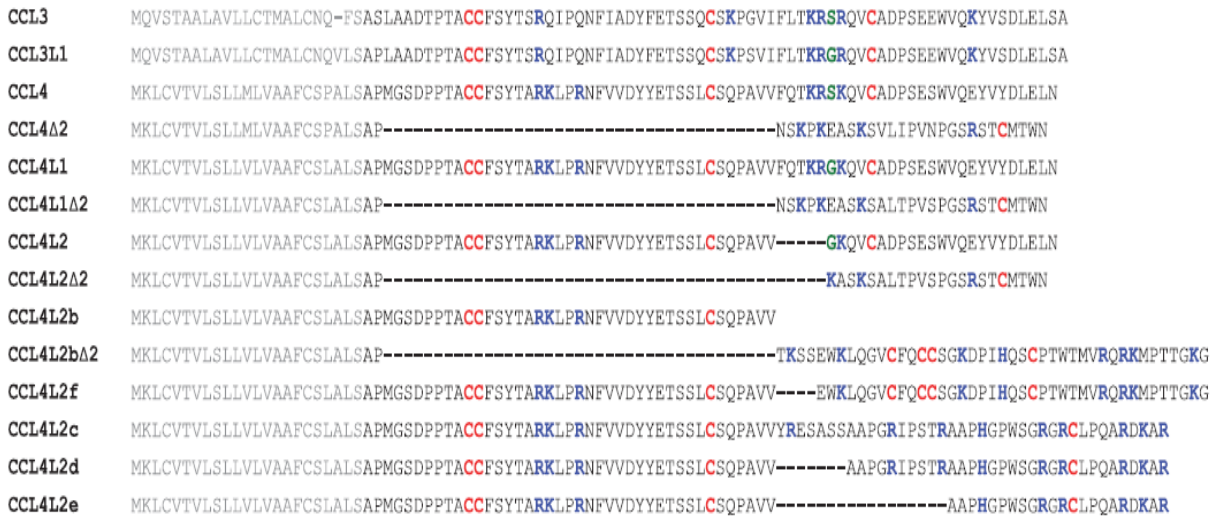


Figure 2.7: Alignment of human *CCL3–CCL4* and *CCL3L–CCL4L* derived proteins. Signal peptides are depicted in grey. Cysteines are depicted in red. Basic amino acids, which are involved in the binding of chemokines to the glycosaminoglycans are depicted in blue. The S/G swap shared between CCL3–*CCL3L* and CCL4–*CCL4L*1/L2 proteins is depicted in green (adapted from Colobran et al, 2010).

The effect of CNV on gene expression has been a crucial question that researchers have been baffled with since the first reports of CNV. The impact of CNV on gene expression varies either positively or negatively based on the gene in question (282). When we examine the *CCL3L* gene, it is noticed that the *CCL3L* gene CNV promotes the production of *CCL3L* protein and mRNA expression, more specifically an increase in the *CCL3L* copy number

results in a strong positive correlation with the *CCL3L* protein and mRNA expression (52, 97, 135, 298).

CCL4L copy number effects on *CCL4L1* mRNA or protein expression has not been fully understood as yet. Initial work has been performed by Townson et al in 2002, whereby the authors state that they were unable to measure a positive correlation between the *CCL4L* copy number and the *CCL4L1* mRNA. What they did find is that subjects that have a single copy of the *CCL4L* gene were found to also have a lower expression of the *CCL4L1* protein when compared to subjects with higher doses of *CCL4L* copy number (293). However, the authors did not take into account the *CCL4L2* variant, which has been shown to produce variations in transcripts and proteins compared to the *CCL4L1* gene. Therefore the lack of correlation that they report could be due to them not quantifying the *CCL4L1* and *CCL4L2* genes independently (293). A further study done by Melzer *et al*. shows a SNP found in close proximity to the *CCL4L1* gene affects the production of the *CCL4L*1 protein, due to the SNP and the *CCL4L* CNV being in linkage disequilibrium (199). Further studies need to be conducted to determine the impact of the *CCL4L* copy number effect on the *CCL4L1* gene and the *CCL4L2* variant mRNA or protein expression.

Gonzalez *et al* has shown that different ethnic groups have significantly different doses of *CCL3L* copy numbers (97). The number of copies range from 0 to 10, with higher numbers in African populations. It was noticed that individuals from African descent have a median *CCL3L*-continaing segmental duplication of four to five (97, 198). A median of 2 copies was observed in Caucasians (97) and Indian individuals (212), a median of 3 copies was found in Hispanic individuals (97) and 5 copies median was observed in Japanese individuals (272).

Gonzalez *et al* have also screened the *CCL3L* copy numbers in chimpanzees, as the animals also follow the differences in *CCL3L* gene copy number from animal to animal. The median copy number, of nine, found in chimpanzees was much higher than in humans (97). The authors hypothesized that the elevated *CCL3L* copy numbers might be the factor that confers HIV resistance. While the findings revealed that the lower number of copies associates with a reduced susceptibility to HIV/AIDS, the authors however caution that the number of copies does not determine HIV susceptibility, but rather the number of copies that an individual has in relation to individuals with the same ethnical background. Therefore the individuals at greatest risk of acquiring HIV are those that have *CCL3L* gene doses lower than the population-specific average (97).

The *CCL3L* CNV was also examined in animals for their associations to rate of progression to AIDS. A study, done by Degenhardt et al, tests the effect of *CCL3L* copy number on disease progression in Rhesus macaques (61). The authors demonstrate that a low *CCL3L* copy number is associated with a faster rate of disease progression to experimental AIDS when the animals were challenged experimentally with SIV. Since the Indian rhesus macaques are known to progress at a faster rate when compared to the Chinese macaques (177), the authors also further suggest that the noticed lower *CCL3L* copy number in Indian rhesus macaques may implicate the faster progression to AIDS in Indian when compared to the Chinese macaques (61). Chimpanzees have also been shown to be affected with variation in *CCL3L* copy numbers, the animals that possess higher copies do not develop to AIDS (52).

In a South African cohort, Kuhn et al have shown that infants with a higher *CCL3L* copy number was found to have a reduced risk of HIV-1 transmissions from their HIV infected mothers (156). Hence, indicating the importance of *CCL3L* copy number in perinatal HIV-1 transmission. Further studies performed have shown an association or lack of an association with disease, table 2.2 summarises these findings.

Table 2.2. Studies showing association of copy number variations for CCL3L-CCL4L genes with disease.

Disease	Gene	Population	N =	Association	Ref.
HIV	*CCL3L*	WHMC (EA,AA,HA) and Argentinean children	4308	Disease association	(97)
	CCL3L	WHMC (EA,AA,HA), MGH, USCF, UCSD	4329	Clinical aspects	(3, 66, 157)
	CCL3L	Ukraine	298	Disease association and clinical aspects	(275)
	CCL3L	South Africa	505	Disease association and clinical aspects	(156, 198, 271)
	CCL3L	Estonia	374	Disease association	(133)
	CCL3L	Japan	300	Disease association	(213)
	CCL3L	Rhesus macaque	57	Clinical aspects	(61)
	CCL3L and *CCL4L*	AA, HA, EA	411	No disease association and clinical aspects	(272)
	CCL3L	Euro-CHAVI, TACC, MACS	2982	No disease association and clinical aspects	(21, 298)
Type 1 diabetes	*CCL3L* and *CCL4L*	British	12,625	No disease association	(79)
	CCL3L	New Zealand (Caucasian)	534	No disease association	(196)
Hepatitis C	*CCL3L*	Germany	464	No disease association	(105)
Systemic lupus erythematosus	*CCL3L*	San Antonio SLE cohort, Colombian SLE cohort, Ohio SLE cohort, Colombia	1639	Disease association and clinical aspects	111, 112
Rheumatoid arthritis	*CCL3L*	New Zealand (Caucasian)	1767	Disease association	(196)
	CCL3L	British (Caucasian)	557	No disease association	(196)
Lung transplantation acute rejection	*CCL4L*	Spain (Caucasian)	161	Clinical aspects	(50)
Kawasaki disease	*CCL3L*	USA	164 + parents	Disease association	(32)
Primary Sjogren's syndrome	*CCL3L*	Colombia	470	Disease association	(188)

WHMC-Wilford Hall Medical Center, MGH-Massachusetts General Hospital, UCSF-University of California San Francisco, USCD-University of California San Diego, TACC-Tri-Service AIDS Clinical Consortium, MACS-Multicenter AIDS Cohort Study, EA-

European American, AA-African American, HA-Hispanic American. Adapted from Colobran, et al., 2010

The copy number of *CCL3L* and *CCL4L* are shown to be highly correlated, with *CCL3L* having a greater copy number than *CCL4L* (275). The distribution patterns of *CCL3L*a and *CCL4L*a has been shown to be very similar to those of the total copy numbers of *CCL3L* and *CCL4L*. The authors also show a strong negative correlation between the copy number of *CCL4L*a and *CCL4L*b. They therefore state that those who had higher 'functional' chemokine-encoding *CCL4L*a copies than *CCL4L*b copies progress better in HIV disease than those individuals that only possessed *CCL4L*b copies (275). In other words a higher gene content of *CCL4L*b or a lower content of *CCL3L*a and *CCL4L*a increases the risk of transmission and an accelerates disease progression (275).

2.1.5.b. Controversies associated with *CCL3L* CNV

Recently a few studies have disputed the previous findings of *CCL3L* CNV having an effect on HIV-1 infection, viral load or disease progression (21, 79, 298). Urban, et al published one of these journal articles and in their attempt to explain their result they remark 'measurement of *CCL3L* copy number variation appears highly susceptible to systematic biases related to the preparation and quality of DNA samples' (298). The authors also make mention of 'batch differences in input DNA amounts between cases and controls can lead to biased copy number estimates by the real-time PCR method used' (298). Of special note Urban and colleagues state that they find a converse result to the one previously reported by Gonzalez et

al, that low *CCL3L* is detrimental for HIV acquisition and disease progression (97), Urban et al demonstrate they find a significant association with higher copy number among HIV infected cases compared with controls, this is however before the dilution of the DNA samples into an appropriate range (298).

Bhattacharya et al indicate that the findings from their research "do not support a relationship between the population specific *CCL3L* gene copy number and HIV/AIDS susceptibility or the response to HAART as previously reported" (21). The authors also further state that despite using a "similar population, used the same method for determining the *CCL3L* gene copy number, used *CCR5*Δ32 as a proxy for the *CCR5* promoter polymorphisms and used a study powered to detect the postulated effect" they were unable to replicate the previous results (21).

A further study, by Field et al, examined the experimental aspects of *CCL3L* copy number quantification in depth (79). The authors examined DNA samples from 5,771 British individuals with type 1 diabetes and 6,854 geographically matched controls for *CCL3L* variation. Using two different assays the authors explain that they notice differences between the results generated by *Taq*Man assay and by an alternative assay called the paralogue ratio test (PRT) (79). When the 5,121 samples were used to compare the different PRT and Quantitative Polymerase Chain Reaction (qPCR) experiments, the results show that 64% were consistent between the two methods, while 25% having an additional copy of *CCL3L* using the qPCR techniques when compared to with PRT. The qPCR assay demonstrated a greater likelihood of possessing higher copy numbers compared to PRT assays; this shift observed the authors explain as an artifact of the qPCR primers which may also be binding a

CCL3L pseudogene, *CCL3L2*, as well as *CCL3L* (79). The authors conclude by indicating that the PRT technique should be used for a more accurate CNV analysis.

In direct response to the above studies He et al, (the study team that initially reported the effects of *CCL3L* CNV on HIV-1 infection, viral load or disease progression) explains the flaws in each of the above studies and point out why the respective authors could not reproduce the results (112). In the Urban et al study, the authors made two major modifications that subsequently resulted in failure to reproduce the previous results. The first of the modifications to the original assay is the change of tetramethylrhodamine (TAMRA) as the quencher, Urban and colleagues used another quencher, minor groove binder (MGB**)**, He et al state this different quencher raises the melting temperature of the probe by 8 ºC higher than the optimal temperature (112, 298). The second modification that Urban and colleagues performed was the quality and the amount of input DNA the authors used variable amounts of input DNA (up to 100 ng) and constructed standard curves of tenfold dilutions from 100 ng to 1 pg, this was in direct contrast to the specifications given in the original assay by Gonzalez et al, (97) requires a fixed amount of high-quality input DNA with a range between 2 and 10 ng, and, consequently, they used qPCR standard curves comprising twofold dilutions from 25 to 0.78 ng (112). He et al performed a comprehensive analysis comparing the two different techniques and show that the conditions from the original Gonzalez et al assay yield more accurate results, therefore stating "that the use of the MGB assay and variable input DNA together may have compromised the results of the association studies" (112).

In response to Bhattacharya and colleague's work (21), He et al indicated that the authors have also modified the qPCR assay (112). The quencher used for their assay is Quencher Series 7 (QSY7), this probe decreases the optimal melting temperature by 8 °C. He et al did not make direct comparisons between the QSY7 assay and the TAMRA assay (112). However, the authors state that this change in the quencher used will affect the assays performance.

He et al also noted, that the Bhattacharya et al study (21) did not adequately account for genetic-epidemiological factors that are pertinent to the full interpretation of their results (112), and indicate three instances, which is briefly summarised here to ensure that researchers are mindful of the delicateness of the *CCL3L* CNV assay and also to guarantee success when replicating the original assay (112). Firstly, since high *CCL3L* gene copy number (GCN) has been previously shown to be associated with a slower rate of progression to AIDS and death, therefore when the population level is considered it is noticed that there is a gradual enrichment over time of HIV-positive subjects with a protective high *CCL3L* GCN (97). With regards to this, individuals that are sero-positive for approximately 7 years, it is noted that after this time the prevalence of *CCL3L* GCN in surviving HIV-positive individuals approaches that of HIV-negative individuals (97, 112). Therefore, He et al argued that since Bhattacharya and colleagues examinined a mainly sero-prevalent cohort (more than three quarters of the population is HIV infected), it may be assumed that the authors have found similar *CCL3L* GCN distributions in both the HIV-infected and uninfected subjects (21, 112).

The second instance that Bhattacharya et al (21) have not fully accounted for as explained by He et al (112), is the number of AIDS events during the 12 year study period. A total of 53 individuals (13%) HIV positive from a cohort of 396 individuals investigated in the Bhattacharya et al study (21) and compared to the Gonzalez et al study (97), which has the similar study period however about 40% of the subjects were infected of the 1, 132 subjects examined. Therefore, He and colleagues stated 'Given the limited AIDS events and small sample size, it is not surprising that strong associations between *CCL3L* GCN and AIDS were not detected' (112).

Thirdly, He et al state that failed detection of the association between *CCL3L* GCN or conjoint genotypes of *CCL3L* GCN and the *CCR5*Δ32 allele with highly active antiretroviral therapy (HAART) responses in the Bhattacharya et al study (21) is possibly due to false negatives (112). Despite the lack of replication found between the *CCR5*Δ32 heterozygosity and improved CD4+ T cell and viral load responses during HAART by the Bhattacharya et al study many other studies have successfully shown this (106, 117, 144, 299, 319), surprisingly this includes individuals from the same study group used by Bhattacharya et al (21, 112). Thus He et al clarified why Bhattacharya and colleagues were unable to replicate their previous findings.

In the third study that He et al (112) responds to, the authors demonstrate a flaw in Fields et al study (79), whereby the authors using the PRT assay only quantify two of the three *CCL3L* genes. The exclusion of a *CCL3L* gene from the analysis could be directly related to the discrepant results observed when comparing the PRT and *Taq*Man qPCR assays. The gene that was excluded from the analysis was the *CCL3L2* gene, which was previously considered

to be a pseudogene, however, a recent report (275) and unpublished data indicate otherwise. Furthermore, Field et al state that the *CCL3L* GCN departed from Hardy-Weinberg equilibrium (HWE), and thus imply that the qPCR assay does not provide robust GCN estimates (79), in response to this He et al state the authors 'Hardy-Weinberg estimates would be valid if the CNVs of the individual *CCL3L* genes were both duplicated and distributed randomly on each chromosome, permitting accurate computation of copies per chromosome' (112).

Using *CCL3L* gene copy number data from 655 parent-child trios, He et al reported it is very difficult to predict the chromosome phase of *CCL3L* GCN, even in individuals with two copies, but especially in those with >2 copies (112). Furthermore, when the expectation maximization algorithm described by Fields *et al* (79), was used by He et al, the authors indicate 'that Hardy-Weinberg estimates in 991 normal donors departed from equilibrium only when they were computed for a *CCL3L* GCN of >2 but not ≤2' (112). Upon examining the data from the 655 parent-child trios, He et al showed 62 trios for whom the *CCL3L* GCN phase of each family member could be resolved, but the expectation maximization algorithm used by Fields *et al* (79) misclassified the *CCL3L* GCN phase in 13% of these 186 subjects (112).

In the final point that He et al make, the authors briefly describe the extensive architectural complexity found in the *CCL3L-CCL4L* region. Firstly, the authors show that there is varying levels of correlations of the GCN between the *CCL3L* and *CCL4L* genes, which is seen in data from the Human Genome Diversity Cell Line Panel (HGDP)-CEPH samples. Of the 55 populations examined, some populations display a strong inverse correlation between the *CCL4L1* and *CCL4L2* GCNs. Also as mentioned above there is a high degree of complexity

in this region, which shows smaller CNVs, embedded within larger ones and with interindividual variation in breakpoints (112). Accounting for complexity when trying to interpret association studies is extremely vital to understanding any study, but here the authors demonstrate that without examining this complexity, the results can yield erroneous results as demonstrated by the three studies (21, 79, 298). Including the previously considered pseudogene *CCL3L2* copy number is also very important for full interpretation of copy number studies.

Relevance to this thesis

Despite the controversy raised primarily by three groups regarding the role of *CCL3L* and HIV-AIDS associations, there have been several studies from independent groups around the world that have showed *CCL3L* CNV having an effect on HIV-1 infection, viral load or disease progression (as shown in table 2.2). Here in studies related to my thesis (97, 3, 66, 157, 275, 156, 198, 271, 133, 213, 61), we examine the effect of *CCL3L* and *CCL4L* GCN with HIV-1 susceptibility in a cohort of highly exposed female sex workers. Furthermore, we also give further insight as to the complexity displayed in the interested gene region. In this thesis we plan to investigate the balance between the doses of *CCL3L* and *CCL4L* genes and their respective components (*CCL3La* and *CCL3Lb*) and (*CCL4La* and *CCL4Lb*) conferring either protective or detrimental effects that impact HIV-1 infection and on the predictors of AIDS risk i.e. CD4 count and viral load.

2.1.6. Duffy Antigen Receptor for Chemokines

Duffy Antigen Receptor for Chemokines (DARC) is also called Glyco-protein D or Duffy Antigen, is another chemokine receptor that has recently gained much interest due to its role in HIV disease (55). The Duffy blood group antigen was first recognised in a haemophiliac patient (named "Duffy") who displayed high antibody titres against the erythrocytes of several individuals (55). The Duffy antigen is located on chromosome 1q21-22 (146). *DARC* is mainly expressed on red blood cells (RBCs) but also found on vascular endothelial and neuronal cells (109). Similar to other members of the G-protein coupled chemokine receptors, *DARC* is organised in seven transmembrane domains. *DARC* has the ability to bind to a wide array of proinflammatory chemokines and has been shown to bind with strong affinity to both CC and CXC chemokine families (109).

The Duffy antigen is an erythrocyte receptor for malarial pathogens such as *Plasmodium vivax* and *Plasmodium knowlesi* (259) (figure 2.9a). Individuals that lack *DARC* on erythrocytes render the individual resistant to *Plasmodium vivax* and *Plasmodium knowlesi*. The vast majority of individuals from African descent lack *DARC* expression, in contrast to Caucasian individuals whereby *DARC* is uniformly expressed (259). Therefore indicating its role in evolution in malaria rich areas of Africa. The lack of *DARC* expression on RBC surface is due to a polymorphism in *DARC* promoter region (figure 2.8). The T-46C mutation disrupts a GATA binding motif (292) and thus the direct result is the loss of expression of DARC on RBCs, which renders the individual resistant to malaria. This mechanism is very similar to that of the CCR5 delta 32 polymorphism (figure 2.8b and c), which when present in the homozygous form protects the individuals from HIV-1 infection. Duffy-negative

individuals are able to express *DARC* in non-erythroid tissues, because GATA binding is not a requirement for *DARC* transcription (292).

The DARC protein has also been shown to play a role in the life cycle of HIV (114, 161), hence its inclusion here in this thesis. HIV has the ability to bind onto RBC's (figure 2.8a) and this is understood to happen by two distinct mechanisms. The first mechanism of HIV-1 binding onto RBC's described by Lachgar and colleagues (161), is mediated by DARC (114). The HIV bound RBC has been detected in HIV infected individuals undergoing HAART with very low to undetectable levels of viral load (122). The second mechanism thought to play a role in HIV binding onto RBC's is an indirect association of HIV/anti-HIV immune complexes and complement factor c3b binding to CR1 (CD35) (15, 128).

Since the T-46C mutation is vital for protection against malaria (292), a study done by He *et al*, further evaluated the impact of DARC on HIV-AIDS pathogenesis, the authors thought to determine the effect of the T-46C polymorphism in *DARC* on HIV-AIDS susceptibility. The results show that the T−46C polymorphism in *DARC* affects susceptibility and disease progression to HIV-AIDS in African Americans (114). They show that the -46C/C genotype (*DARC* negative phenotype) increases the likelihood of HIV infection by 40% when compared to those lacking this genotype (114). They also consider the risk of acquiring HIV associated with possession of the *DARC* −46C/C in subjects of African descent. An estimated 11% of the HIV burden in Africa may be due to -46C homozygosity (114).

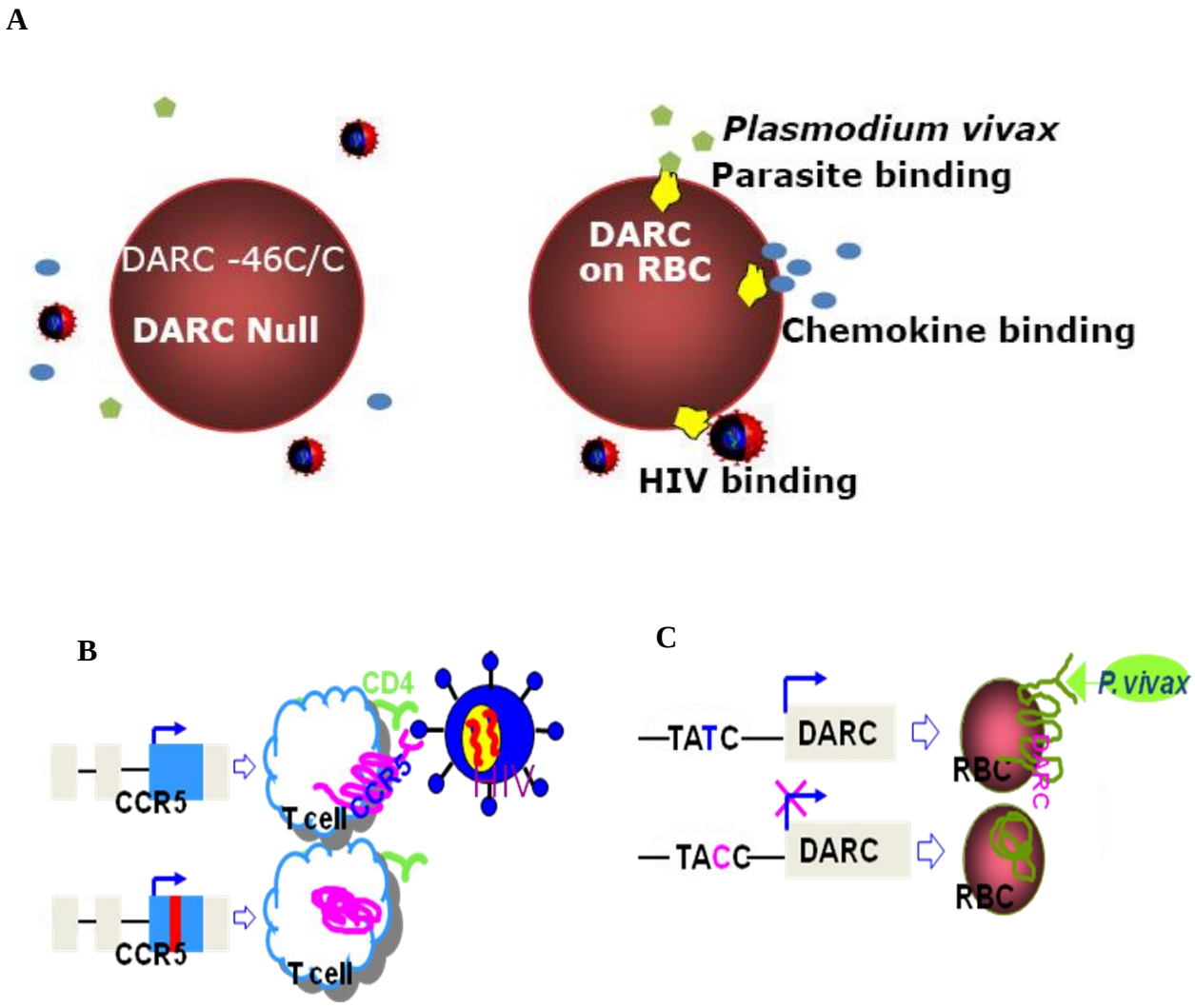


Figure 2.8: (**A**) Schematic diagram on the left indicates the presence of DARC -46C/C, which results in lack of DARC expression on the RBC cell surface. The diagram on the right shows the DARC proteins bound onto a red blood cell. The DARC proteins (shown in yellow) is found on the surface of the RBC, the pathogen *Plasmodium vivax* (shown in green) binds onto the surface DARC to get into the cell. Several chemokines (blue dots) also bind onto DARC. HIV (shown as the red spiked sphere) has also been shown to attach onto the DARC protein (114), this then transfers the HIV bound particle to cells where HIV replication can occur. (**B**) Presence or absence of the CCR5 delta 32 polymorphism directly results whether the protein will be expressed on the cell surface. (**C**) Similar to the CCR5 system the DARC protein is not expressed on the RBC surface based on the presence or absence of the -46C/C mutation.

He *et al*, also reported that the *DARC* negative genotype was associated with delayed development to AIDS related dementia, delayed CD4+ T cell loss, and longer survival. The authors demonstrate that HIV-1 binds to *DARC* on the surface of RBCs and transfers virus to target cells, therefore implicating the RBCs as the carrier of HIV-1 particles to susceptible cells such as CD4+/CCR5+ T lymphocytes (114). Levy et al, in their review illustrate that HIV binds onto RBCs, dendritic cells, lymphocytes, neutrophils, and platelets (172). The HIV bound cell then transfers the virus to target cells; T-cells, macrophages, and mucosal cells (114).

The *DARC* -46C/C genotype is responsible for lower WBC count (140, 214, 250). Reich *et al*, have shown that more specifically lower neutrophil and monocyte counts contribute to the decreased WBC. *DARC* -46C/C associates with low neutrophil and monocyte count (250). It is known that persons of African ancestry have, on average, a low neutrophil cell count, a condition designated as "benign ethnic leukopenia/neutropenia" (86, 108, 132, 158). The time to death in HIV infected persons is significantly shorter in those with a low WBC than those individuals with high WBC (158). Given the association between the *DARC* -46C/C genotype and low neutrophil and monocyte cell count in HIV-positives in studies related to this thesis, the hypothesis was tested that a low neutrophil count is associated with an increased risk of acquiring HIV infection, and further hypothesized that interaction between *DARC* -46C/C and neutrophil counts will influence HIV risk.

2.1.6.a. Controversies associated with *DARC* genotypes and HIV-AIDS susceptibility

After the publication by He *et al*, (114), there has been four studies showing contrasting associations (129, 140, 308, 316), indicating that the *DARC* -46C/C genotype does not influence HIV-1 disease progression/ acquisition. In one of these studies, which was performed by Winkler *et al*, (316), it is stated that the population substructure is considered to be the major confounding factor, particularly due to the factor that the DARC -46C genotype is considered to be a marker for African ancestry. The authors further explain that after performing a principle components analysis (PCA) using the EIGENSOFT program with 70 independent, ancestry-informative markers (AIMS) the results showed that there is minimal population substructure stratification using a cohort of injecting drug users enrolled in the ALIVE cohort in Baltimore, Maryland, between the 454 HIV-infected and 425 HIV-uninfected persons (316).

He *et al*, (113) explain in their response to Winkler *et al*, (316), the 11 markers used in the original study predicted self-reported ethnicity with >98% accuracy, however the authors further explore 96 well-characterized AIMS (154). Using a total of 360 randomly selected European Americans (EAs) and African Americans (AAs) individuals, in addition population reference groups from the HapMap were included, European [CEU, n =60] and African [YRI, n = 60] (113). The DARC -46T/C polymorphism was not selected in the 96 genetic marker set. The results of the principal component (PC) using EIGENSOFT displayed the distinct separation between individuals of European and African ancestry ($p < 0.0001$). The PC analysis between the AAs from HIV-positive and HIV-negative individuals demonstrated no

significant differences in the population stratification p = 0.43 and 0.48 with 96 and 11 markers, respectively. Furthermore, there was a strong correlation noticed between the 96 and 11 markers. The authors indicate that the association noticed between the DARC genotype and HIV risk subsequent to adjusting for the PC of the 11 markers (OR = 1.56; 95% CI = 1.14–2.13, p = 0.005). Therefore the authors (He et al) state that their original findings were not the result of population stratification (113).

He et al (113), highlight the inconsistencies demonstrated by the four studies (129, 140, 308, 316) and therefore state "The epidemiological features (e.g., representativeness of subject selection and controls, sample size, route of infection, endpoints analyzed) of the four studies reported in these correspondences may limit direct comparability" (113). The author's also further comment on the four studies selected cohorts, some of these comments are based on the sample size chosen, as well as the cohort selected i.e. injection drug users. Previous studies indicate host genotypes may differentially influence HIV acquisition based upon the route of infection (192). He et al (113), comment on the choice of samples from the Julg et al study (140) by stating the authors 'investigated 381 subjects comprising an untreated subset of a larger seroprevalent cohort. This may preferentially result in enrolment of slow versus rapid disease progressors, as the latter would more likely be placed on therapy, rendering them ineligible for study participation' (113). Implying that researchers should be mindful of the selection of the cohorts when conducting host genetic association studies.

The authors reiterate their previous findings that the *DARC* -46C/C genotype does associate with variability to HIV/AIDS susceptibility, however further work needs to be done to fully understand the effect that this genotype plays.

Relevance to this thesis

To achieve a better understanding of DARC genotype with HIV/AIDS, which might help clear up the controversies attached with this subject. Here in studies related to my thesis, we plan to explore the intricate interaction between DARC -46C/C and with a low WBC count, found by Julg et al (140) and He et al (113) which influences HIV-AIDS susceptibility (Kulkarni et al., 2009). In this study we utilise the cohort probably best suited for this type of analysis i.e. a highly exposed persistently seronegative cohort of female sex workers from South Africa.

2.1.7. Host antiviral genes

Apart from the genes that restrict viral entry, there are various other genes that are critically involved in HIV-1 restriction and hence could influence the immune response. These genes include, the virus restriction factor TRIM5α and the antiviral APOBEC3G gene, which is described briefly here.

TRIM5α

Tripartite motif-containing 5 alpha (TRIM5α) is a protein of 493 amino acids long and found in most cells of primates. This antiviral factor has also been shown to act as a defence mechanism in both humans and and non-human primates against a number of retroviruses (146). The TRIM5α acts as an intrinsic immune factor, which is critical in the innate immune defence system. The mechanism of the TRIM5α protection occurs as the viral capsid uncoats, TRIM5α, found in the cytoplasm, recognises the motifs within the viral capsid and binds to them. Thus inferring with the viral uncoating process and preventing successful reverse transcription (146).

The TRIM5α protein is not efficient against HIV-1. There have been two polymorphisms described, which affects its antiviral activity. The H43Y and R136Q mutations have shown to play a role in viral activity (302). The H43Y allele is thought to weaken the E3 ligase activity of TRIM and hence leds to progression to AIDS. The R136Q mutation is thought to

be related with anti-HIV activity (146). This mutation has also been reported to be more frequent in high-risk seronegative African Americans as compared with HIV seropositive patients. In the cohort examined in this study, it was noticed that HIV-1 negative participants had a higher expression of the TRIM5α protein than those infected with HIV-1 (270).

APOBEC3G

Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G (APOBEC3G) exists as a tandem array of seven APOBEC genes or pseudogenes. The APOBEC3G is a host antiviral factor that hypermutates the retroviral DNA G-to-A; i.e. it is a cytidine deaminase (324). This therefore leads to instability of the nascent viral transcripts or lethal mutations. However, this APOBEC3G-mediated innate immune mechanism is counteracted by the HIV-1 vif, thus allowing the virus to escape through mutations and even develop drug resistance. A polymorphism, H186R, has shown to be associated with HIV disease progression (6). Individuals that posses the mutated allele progress to AIDS at a faster rate. This result has been duplicated in the cohort examined in this study (248).

2.2. Epigenetics

In 1942, Conrad Waddington referred to epigenetics as ‘the process whereby genotypes give rise to phenotypes during development’ (23, 307). Robin Holliday defined as epigenetics "the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms" (124). Another definition by Arthur Riggs and his group in 1996 referred to epigenetics as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (264). A further definition of epigenetics is given by Adrian Bird as the: “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states” (23). The term ‘epigenetics’ translates to, above/over (*epi*-) the genome (-*genetics*) and refers to the changes in gene/protein expression, from the changes in the DNA sequence (23, 72, 164, 285). There are a number of varying definitions of epigenetics, however despite the number of different definitions given it is still accepted that epigenetics alters gene expression is passed on through cell divisions.

A change in gene expression can result from alternate states in chromatin structure, histone changes, non-coding RNA (e.g. siRNA, miRNA, piwiRNA), polycomb/trithorax protein complexes, ATP-dependent chromatin remodeling complexes and cytosine-5 DNA methylation at CpG dinucleotides (23, 72, 77, 87, 164, 227, 285). These epigenetic mechanisms are directly related to development and differentiation and can arise via selected pressure from the environment or random change (136). Epigenetic mechanisms can also serve as a form of protection from viral genomes which are able to hijack cellular functions for their propagation (136, 137).

The important regulators of gene expression which utilises a series of well-known chemical modifications are the Histones. These modifications include acetylation, phosphorylation, methylation, and ubiquitylation. Alteration of the chromatin structure and the gene expression can be demonstrated by an example of methylating and/or acetylating specific lysine residues of the nucleosomal cores of the histones (136, 137).

The genetic variation and environment affect phenotype is a well known model. A further genetic and epigenetic model of common diseases suggests that the epigenotype influences disease (Figure 2.9). The epigenotype, in turn, is affected by the environment, the epigenotype of the parents, age, and the genotype at loci that regulate DNA methylation and chromatin (101).

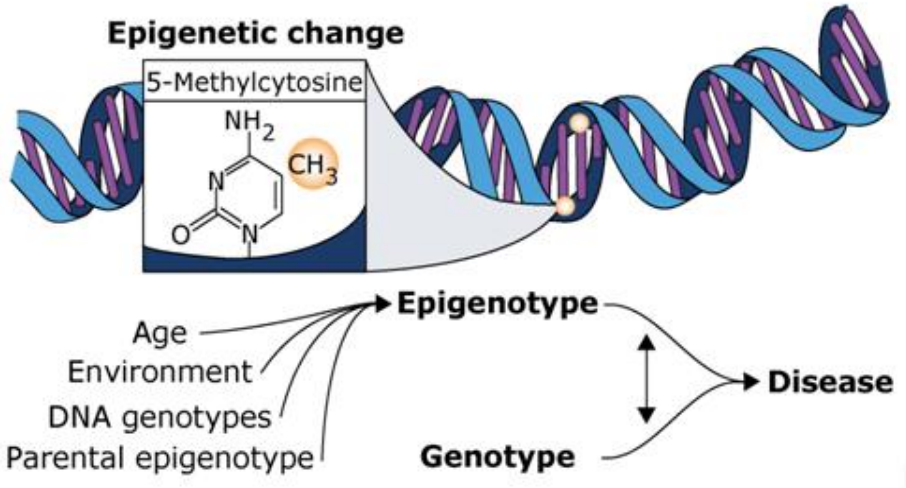


Figure 2.9. Influence of genetic and epigenetic factors on disease (Adapted from Gosden, and Feinberg, 2007).

2.2.1. DNA Methylation

In 1975 DNA methylation was thought to be responsible for a certain gene expression pattern throughout mitotic cell division (125, 136, 257). Currently DNA methylation is considered to be the major contributor to the stability of gene expression states (136).

Chromatin structure is a dominant epigenetic characteristic, and is regulated by a number of factors including cytosine methylation and post-translational histone modifications (PTM). Cytosine methylation is a vital DNA modification that is essential for normal development, chromosome stability, maintaining gene expression states and proper telomere length (23, 72, 164, 285). DNA methylation involves transfer of a methyl-group to the fifth position of the cytosine in a CpG dinucleotide (22) via DNA methyltransferases (DNMT) that create DNMT3A, 3B or maintain DNMT1 methylation patterns (Figure 2.10). Methylation of other dinucleotides such as CpNpG is rare and only has been found in plants or in mouse embryonic stem cells.

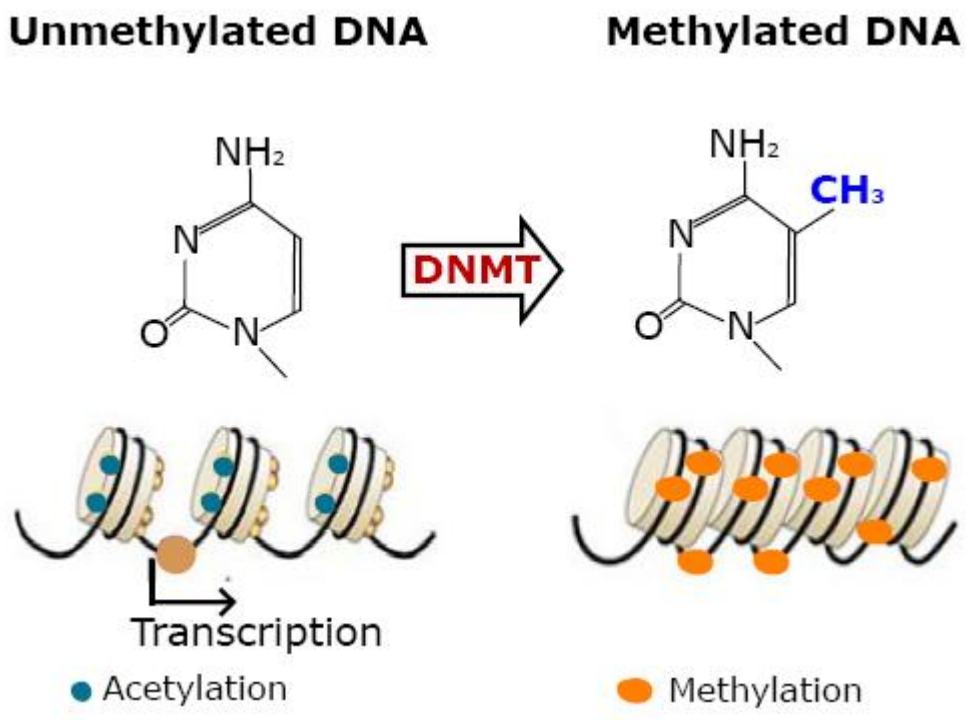


Figure 2.10. Unmethylated DNA undergoes methylation by DNA methyltransferases (DNMT) at the CpG site. DNA demethylation relaxes chromatin structure which allows histone acetylation and the binding of transcriptional complexes. The methylated DNA results in a tight compact chromatin structure which does not allow the transcription machinery access to the DNA hence resulting in lack of transcription.

Three DNA methyltransferases have been discovered in mammals, the first DNMT1 has become known as the maintenance methyltransferase, due to its role in maintaining the DNA methylation patterns through successive rounds of DNA replication (139, 313). The second methyltransferase is DNMT2, (20, 139) and is the most conserved of all DNMT family members. This enzyme was initially thought to lack DNA methyl transfer activity; however, recent data has shown it to be an active enzyme (121). The DNMT3 is the third DNA methyltransferase and has two related enzymes, DNMT3a and 3b. These enzymes are responsible for establishing the somatic pattern of DNA methylation during embryogenesis and also play a role in *de novo* DNA methylation (224).

The DNA methylation levels, measured by PCR-based techniques can be described as a one-dimensional measurement of a multidimensional state. Furthermore, recently new studies show incorporate an integrated analysis of epigenetic states (110, 322) such as combinations of DNA methylation and transcription factor binding (311), gene expression and DNA methylation (207), DNA methylation and histone modifications (220, 311), combinations of gene expression and DNA methylation with the use of epigenetic inhibitors (92, 173, 284) and also triple combinations of gene expression, DNA methylation and histone acetylation (274).

DNA methylation patterns vary between cell types and individuals (84, 266), potentially stemming from environmental exposure (317), stochastic methylation events (89) or heritability(141), and influence development of disease (e.g. cancer, autoimmune diseases), and also time of disease onset (2, 34, 39, 70, 71, 81, 95, 98, 101, 111, 138, 160, 178, 216, 238, 241, 283, 286, 295, 300, 304, 315). Based on several examples of epigenetic inheritance, including from twin-based studies (26, 44, 115, 205, 233, 244, 245), there is little doubt that epigenetic (e.g., DNA methylation) heritability exists. Furthermore, the notion that there are inter-subject differences in epigenetic patterns (81, 251, 285), including among monozygotic twins, is well established.

Normal DNA methylation patterns may vary among individuals (84, 266), potentially stemming from environmental exposure (317), stochastic methylation events (89) or heritability (141), and have been implicated in the development of disease, and also the time of disease onset (81, 238, 304). Alterations in DNA methylation content have been found in diverse clinical entities that range from cancer, imprinting disorders, and autoimmune,

neurological and cardiovascular diseases (34, 70, 71, 111, 138, 241, 283, 304, 315)). Of note, it has been shown that HIV-1 infection may directly modify the host epigenome by increasing the DNMT1 expression levels which in turn resulted in *de novo* increase of global methylation content (74) as well as an increase in CpG methylation of certain genes such as IFNγ (201). Outside the HIV-AIDS field, the importance of epigenetics in disease susceptibility has been highlighted by many studies, spawning the field of "epigenetic epidemiology" (81).

In this emergent field, the attention has been placed on two major disorders: cancer and mental retardation syndromes (reviewed in (81)). Initially it was conjectured that "environmental noise" could account for the phenotypic variation observed in certain traits between monozygotic twins (MZT) (246) as well as for different disease outcomes. Once the bisulphite modification and high throughput methodologies were incorporated in the field to assess methylation content, it was recognized that indeed the environment exerts its disease-modifying effects *via* epigenetic modification.

Just 7% of all CpGs are within CpG islands and most of the CpG sites (75%) are methylated in mammalian genomes (258). Due to the higher deamination rate of the 5-methylcytosine dinucleotide (CpGs are hotspot for mutations), the content of CpG is underrepresented in eukaryotic genomes. However, CpGs that have resisted this methylation-induced deamination, are clustered around 5' ends of first exons and in promoter regions, are usually hypomethylated and because they are found at the expected C+G frequency, they are called "CpG islands". The definition of "CpG island" has suffered in the past decade several changes, but currently, it is accepted as a DNA fragment that has G+C content that is above

50%, an observed/expected ratio of CpG of more than 0.6 and a minimal size of 200bp (287). These elements are usually associated with open chromatin domains (which are DNAse I sensitive) and are linked to gene activation (9). Interestingly, during carcinogenesis or imprinting, CpG islands have been found to be hypermethylated but there are cases where they are hypermethylated even in somatic tissues of normal individuals (273).

As mentioned earlier DNA methylation involves the covalent addition of methyl group to cydines in the CpG context. This "transmethylation" reaction requires the transference of methyl group from the universal methyl donor S-adenosyl methione (SAM), and is catalyzed by DNA methyl transferases. One-carbon biochemical reactions depend on the availability of methyl donor compounds such as methionine and choline and cofactors such as vitamin B12, folate and phosphate of pyridoxal, and hence efficiency of methylation can be influenced by dietary factors or by the environment such as diet, age, other disease, etc. (reviewed in (22) and references therein).

Enzymes that have been linked to the establishment or maintenance of DNA methylation patterns are DNMT1, DNMT2, DNMT3A, DNMT3B as well as DNMT3L. Whereas it has been demonstrated in vitro the role of DNTM1, DNT3A and DNMT3b as "the maintenance" and "the novo" methyl transferases, respectively, the exact biochemical function of DNMT2 and DNMT3L remains obscure. Intriguingly, DNMT2 has been implicated in methylation of RNA [(i.e. cytosine 38 methylation of tRNA(Asp)] since it has been shown that treatment of cell lines with ribonucleoside azacytidine prevents RNA-methylation directed by DNTM2 (267). Experiments in mice have demonstrated that the knockout of DNMT1 or DNMT3A/3B is detrimental for embryonic viability, highlighting the role that DNA

methylation plays in development. Indeed, only conditional knock out (employing Cre-lox methodology) strategies have been used to demonstrate that DNA methylation represses IFNgamma and other gene expression in certain stages of T-cell ontogenesis (10).

DNMT1 propagates DNA methylated patterns mitotically and meiotically (150). It catalizes the addition of methyl group from SAM to the hemimethylated template, which is generated after the S-phase of DNA replication. It "travels" along the replication fork in association with other proteins of the "replicasome" (e.g. PCNA) and hence it passes the epigenetic information from mother to daughter molecule of DNA. Thus, it has been coined the term "maintainance" transferase. DNMT3A and DNMT3B are known as "*de novo*" methyl transferases because they add methyl tags to completely demethylated templates, as shown by *in vitro* assays. They have been involved with centromeric methylation as well as with the establishement of the methylation marks early in embryogenesis. These enzymes seem to have substrate specificity (they are usually found in regions of low complexity such as alpha-satellite DNA) and require the cooperatation with DNMT3L. Neverthess, *in vitro* data suggest the roles of DNMT1, DNMT3A and 3B are probably interchangeable and each of these players might compensate the other one in certain normal as well as pathological situations.

Genomic methylated motifs can be recognized by a plethora of nuclear proteins such as DNA methyl-binding domain proteins (MBDs 1-4) or methyl CpG-binding proteins (MeCP2), which are attracted to methylated DNA. By binding to methylated DNA, they create the scaffold of repressed chromatin along with other correpresors such as HDACs, proteins from the polycom/trithorax group, ATP-remodeling complexes, chromodomain binding proteins

and heterochromatin factors (e.g. HP1α). In addition, some Zn-finger motif-possessing factors such as Kaizo or ZBTB4/ZBTB38 proteins can bind to methylated CpG and prevent RNA transcription (80); for example Kaizo can bind to methylated CGCG motif through its three Kruppel-like C_2H_2 zinc fingers and repress gene expression.

CpG methylation is mainly thought to be a "repressive signature" and the repression is thought to be mediated through recruitment or exclusion of the regulatory proteins. As explained in the paragraph above, methyl-CpG binding proteins interact with histone deacetylases (HDACs) to create a repressive chromatin state (22). Methylation may also preclude binding of activating transcription factors (such as NF-kB p50 subunit, CTCF or c-myb, reviewed in (22)) and thus inhibiting transcription or conversely in rare cases might activate transcription by excluding repressive proteins (22).

How does DNA demethylation occur? Even though the exact biochemical mechanism in mammals is still unclear, two mechanisms have been ruled in. A passive mechanism, described for the regulation of *IL-4* or *IFNγ*, can be speculated to occur by merely preventing the propagation of this epigenetic mark mitotically or meiotically. Simply, a trans-factor could "evict" DNTM1 from its target site and in this way, could prevent the maintenance of the DNA methylation patterns. The second mechanism postulates the existence of an "active DNA demethylase" and it is based on the assumption that a similar mechanism of DNA ß-glycosylation (*via* GADD45a) that occurs in plants could also be operative in mammals (16). Very recently, it has been found that the activation induced cytidine demainase (AID) known as "Aid" (Apobec family) was capable of deaminating rU in RNA as well as 5-methyldC in DNA (206).

2.2.2. Epigenetics and HIV-1/AIDS Pathogenesis.

Even though the international Human Epigenome Project (HEP) has been launched (1, 279) and there is a burgeoning interest in epigenetic epidemiology (81) few studies have examined the direct association between host epigenetics and infectious diseases.

Some pioneer studies have demonstrated that in the battlefield of HIV-1 and host, DNA methylation can be employed to favor viral latency or immune evasion. For example, it has been shown that HIV-1 infection may directly modify the host epigenome by increasing the DNMT1 expression levels (probably by Tat-mediated Erk inactivation) which in turn resulted in *de novo* increase of global methylation content (74) as well as an increase in CpG methylation of certain genes such as interferon gamma (201).

Although still controversial, it appears that HIV life cycle is also regulated by DNA methylation (107, 145, 174). This epigenetic mark elicits two effects on HIV-1 biology: it represses transcription and maintains latency. In fact, many endogenous retroviral elements (e.g. H, K and W families (168, 193)) as well as infective retroviruses (e.g. Rous sarcoma virus (116)) exploit 5' LTR CpG methylation to accrue transcriptional silencing. Very recently, Blazkova *et al. {Blazkova, 2009 #127}* showed that DNA methylation of HIV promoter can lastingly suppress HIV expression. In this study they used a two-fold approach: *in vitro* they analyzed the relation of CpG methylation of the HIV-1 5′ LTR to transcriptional suppression and reactivation of HIV-1 provirus in a model consisting of Jurkat clonal cell

lines harboring a latent HIV-1 LTR–driven retroviral vector; once established this *in vitro* model of epigenetic latency, they compared these findings with the relationships between the methylation pattern, latency, and reactivation of latent HIV-1 provirus in $CD4^+$ T cells of long-term aviremic and viremic patients {Blazkova, 2009 #127}. The authors concluded that "CpG methylation of the HIV-1 5′LTR could be an important epigenetic mechanism that maintains the latency of HIV-1 proviruses by preventing their reactivation" and they further proposed a two-step model of epigenetic control of HIV-1 latency. Interestingly, they showed that treatment with a strong HDAC inhibitor such as suberoylanilide hydroxamic acid (SAHA) could reactivate the latent virus. These results instilled the impetus to investigate whether, in addition to the current HAART protocols, "epigenetic drugs" should be employed to eradicate viral reservoirs (48, 252).

HIV integration sites in activated and resting CD4+ T-cells also show a preferential targeting of the active chromatin compartments (i.e. CpG-island gene-rich domains), as shown by 454 sequencing (27).

Most of the studies conducted so far are either in mouse models or *in vitro* and are circumscribed mostly to genes that are hallmarks of Th_1-Th_2-Th_{17} differentiation/polarization pathways. However, there is a lack of *in vivo* studies on epigenetic regulation of HIV-1 host-restriction factors. To emphasize the paucity of information regarding the role of DNA methylation of genes that encode host factors in HIV-1/AIDS pathogenesis, it can be mentioned that no single publication was retrieved as at the date of the preparation of this Ph.D. dissertation when the key terms: "epigenetic epidemiology" and "HIV/AIDS" were utilized in a Pub Med Search.

Why is the need to explain HIV-1/AIDS pathogenesis in light of epigenetics? After ~25 years of extensive research, explaining and predicting HIV-associated CD4 cell loss, the pathogenesis of HIV disease still is a puzzle. Several lines of evidence from human and non-human primate studies indicate that viral entry-dependent mechanisms cannot account for the full extent of CD4+ T cell loss observed during HIV infection.

First, viral load (VL), a parameter of the extent of viral entry and replication is insufficient in explaining fully time to AIDS or rates of CD4+ T cell declines. For example, Dolan, Ahuja *et al.* found that ~9% and ~4% of variability in time to AIDS and CD4+ slope was explained by steady-state VL, respectively(66). Likewise, ~6% and ~3% of variability in time to AIDS and CD4+ slope, respectively, was explained by polymorphisms in *CCR5* and *CCL3L*. Second, during chronic untreated HIV infection, relatively few HIV infected CD4+ cells can be detected, even in subjects with a high VL (118). Third, SIV infection of several non-human primate species does not cause AIDS despite high levels of viral replication (276). Together, these three observations suggest that host factors might have a comparable or even greater impact on disease progression than does the extent of viral replication, such that indirect, viral-entry independent mechanisms (those not attributable to HIV replication) might contribute significantly to HIV-AIDS pathogenesis. CMI conferred by the *CCL3L-CCR5* axis was identified by our group as one of the VL-independent parameter of HIV/AIDS pathogenesis; however, mathematical analysis suggested the presence of other VL-independent factors (66). Concordantly, recent genome wide association studies found genetic variants in the MHC locus that explain only ~10% of the variability of VL setpoint

(75). In summary, these findings suggest that factors other than host genetic variation impact on VL setpoint; we believe that epigenetics could fill in the gap.

Relevance to this thesis

Epigenetics has the ability to alter gene expression, as demonstrated in other disease models. A change in the expression of the CCR5 gene is critical for HIV susceptibility/protection as demonstrated by many previous studies. Measuring DNA methylation could help explain the changes in the CCR5 gene expression, since methylated DNA does not allow the transcription machinery to bind therefore resulting in lack of gene expression. Furthermore, in this cohort a unique opportunity is provided by measuring the DNA methylation levels in HIV negative individuals and following them over time to the point of HIV seroconversion, and thereafter followed up regularly during the course of HIV disease progression. To our knowledge there have been no studies exploring the relationship between epigenetics and HIV/AIDS.

2.2.3. CCR5 Expression

Expression levels of CCR5, the major coreceptor for cell entry of HIV-1, on T cells are a central determinant of nearly all facets of HIV/AIDS pathogenesis including HIV cell entry (148, 175, 176, 225, 230, 318) and replication (90, 256), HIV (18, 230, 253, 288) –AIDS (90,

255) susceptibility, and CD4 cell recovery during HAART (90, 117, 144, 254, 299, 305, 319). In each case, low CCR5 surface expression levels are associated with a protective phenotype.

Defining factors that fine-tune CCR5 levels is of critical importance because there is a 'threshold' of CCR5 surface expression that is permissive for cell entry, and small changes in CCR5 density are associated with large increases in HIV infectivity and efficacy of CCR5 blockers (119, 120, 155, 171, 237, 254, 255). Kuhman *et al*, and Platt *et al*, have shown that small changes in CCR5 density are associated with an exponential increase in HIV infectivity *in vitro* (155, 237). Accordingly, small changes in CCR5 expression are exponentially correlated with VL and disease progression (119, 171, 254, 255). Additionally, *in vitro* data shows that small changes in CCR5 levels have large effects on efficacy of CCR5 blockers/entry inhibitors (119, 120).

Although there is remarkable variation in CCR5 levels between persons, longitudinal examination of CCR5 expression demonstrates that within a healthy person it is remarkably stable over time (36, 37, 58, 83, 149, 181, 195). Although the HIV field is heavily focused on CCR5 density (number of CCR5 molecules) per cell, the proportion of CCR5 expressing cells also has critical importance for two reasons: First, measuring numbers of CCR5-expressing T cells vs. simply CCR5 density also has the advantage of directly relating CCR5 expression to variation in PBMC levels among subjects or within each subject during disease progression, factors which are frequently not considered. Second, although there is a strong correlation between CCR5 density and %CCR5-expressing cells, it is not 100%. This suggests that

although density and %CCR5-expressing cells are inter-related, both common and specific molecular mechanisms may underlie their genesis.

The mechanism after infection is very specific as demonstrated by O'Brien et al (223)"when HIV-1 infects a person, it seeks out tissue cell compartments where it can replicate, including lymph nodes, neural tissue, epithelium in gut or vagina, spleen, testes, kidneys and other organs. HIV's principal factories are macrophages, monocytes and T-lymphocytes, all of which carry the CD4 cell surface protein. HIV-1 enters various cells by co-opting two receptor proteins on the cell surface" (223). HIV also selects specific cells for infection, as explained above; CCR5 is required for viral entry and the more CCR5 available to bind to subsequently results in greater chance for HIV acquisition (119, 120, 155, 171, 237, 254, 255).

CCR5 also plays a key role in several facets of T-cell immunity (204), including T-cell differentiation (185), proliferation (35, 143), and activation-induced cell death (209, 236). Consistent with this, prior studies in CCR5 knockout mice and HIV-negative subjects revealed that CCR5 expression influences T cell immune responses *in vivo* (65). Additionally, levels of CCR5 and T cell activation are highly correlated (18, 219, 228, 278, 325), and CCR5 expression is closely tied to the type of effector function acquired during T cell activation, that is, a Th1 response (180, 185). Although IL-2 impacts on CCR5 levels (312, 321, 327), we found that CCR5 expression influences NFAT translocation, IL-2 production, and subsequent signaling events during T lymphocyte activation (e.g. T cell differentiation) (35). These data demonstrated a link between CCR5 and pathways

downstream of T cell receptor (TCR) that influence IL-2 expression, and that the expression levels of CCR5 and IL-2 are coregulated.

CCR5 ligands such as *CCL3L* can influence CCR5 surface expression by inducing ligand-mediated internalization and subsequent recycling/degradation of the chemokine receptor (Figure 2.11). CD26/dipeptidyl peptidase IV alter the biochemical structure of chemokines, affecting their coreceptor specificity/agonistic activity, disrupting the network of chemokine receptor/ligands and thus, indirectly, influencing CCR5 cell expression. Atypical receptors such as D6 or DARC, by binding to "free" circulating chemokines can buffer the availability of ligands that can internalize or desensitize CCR5. In terms of CCR5-dependent factors, *CCR5* locus (defined as *CCR5* haplotypes) result in differential DNA-binding protein profiles that can impact on CCR5 transcriptional ability. The goal of thisThesis is to study how various factors can regulate CCR5 expression and hence link this with HIV-AIDS susceptibility or resistance.

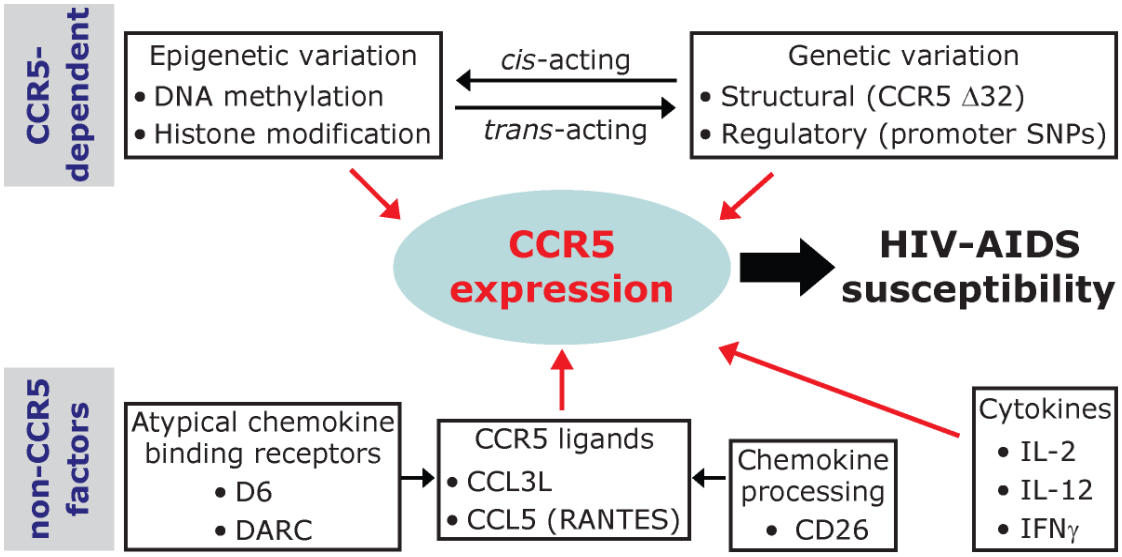


Figure 2.11: Non-CCR5 factors and CCR5-dependant factors directly and indirectly affecting CCR5 expression, which leads to HIV-AIDS susceptibility (unpublished).

2.2.4. CCR5 Promoter regions

CCR5 contains 3 exons: 1 coding exon, 2 non-coding exons designated as exon 1 and exon 2A/2B. CCR5 has 2 promoters (Pr): Promoter 2 (Pr2) and Pr1 (figure 13) drive production of alternatively spliced mRNAs with distinct 5' untranslated exons, named as "full length" or "truncated" isoforms, respectively. All transcripts have an identical open reading frame (ORF) that resides in exon 3 but on the basis of whether these transcripts contained or lacked exon 1 they were designated as full length or truncated transcripts, respectively (figure 2.12). Promoter 1-driven, exon 1-*lacking* transcripts are present even in cell types that do not have surface CCR5 (e.g. naïve CD4+ T cells), therefore we conceptualized Pr1 as the "constitutive" *CCR5* Promoter. This establishes the following relationship: "Pr1 — exon 1-lacking, truncated isoforms — absent or low CCR5 surface expression". On the other hand, production of promoter 2-driven, exon 1-containing full length transcripts correlates with CCR5 surface expression and T cell activation (therefore, Pr2 represents the *CCR5* "inducible promoter*"*), establishing the following relationship: "Pr2 — exon 1-containing, full-length isoforms — presence of CCR5 surface expression."

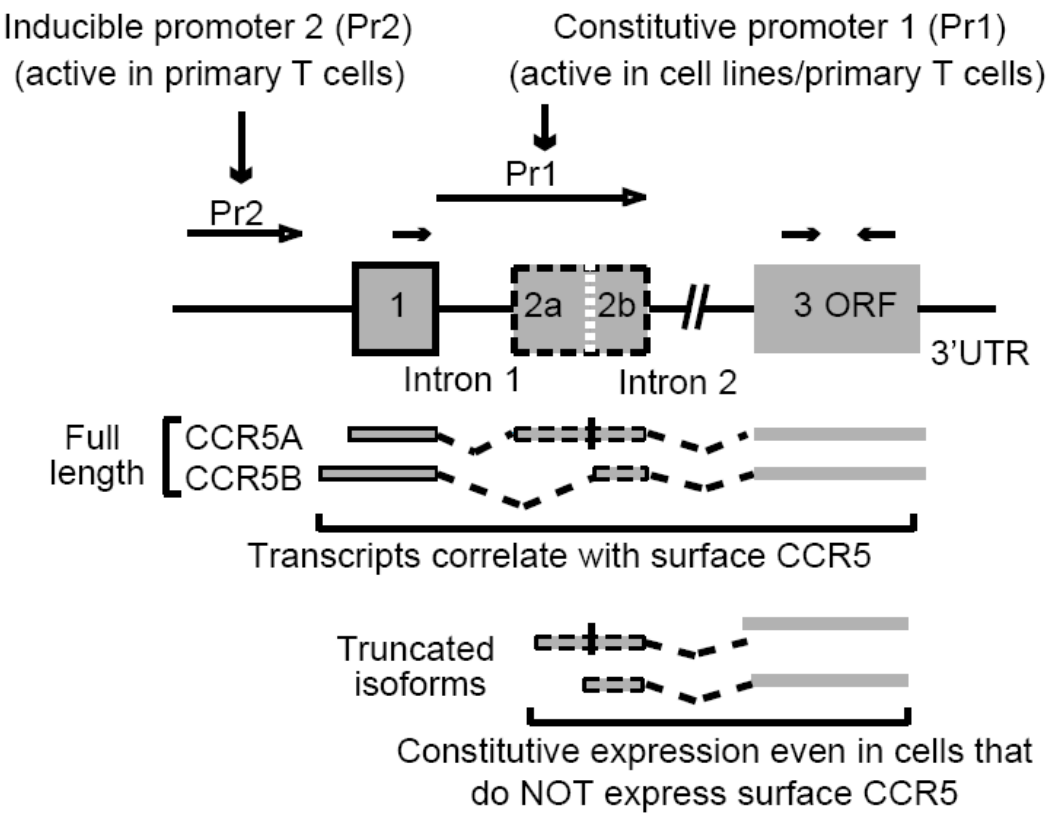


Figure 2.12: CCR5 gene/mRNA structure and its 2 promoters (unpublished).

Initial screening of CpG sites in the CCR5 promoter and intron gene regions was performed on bisulphite modified DNA (EZ kit, Zymo Research) (unpublished). A total of 31 CpG sites were located upstream of the CCR5 ORF (Figure 2.13). These CpG sites were PCR amplified, cloned, multiple plasmids were picked at random and sequenced. A cost effective method was established whereby, certain positions which showed the most significant changes in methylation, assessing the CD4+CD45RO+/- activity, were selected. For the CCR5 Intron 2 region, the CpG positions 1 to 6 were selected. While for the CCR5 promoter 2 region the CpG positions -28 to -31 were selected (Figure 2.13) (99).

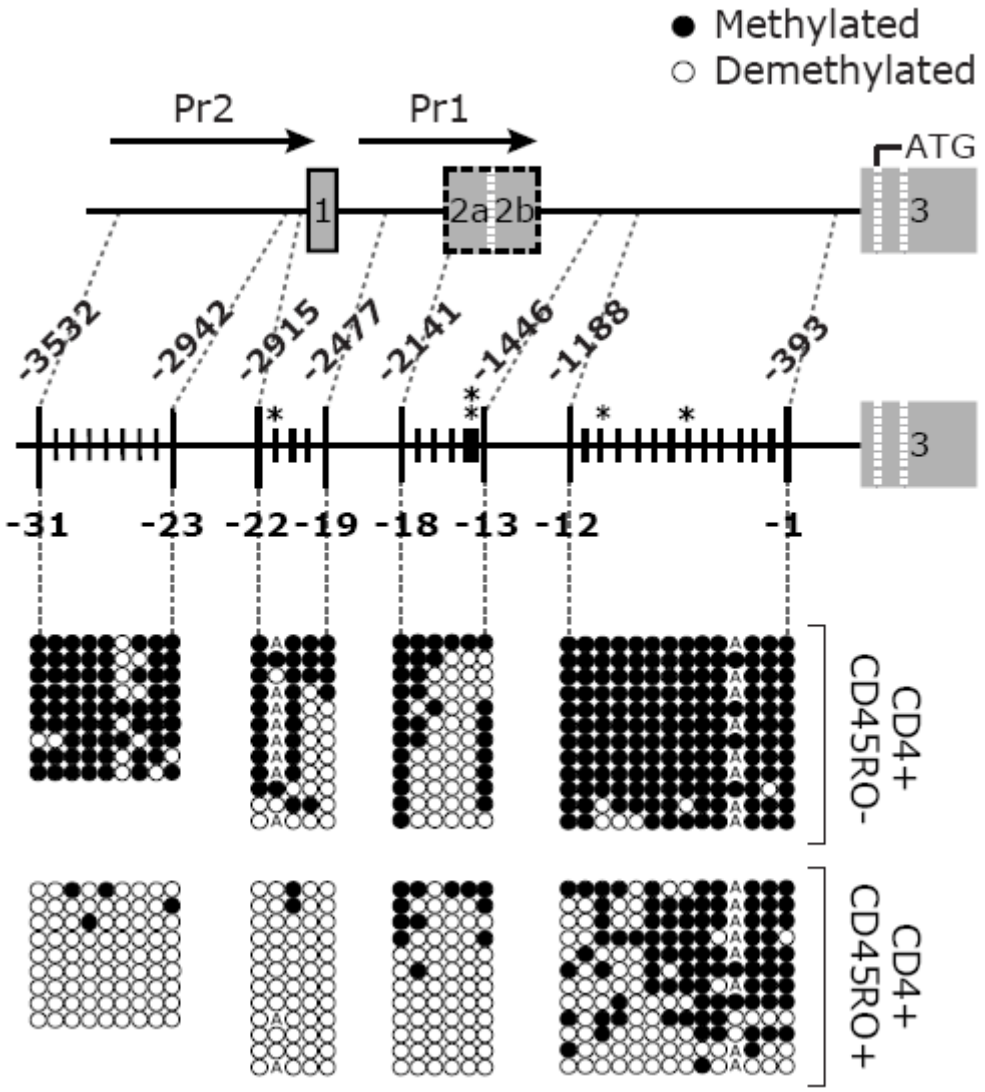


Figure 2.13: Methylation (CpG) sites in the CCR5 Promoter and intron regions (unpublished). Black circles represent methylation and white circles represent demethylation. Each circle represents a single PCR clone picked at random.

The CpG site that resides closest to the *IL-2* transcription start site (TSS) has been previously shown to closely track IL-2 expression (31). DNA methylation profile of IL2 shows that this gene is composed of 4 exons (black boxes in Figure 2.14). Methylation statuses of six CpGs upstream of the IL2 transcriptional start site were determined. These CpG sites were selected because they play a role in IL-2 gene regulation.

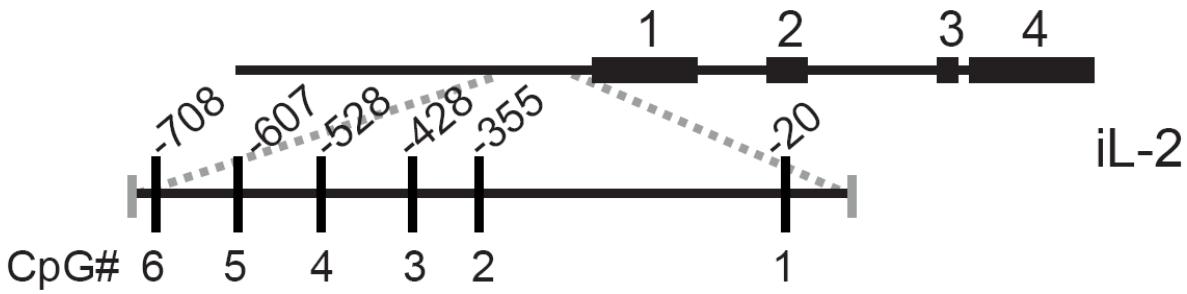


Figure 2.14: IL-2 region demonstrated with six CpG sites located in the promoter region (unpublished).

Gonalusse *et al* , {Gornalusse, 2011 #411}also suggests that CCR5 Promoter 2 and other regulatory regions such as intron 2 (a genomic region that we found also influenced transcriptional activity) is methylated in naïve T cells, which therefore provides the basis for repression of exon 1-containing transcripts and minimal CCR5 surface expression on naïve T cells. The data also shows that CpGs in Pr1 is demethylated even in naive T cells, providing a basis for the constitutive expression of exon 1-lacking CCR5 mRNA isoforms in naïve T cells without concomitant high CCR5 surface expression in these cells.

In this study we examine the role of CCR5 DNA methylation, located in the promoter and intron regions, and CCR5 expression looking at both the density per cell and the number of cells expressing CCR5. Furthermore, we look at the effect of DNA methylation on HIV acquisition and disease progression. We also highlight confounding effects of non-similar sources of DNA when measuring DNA methylation.

CHAPTER 3 - MATERIALS AND METHODS

3.1. CAPRISA cohort

The HIV infected and exposed uninfected samples used in this study were obtained from Centre for the Aids Programme of Research (CAPRISA), Doris Duke Medical Research Institute (DDMRI), University of KwaZulu-Natal (UKZN), Nelson R. Mandela Medical School and the CAPRISA Vulindlela Research Facility. Ethics for this study was approved by the UKZN ethical review committee. The ethics number for the study is E013/04

The CAPRISA 002 Acute Infection study is a longitudinal cohort study and consists primarily of high risk HIV-uninfected women working as female sex workers (FSW), recruited from a large urban area (301). This represents the first acute infection study in Southern Africa to document acute infection in a prospective cohort with extensive follow-up on the natural history of HIV-1 subtype C infection. A paper published by van Loggerenberg et al describes the details of establishing the CAPRISA 002 study (301). This thesis contains a brief description of the study with some additional details.

The target group for the study is women who are at greater risk of HIV acquisition, and data has shown that younger women in urban South Africa have an increased risk for HIV infection (247). To be enrolled in the study individuals must be self-identified FSW or women reported as having at least three sexual partners in the last three months prior to recruitment or at least 2 sexually transmitted diseases (STDs) episodes in the last 3 years were screened for study participation. Eligible women had to be at least 18 years of age and not pregnant at the time of screening and assurance that adequate birth control measures will

be used for the duration of the study. Individuals that planned on travelling away from the study site for more than three months were not included in the study. Individuals that are negative for the ELISA and Western Blot HIV-1 tests.Also are able to provide written informed consent to have samples stored. HIV negative individuals were followed up and received counseling; however when an individual did seroconvert they were recruited and followed up.

3.1.1. Additional HIV Infected recruitment

In order to increase the number of HIV- infected individuals study participants, with acute HIV infection , meeting one of the following criteriawere recruited:HIV antibody seropositive with a documented HIV seronegative test in the previous three months; HIV antibody negative with HIV infection demonstrated by HIV-1 RNA testing; Willingness to adhere to the evaluation schedule; Hemoglobin > 10.0 g/dl at screening; Age: 18-60.

Participants who met any of the following criteria were NOT eligible for this study: Subjects who planned to travel away from the recruitment area for > 3 months during the 24 months following screening; History of immunodeficiency, chronic illness, malignancy, autoimmune disease, or use of immunosuppressive medications; Individuals with a history of cancer were excluded unless there has been surgical excision followed by sufficient observation period to give a reasonable assurance of cure: Medical or psychiatric condition or occupational responsibilities, which precluded subject compliance with the protocol; Prior receipt of HIV-1 vaccines; Pregnant or lactating women; Subjects who had any other condition that, in the

opinion of the Principal Investigator or designee would preclude provision of informed consent, or make participation in the study unsafe. Antiretroviral therapy (ART) was disallowed in the study . The initiation of ART was an endpoint for this study and the individual was not followed up. Concomitant ART was not permitted while on study.

3.1.2. Date of infection calculation

Date of HIV infection was estimated using the following algorithm: If no previous HIV serology result was available and a positive RNA was available on the same date as a negative HIV enzyme immunoassay (EIA), the HIV infection was estimated at 14 days prior to the positive RNA EIA test date. The second method used for date of HIV infection calculation was estimated as the midpoint between the last documented HIV negative EIA and any of the first positive EIA.

3.1.3. Sites of the study populations

This prospective observational cohort study was conducted at the following sites:

(i) Doris Duke Medical Research Institute (DDMRI) at the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban.

(ii) CAPRISA Vulindlela Research Facility. The CAPRISA rural facility in Vulindlela adjoins the Mafakathini Primary Care Clinic. In this setting, adolescents and particularly young women are a vulnerable group at high risk of acquiring HIV. Women acquire HIV infection at a younger age, at least 5-10 years earlier than men. Adolescents utilizing the primary health care services for antenatal, family planning or sexually transmitted infections (STI) services and who met all study eligibility criteria were invited to participate in this study.

(iii) CAPRISA eThekwini Clinical Research Site is based the Prince Cyril Zulu Communicable Disease Clinic, Durban. The Prince Cyril Zulu Communicable Disease Clinic (CDC) is a large local government clinic for the diagnosis and treatment of STIs and TB, for which it provides free treatment. Given the high HIV prevalence of 63% in this group of STI patients and the strong association between STIs and HIV, these patients are a key risk group for acquiring and transmitting HIV. HIV negative male and female patients seeking STI care at the Prince Cyril Zulu clinic and who met study eligibility criteria were invited to participate in this study.

3.1.4. Phases of the study

The individuals were grouped in different phases based on their infection and progression status. Phase I refers to HIV negative individuals, Phase II is HIV positive within three months of infection, Phase III is HIV Positive between three to twelve months, and Phase IV is HIV positive greater than 12 months. Table 3.1 summarizes the phases of the study and describes the criteria for enrolled, the period enrolled and frequency of the visits.

Table 3.1: Summary of the Phases and follow-up of the study, indicating both serostatus at enrollment into each phase and their follow-up schedule, as well as time from enrollment (into Phase II).

Study Phase	HIV Status	Months from Infection	Visit Frequency
I: HIV negative	HIV Antibody negative and HIV RNA negative	N/A	**Monthly**
II: Acute Infection	HIV antibody negative and HIV RNA positive; OR HIV antibody positive with a documented HIV negative within the previous 3 months	0 - 3	**Weekly for 3 weeks, fortnightly until 3 months post enrollment**
III: Early Infection	HIV seropositive	3 - 12	**Monthly, post diagnosis**
IV: Established Infection	**HIV seropositive**	**> 12**	**Quarterly, post enrollment**

3.1.5. Study Design

A total of 775 women were screened over a 6-month period, of these 462 had anHIV infection and hence ineligible for the study. A further 68 women were also not eligible due to various reasons e.g. some were pregnant, relocating to a different area, work commitments etc. Therefore, 245 HIV high risk women were recruited into the study and were enrolled in phase I (refer to figure 3.1). These women were followed up for two years, during which 28 seroconverted to HIV positive status and were hence enrolled into phase II. To increase the number of HIV infected samples we recruited a further 44 seroincident cases from other similar CAPRISA studies mentioned above.

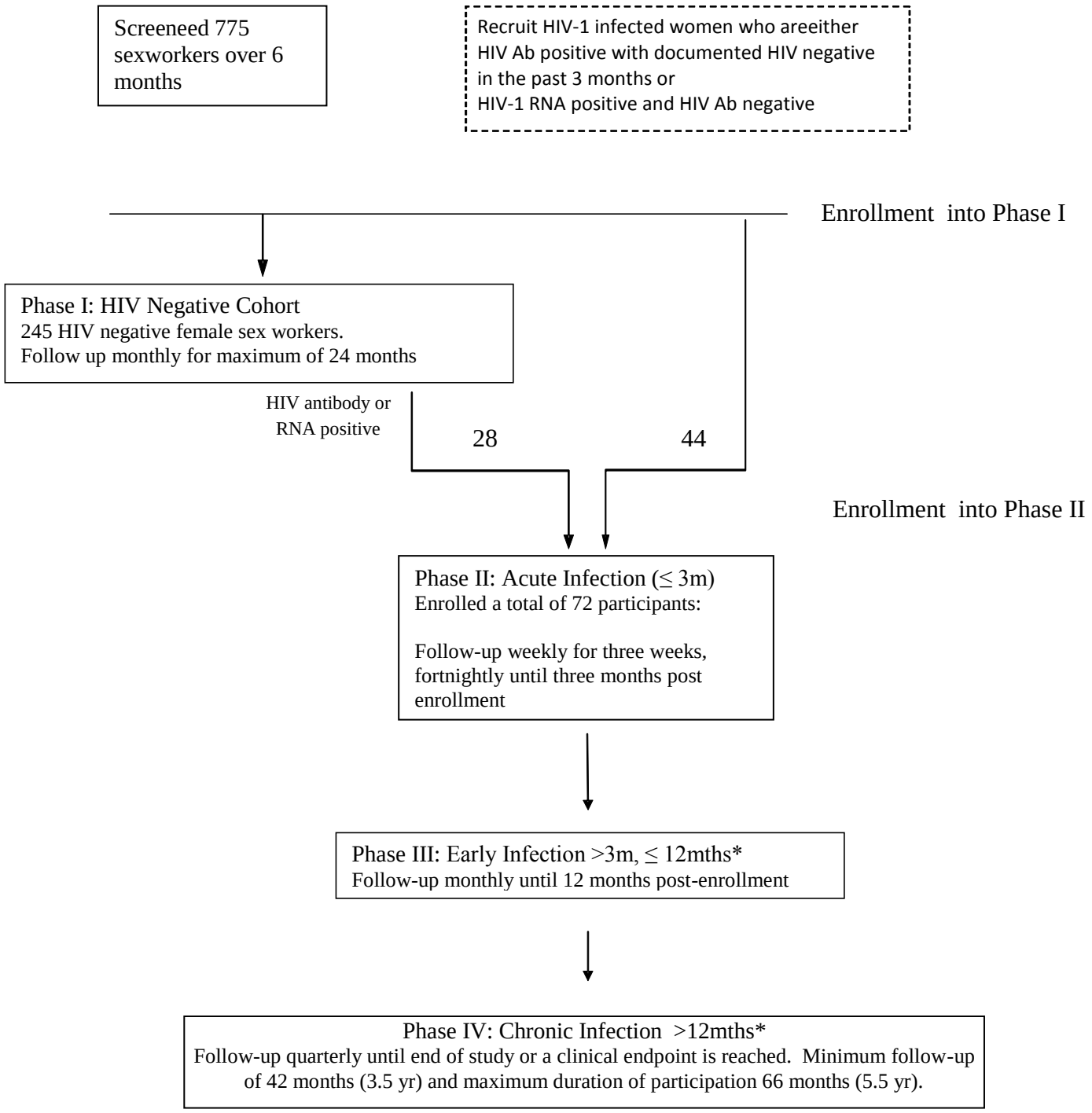


Figure 3.1: Flow diagram of the screening and recruitment for the CAPRISA study with the breakdown of each phase.

3.1.6. Laboratory Assays and Procedures

Sexually transmitted infections were used as a biological marker for behavioral risk. Blood specimens were collected and tested for syphilis, Hepatitis B virus (HBV) and HSV-2 serology. Vulvovaginal swabs are used to test bacterial vaginosis, *N. gonorrhoeae*, *C. trachomatis*, *T. vaginalis*, *M. genitalium* and HSV-2. Cervicovaginal lavages were also collected to test for human papilloma virus (HPV) and DNA methylation levels measured. Individuals possessing genital ulcers had specimens taken for DNA extraction and PCR testing for *T. pallidum*, *H. ducreyi*, *C. trachomatis* (LGV types), *C. granulomatis* and HSV-2. Blood specimens were collected for; hematology assays (full blood count), biochemistry assays (urea and electrolytes, liver function tests, calcium, magnesium, phosphate, Vitamin B12, iron, folate, blood glucose, cholesterol, LDL, triglycerides), on-site rapid testing for HIV antibodies. Hemoglobin at enrollment into phase II was determined using a hemoglobinometer. Urine specimen collected for on-site β-HCG pregnancy testing as well as Uricheck Dipstick.

CD4 + T cell counts were measured using the FACSCalibur flow cytometer (Beckton Dickinson). The procedure is as follows; whole blood is stained using the Beckton Dickinson Tritest (Becton Dickinson 342414). The reagent contains a monoclonal antibody cocktail composed of CD4 FITC, CD8 PE and CD3 PerCP. When a known volume of whole blood is added to the staining reagent, the monoclonal antibodies bind to specific cell markers. When these pass through the flow cell of the FACSCalibur flow cytometer, they are excited by the laser beam and fluoresce. Staining takes place directly in a TruCOUNT tube containing a lyophilized pellet. The pellet dissolves liberating a known concentration of fluorescent beads.

The absolute number of positive cells is determined automatically by the software (Multiset) by comparing the number of cellular events with the number of bead events.

The COBAS AMPLICOR™ HIV-1 MONITOR Test, v1.5 (Standard or the Ultrasensitive, Roche Diagnostics) was used to measure viral loads. The detection of HIV-1 RNA in pooled samples the AMPLISCREEN^{Tm} HIV-1 Tests v1.5 (Roche Diagnostics) were used.

3.2. Durban HIV negative cohort

Control samples with lower HIV exposure and HIV negative that are also age matched, were recruited from the general population, as a comparative group to the CAPRISA cohort. The ethics for this study was approved by University of KwaZulu-Natal ethics number is BE081/09. Inclusion criteria were as follows; firstly only black females from the similar area as the CAPRISA cohort were considered, secondly, women between the18-60 age group; and thirdly, HIV negative according to an ELISA test.

Health care workers, predominantly nurses, were screened for HIV-1 infection at Inkosi Albert Luthuli and King Edward VIII hospitals. Eighty four individuals were HIV-1 screened, of these 24 were HIV infected and not considered for study inclusion. The study participants enrolled totalled 58. PBMC's were collected, processed and stored at the HIV Pathogenesis Programme (HPP), Nelson R Mandela Medical School, University of KwaZulu-Natal.

3.3. DNA Methylation

3.3.1. Sodium Bisulphite Mutagenesis.

Genomic DNA was bisulphite-modified and PCR amplified and individual clones were sequenced according to the original procedure described by Grunau *et al.* (104). Concentration of $NaHSO_3$ was increased to 5 M and time reduced to 5 hrs to increase kinetics of reaction and decrease template degradation. Alternatively, bisulphite modification was performed using the commercially available kit (EZ DNA methylation Direct, Zymo Research), following the manufacturer's instructions. The bisulphite method of mapping 5-methylcytosine exploits the fact that sodium bisulphite ($NaHSO_3$) reacts readily with the 5, 6-double bond of cytosine, but poorly with methylated cytosine. Cytosine reacts with the bisulphite ion to form a sulfonated cytosine reaction intermediate which is susceptible to deamination, giving rise to a sulfonated uracil. The sulfonate group can be removed under alkaline conditions, resulting in the formation of uracil. Uracil is recognized as a thymine by *Taq* polymerase and hence upon PCR the resultant product will contain cytosine only at the position where 5-methylcytosine starts in the DNA template. When sequenced, originally unmethylated cytosine appears as thymine, while methylated cytosine retains its identity. Methylation status of CpG sites in *CCR5* and *IL-2* was assessed by sequencing and pyrosequencing as described below.

3.3.2. PCR Amplification of *CCR5* promoter 2, intron 2 and IL-2

A 173bp fragment encompassing four CpG sites, located in CCR5 Pr2 at -776, -768, -702 and -681 positions was amplified for the CCR5 Pr2 methylation analysis. For quantification of CCR5 intron2 methylation, a 296-bp amplicon containing 6 CpGs in intron2: 2169, 2187, 2258, 2311, 2342 and 2365 was designed. Primer sequences are listed in table 3.2. The PCR reaction mix consisted of PCR buffer, 3 mM $MgCl_2$, 200µM of each dNTP, 0.2µM of forward primer, 0.2 µM 5' biotinylated reverse primer, 2.5U *Taq* polymerase and 2-4 µL of bisulphite treated DNA. PCR was performed under "hot-start" conditions to reduce non-specific amplification. For amplification of Pr2, PCR cycling conditions were: initial stage of 80°C for 4 minutes and 94°C for 2 minutes, followed by 45 cycles of (94°C for 10 sec, 55°C for 30 sec and 72°C for 30sec) and a last stage of 72°C for 10 min. For amplification of int2 fragment, similar conditions were followed with the exception of the annealing temperature, which was 60°C. Successful PCR amplification was confirmed by gel electrophoresis. Five microliters of PCR reaction mixtures were mixed with 1µl of xylene cyanol sample loading buffer and loaded onto a 3% w/v agarose gel.

Table 3.2: Primers used for Pyrosequencing Analysis.

Primers	Oligonucleotide sequences
Pr2 forward	5'-TTTGAATTGTATATATGGGATGAA-3'
Pr2 reverse	5'-Biotin-TACTTAAAAAAAACCAAAACAATATAA-3'
Pr2 seq1	5'-ATATATGGGATGAATTAGAA-3'
Pr2 seq2	5'-TGAGTTTTGTTGTAGTATAG-3'
Int2 forward	5'-GGTTAGTATTTTAGGAGGTTGAGGTAG-3'
Int2 reverse	5'-Biotin-CAAACATAATACAACTCAACCTTTTTA-3'
Int2 seq1	5'-TTAGGAGGTTCAGGTAGG-3'
Int2 seq2	5'-TAGTTTGGTGTGGTGG-3'
Int2 seq3	5'-TAGGTTGTAGTGAGTTATGA-3'

3.3.3. Pyrosequencing

Pyrosequencing was employed as a strategy to quantify methylation levels in the genomic DNA and was performed by a commercial firm (Epigendx, MA). PCR products were pyrosequenced with the PSQ 96HS system (Biotage AB, Uppsala, Sweden), according the established method (69), using sequencing primers shown in table 3.2. Pyrosequencing is a sequencing method based on real-time monitoring of DNA synthesis (73, 290, 291) and contains four coupled biochemical reactions that result in a bioluminescence signal. The four enzymes used are *Klenow* fragment of DNA polymerase, ATP sulfurylase, luciferase and apyrase. Other reagents include enzyme substrates such as adenosine phosphosulfate and D-

luciferin and sequencing template with an annealed primer, which acts as starting material for DNA polymerase. We validated the sensitivity of the pyrosequencing technique by mixing different proportions of universal methylated and universal unmethylated templates and performed pyrosequencing in order to obtain highly corroborative results.

3.4. CCR5 Protein Expression

The surface expression of CCR5 was measured as described by Hladik F, et al 2005 (123), with a few modifications. The density of CCR5 was determined using the fluorescence-activated cell sorting FACSCalibur and Dako Qifi Kit.

In preliminary experiments, the saturating concentrations of the anti-CCR5 monoclonal antibodies were determined (2.5 g/ml) and used throughout the study. PBMC were incubated with either anti-CCR5 (clone 2D7), or immunoglobulin G2a (IgG2a) isotype control (all BD Pharmingen) in phosphate-buffered saline supplemented with 1% bovine serum albumin and 0.1% sodium azide (FACS buffer) for 30 min at 4°C. After two washes in FACS buffer, cells were incubated in fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG Fc F(ab)2 (1:100; Qifikit; DAKO) for 20 min at 4°C. Cells were washed twice, and a final incubation step was performed with a combination of antibodies shown in table 3.3. Phycoerythrin (PE)-conjugated anti-CD3 or HLA-DR or control IgG2a. Allophycocyanin (APC)-conjugated anti-CD4 or anti-CD14 (all BD Pharmingen). Red Live/dead (LD) marker was included (Invitrogen) (1 g/ml) during this step to mark dead cells. Two final washes were performed in FACS buffer. Cells were fixed in 2% paraformaldehyde containing 5 g/ml actinomycin D and analyzed on a Calibur flow cytometer (Becton Dickinson) within 24 hours after staining (Figure 3.2). Using the LD marker, dead cells were excluded from the analysis, and the percentages of CCR5 expressing CD3, HLA-DR, CD4-T-helper lymphocytes and CD14 monocytes were determined (123).

Table 3.3: The staining panel used for each sample to calculate the CCR5 density on a cell.

Tube 1	**Tube 2**
CCR5 – 0.25 µl	IgG2a – 0.25 µl
GAM-FITC - 0.5 µl	GAM-FITC - 0.5 µl
HLA-DR – 5 µl	IgG2a-PE – 5 µl
LD – 1 µl (1:1000)	LD – 1 µl (1:1000)
CD4 -APC– 2 µl	CD4 -APC– 2 µl
Tube 3	**Tube 4**
CCR5 – 0.25 µl	IgG2a – 0.25 µl
GAM-FITC - 0.5 µl	GAM-FITC - 0.5 µl
CD3-PE- 1 µl	CD3- PE- 1 µl
LD – 1 µl (1:1000)	LD – 1 µl (1:1000)
CD14 -APC– 0.5 µl	CD14 -APC– 0.5 µl

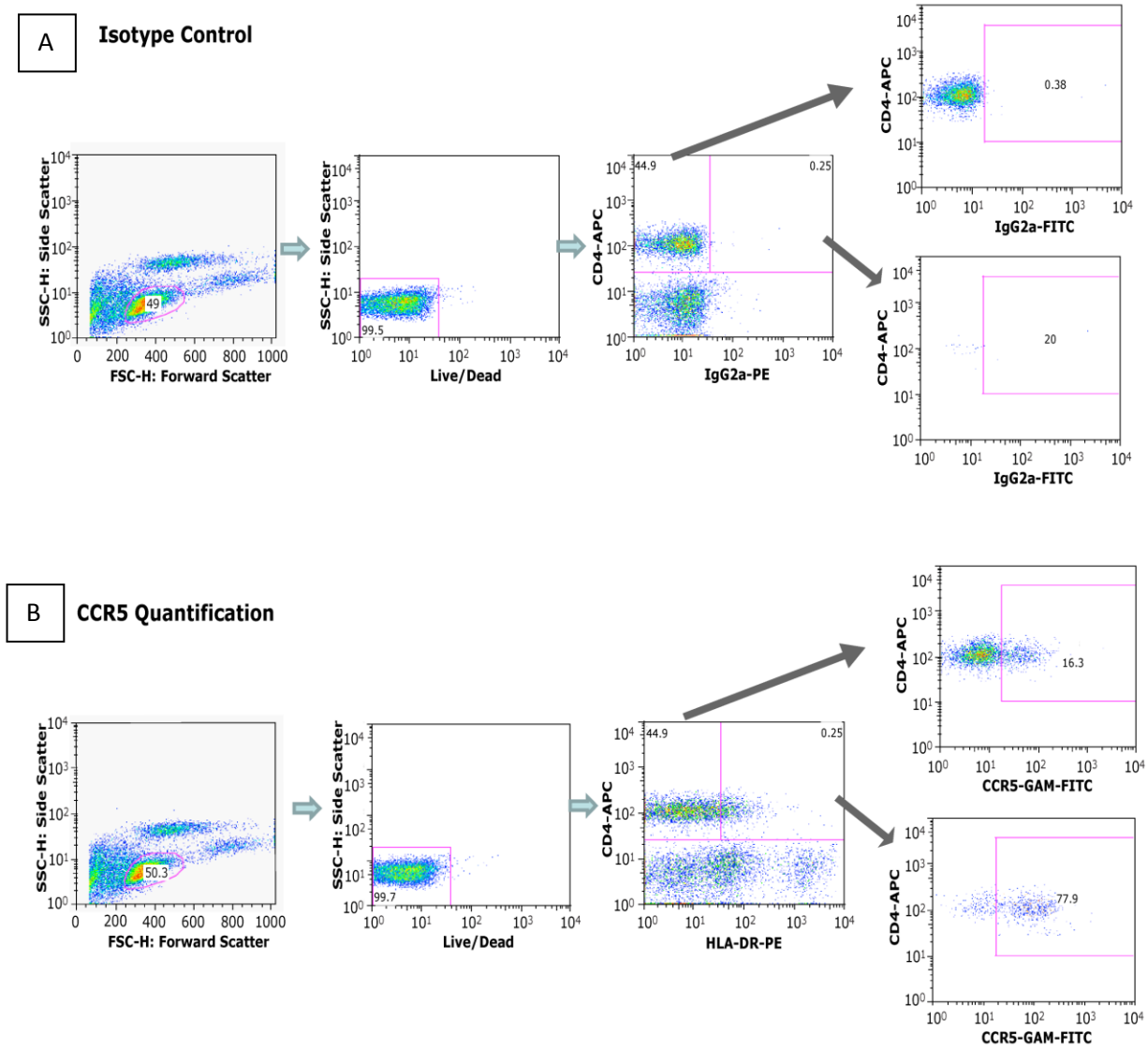


Figure 3.2: Gating strategy used for CCR5 quantification. (A) The isotope control used for drawing gates. IgG2a-PE is the control for HLA-DR-PE, and IgG2a-FITC is the control for CCR5-GAM-FITC. (B) To calculate the CCR5 quantification lymphocytes were gated in the forward against side scatter, the second panel live cells were gated. In the third panel CD4+ cells were gated with activated and unactivated cells using HLA-DR-PE. CCR5 positive cells are labelled with CCR5 Goat-anti-Mouse antibodies, CCR5 quantification was determined for both the activated (HLA-DR+) and unactivated (HLA-DR-) cells.

CCR5 receptor densities per cell were quantified by comparing FITC staining of cells to a standard curve. The standard curve was created by analysis of a mixture of five calibrated bead populations, which are coated with five different quantities of immunoglobulins/bead

(Qifikit). Beads were treated like cells are hence followed the similar staining and incubation steps. When using the fluorescence intensity of each bead population, this was plotted against the designated number of immunoglobulins/bead. Thereafter, five data points were interpolated by linear regression. We included the HLA-DR marker so as to differentiate between activated and non activated cell (Figure 3.2). Therefore, using CD3/HLA-DR CD4-T-lymphocyte and CD14 monocyte populations were defined as CCR5 by comparison to the corresponding isotype control. To calculate mean fluorescence intensity the regression formula was used to convert the geometric mean of the fluorescence intensity of these CCR5 cell populations to the mean density of receptors per cell (123).

3.5. Genotyping Methods

3.5.1. Copy number variation of CCL3L and CCL4L

Nomenclature of *CCL3L* and *CCL4L* chemokine genes are as illustrated (figure 7) and described above (Introduction section on page 28). DNA was isolated from EDTA-treated whole blood by using a Qiagen kit. The copy number of *CCL3L* (previously designated as the *CCL3L*-containing segmental duplication and includes *CCL3L/2/3*) was genotyped as described previously (97). Similar Taqman - based Real-time PCR assays were developed to genotype the copy number of the other chemokine genes. Briefly, the human housekeeping *beta-globin* (*BGB*) gene was used as a control, as this gene possess two copies per diploid genome (pdg). The A431 cell line was used as a standard for *CCL3La*, *CCL4L* and *CCL4La* with each gene possessing two copies pdg ((293) and data not shown). The K562 cell line was used as a standard for *CCL3Lb* and *CCL4Lb* with one and two copies pdg, respectively. *CCL4La* and *CCL4Lb* were run in a duplex assay whereas the others were run in single-plex. Each sample was run in triplicate in three separate 384-well plates. Rounded average numbers were used for analysis, and quality control procedure was applied as described previously (97). Primer and probe sequences are listed in table 3.4.

Table 3.4: Primer and probes sequences used to calculate the CNV for *CCL3L* and *CCL4L* genes and their respective components.

Gene	Primers/Probes	Oligonucleotide sequences*
CCL3L	Sense primer	5'-TCTCCACAGCTTCCTAACCAAGA-3'
	Antisense primer	5'-CTGGACCCACTCCTCACTGG-3'
	Probe	6FAM-AGGCCGGCAGGTCTGTGCTGA-MGBNFQ
CCL3La	Sense primer	5'-GGGTATGACTTCTTGAACCGACAAA-3'
	Antisense primer	5'-GGTTCTCTGTTTCTCTATGTGATCCA-3'
	Probe	6FAM-CAAAGTGTTGGGATTACA-MGBNFQ
CCL3Lb	Sense primer	5'-CATCCACTCGCTCACACCTGTA-3'
	Antisense primer	5'-GCGGTCGGCGTGTCA-3'
	Probe	6FAM-AGAGTTGGGCTTATTCT-MGBNFQ
CCL4L	Sense primer	5'-CATGGTCAGGCAGAGGAAGATG-3'
	Antisense primer	5'-GCTTGCCTCTTTTGGTTTGGAAT-3'
	Probe	6FAM-TACCACAGGCAAGGGAT-MGBNFQ
CCL4La and *CCL4Lb*	Sense primer	5'-GGAAGATGCCTACCACAGGC-3'
	Antisense primer	5'-GCGCAGACTTGCTTGCC-3'
	CCL4La Probe	VIC-CTTGTTCTACaGATTCC-MGBNFQ†
	CCL4Lb Probe	6FAM-CTTGTTCTACgGATTCC-MGBNFQ†
BGB	Sense primer	5'-TCGCTTTCTTGCTGTCCAATTTCTA-3'
	Antisense primer	5'-ATGCTCAAGGCCCTTCATAATATCC-3'
	Probe	VIC-CCTAAGTCCAACTACTAAACTG-MGBNFQ

* VIC and FAM, probe reporter fluorescent dyes. MGBNFQ, Molecular-Groove Binding Non-fluorescence Quencher hybridization probes which allow for using probes at lower melting temperatures. †Probe sequences in lower case represent difference between *CCL4La* and *CCL4Lb*.

3.5.2. Single Nucleotide Polymorphisms

The SNPs for CCR5 was genotyped by TaqMan allelic discrimination assays. Briefly, a real time PCR assay is used to detect the presence of a SNP that can either be a homozygous mutant, heterozygous, or homozygous wild type. The mutant probe sequence binds onto the mutant polymorphism, while the wild type probe sequence binds onto the wild type polymorphism. Each probe has a dye attached to it and when amplified the probe fluoresces, hence the SNP can be scored accordingly. Table 3.5 shows the list of primers used for the CCR5 polymorphisms.

Table 3.5: Primer and Probe sequences used to determine the CCR5 SNPs

Gene	Primers/Probes	Oligonucleotide sequences*
CCR5 -29	Sense	5'-TCA TGT GGA AAA TTT CTC ATA GCT TCA GA-3'
	Antisense	5'-GAG GAC TCA CAC TAT GCC AGA TAC-3'
	Wild Type Probe	VIC -AGT GAA GAA TCC TGC CAC C- MGBNFQ
	Mutant Probe	6FAM -TGA AGG ATC CTG CCA CC- MGBNFQ
CCR5 -676	Sense	5'-CCA GAG ATC TAT TCT CTA GCT TAT TTT AAG C-3'
	Antisense	5'-TGT ATT GAA GGC GAA AAG AAT CAG-3'
	Wild Type Probe	VIC -CAA CTT AAA AAG AAG AAC TGT- MGBNFQ
	Mutant Probe	6FAM -AAC TTA AAA GGA AGA ACT GT- MGBNFQ
CCR5 -927	Sense	5'-CCT GTT AGT TAG CTT CTG AGA TGA GTA AA-3'
	Antisense	5'-CCA AAC TGT GAC CCT TTC CTT ATC T-3'
	Wild Type Probe	VIC -TTT GCC AAA TGT CTT CT- MGBNFQ
	Mutant Probe	6FAM -TTT GCC AAA TAT CTT CT- MGBNFQ
CCR5 - D32	Sense	5'-CAT TAC ACC TGC AGC TCT CAT TTT-3'
	Antisense	5'-CGA GTA GCA GAT GAC CAT GAC AA-3'
	Wild Type Probe	6FAM -CAT ACA GTC AGT ATC AAT TC- MGBNFQ
	Mutant Probe	VIC -CCA TAC ATT AAA GAT AGT C- MGBNFQ
CCR5 -208	Sense	5'-CCG TGA GCC CAT AGT TAA AAC TCT T-3'
	Antisense	5'-CAC AGA TGC TCA CCA CCC AAT ATT A-3'
	Wild Type Probe	VIC -CAA CAG GTT GTT TCC GT- MGBNFQ
	Mutant Probe	6FAM -ACA ACA GGT TTT TTC CGT- MGBNFQ
CCR5 -303	Sense	5'-GGG TGG TGA GCA TCT GTG T-3'
	Antisense	5'-GCC AAC TTA AAC CAA CTT TAA ATG TAG AGG-3'
	Wild Type Probe	VIC -CCT GTG CCC CCT TT- MGBNFQ
	Mutant Probe	6FAM -CCC TGT GTC CCC TTT- MGBNGQ
CCR2 V64I	Sense	5'-TCT TTG GTT TTG TGG GCA ACA TG-3'
	Antisense	5'-AGG TAA ATG TCA GTC AAG CAC TTC A-3'
	Wild Type Probe	VIC -CTG GTC GTC CTC ATC- MGBNFQ
	Mutant Probe	6FAM -CTG GTC ATC CTC ATC- MGBNFQ
CCR5 630*	Sense	5'GGTTAATGTGAAGTCCAGGATCC-3'
	Antisense	5'AACAGTTCTTCTTTTTAAGTTGAGCTTAAAATAAGCTAGA GAATAGATCTCTGGTTT 3'

***** The real time PCR assay for the CCR5 C630T polymorphism was not designed therefore a restriction fragment length polymorphism (RFLP) assay was used, the conditions for the assay are described previously (96). *DraI* was the restriction enzyme used.

DARC T-46C (rs 2814778) was run using the above explained (page 94) TaqMan allelic discrimination assay. The primers and probes used for this assay are as follows (114):

DARC-46-Sense: 5'-CTGATGGCCCTCATTAGTCCTT-3'

DARC-46-Antisense: 5'-AGGCCCGCAGACAGAAG-3'

DARC-46-Wt Probe: VIC 5'_CTTCCAAGaTAAGAGCC_3'

DARC-46-Mut Probe: FAM 5'_CCAAGgTAAGAGCC_3'

RESULTS

CHAPTER 4 - PAPER ONE

Duffy-Null-Associated Neutropenia Influences HIV-1 Infection Rate in High-Risk South African Black Women

Paper accepted for publication in Journal of Clinical Infectious Diseases

Ramsuran, V. *et a*l. (2011). Duffy-Null-Associated Low Neutrophil Counts Influence HIV-1 Susceptibility in High-Risk South African Black Women. Clinical Infectious Diseases. **In Press**.

Duffy-Null-Associated Neutropenia Influences HIV-1 Infection Rate in High-Risk South African Black Women

Veron Ramsuran[1,2,3,4]*, Hemant Kulkarni[3,4]*, Weijing He[3,4]*, Koleka Mlisana[2], Edwina J. Wright[5], Lise Werner[2], John Castiblanco-Quinche[3,4], Rahul Dhanda[3,4], Matthew J. Dolan[6], Weihua Guan[7], Robin A. Weiss[8], Robert A. Clark[3,4], Salim S. Abdool Karim[2], Sunil K. Ahuja[3,4†], Thumbi Ndung'u[1,2†]

[1]HIV Pathogenesis Programme, [2]Centre for the AIDS Program of Research in South Africa (CAPRISA), Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, 4013, South Africa
[3]Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, San Antonio, TX 78229
[4]Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229
[5]The Alfred Hospital, Melbourne, Victoria Australia 3004, The Burnet Institute, Melbourne Victoria Australia 3004, and Departments of Medicine, Nursing Health Sciences, Monash University, Clayton, Victoria Australia 3800
[6]Henry M. Jackson Foundation, Wilford Hall United States Air Force Medical Center, Lackland AFB, TX 78236
[7]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455-0378
[8]Medical Research Council Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London, London, W1T 4JF, United Kingdom

*Equal contributions. †To whom correspondence should be addressed: ahujas@uthscsa.edu (S.K.A.) or ndungu@ukzn.ac.za (T.N.)

Author Contributions

Conception and design of the study	- VR, HK, WH, SKA, TN
Performed Lab work	- VR
Analysis and interpretation of data	- VR, HK, WH, RD, SKA
Collection and assembly of data	- VR, WH, JCQ
Drafting of the article	- VR, WH, SKA
Statistical expertise	- HK, LW, RD, WG
Critical revision and approval	- VR, KM, EJW, MJD, RAW, RAC, SSAK, SKA, TN

ABSTRACT

Apart from the previous observation in the Multicenter AIDS Cohort Study that Caucasian men who resisted HIV-1 infection had higher blood neutrophil and CD8+ T-cell counts than those who became infected, the influence of levels of peripheral blood cells on HIV risk is unknown. We evaluated this question in a prospective study of Black South African high-risk women, including commercial sex workers. Pre-seroconversion neutrophil counts in women who subsequently seroconverted were significantly lower, whereas platelet counts were higher, compared with those who remained HIV-negative. Comprising 27% of the cohort, subjects with initial neutrophil counts of <2,500 cells/mm^3 had a ~3-fold greater risk of acquiring HIV. In genome-wide association analyses, an African-specific polymorphism (rs2814778) in the promoter region of Duffy Antigen Receptor for Chemokines (*DARC −46T>C*) was highly predictive of neutrophil counts (P_d=7.9x10^{-11}); a statistically significant association for platelet counts was not detected. Consistent with the prevailing viewpoint that the *DARC −46C/C* genotype that imparts the Duffy-null state on erythrocytes is a strong determinant of ethnic neutropenia of African ancestry, only *DARC −46C/C*-bearing study participants had initial neutrophil counts of <2,500 cells/mm^3. Before or after adjustment for platelet counts and population admixture, the risk and rate of acquiring HIV were four-fold (P=0.007) or three-fold (P=0.005) higher, respectively, in those with the trait of Duffy-null-associated low neutrophil counts compared with all other study participants. Due tothe high prevalence of this trait among persons of African ancestry, it may be a major contributor to the dynamics of the HIV epidemic in Africa.

INTRODUCTION

Over 15 years ago, Detels *et al*, [1] demonstrated the influence of circulating levels of leukocyte subsets on resistance to acquiring HIV-1 in men followed prospectively in the Multicenter AIDS Cohort Study (MACS). They showed that resistance to acquiring HIV despite documented heavy viral exposure, was associated with high neutrophil and CD8+ T cell counts [1]. Although significant attention has been placed on investigating the role of CD8+ T cells, the observed association of low neutrophil counts with increased HIV risk has been largely ignored.

To determine the association of levels of peripheral blood cell (PBC) counts with HIV risk in subjects of African descent, a well-characterized prospective cohort of high-risk women (HRW) including commercial sex workers (CSW), from South Africa was evaluated [2]. We surmised that if this association were to be affirmed it might have special relevance to the HIV epidemic because many people of African ancestry have low white blood cell (WBC) counts, mostly attributable to low neutrophil counts, a trait designated 'ethnic leukopenia/neutropenia' [3-5].

RESULTS

Hematologic parameters and risk of HIV infection

Evaluation of the association of the initial baseline values of the three major PBCs, namely WBC, red blood cells (RBC), and platelets, with the future risk of acquiring HIV infection revealed that a higher baseline WBC count was associated with a reduced risk of acquiring

HIV (Table 1, left). Specifically, each 1,000 cell/mm^3 increment in the baseline WBC count was associated with a 28% reduction in the risk of subsequently acquiring HIV [OR=0.72, 95% confidence interval (CI) = 0.55–0.93]. By contrast, an initial high platelet count was associated with an increased future risk of acquiring HIV with each additional 100,000 cell/mm^3 increment in the platelet count associating with a ~2 fold increased risk of acquiring HIV (Table 1, left). A significant association between RBC counts with HIV risk was not detected (Table 1). Similar results were obtained after removing subjects classified as outliers from the multivariate model (Table 1, right).

Table 1: Association of major peripheral blood cell (PBC) components with future risk of acquiring HIV.[1]

	All Subjects			**Excluding Outliers**		
PBC	**OR**[2]	**95% CI**	***P***	**OR**[2]	**95% CI**	***P***
RBC	1.08	0.29 – 3.95	0.911	1.19	0.30 – 4.76	0.801
Platelet	2.31	1.31 – 4.08	0.006	2.61	1.38 – 4.93	0.003
WBC	0.72	0.55 – 0.93	0.014	0.68	0.51 – 0.91	0.009

[1]Unconditional multivariate logistic regression modeling using baseline PBC values from 142 individuals; an outlier was defined as a subject with at least one PC that was beyond six standard deviations from the population mean for that PC. . [2]OR, odds ratio (an odds ratio that is >1 or <1 indicates that a higher value of the hematological parameter is associated with a higher or lower risk of acquiring HIV infection, respectively). ORs are per 10^6 RBCs, and 10^5 platelets or 10^3 WBCs (all in cells/mm^3). CI, confidence interval. *P*, significance value.

We next determined which specific leukocyte subset contributed to the association observed for the WBC count. To address this, we replaced the baseline WBC count in the multivariate logistic regression model with the baseline counts of its major components (e.g., neutrophils, lymphocytes, monocytes, eosinophils and basophils). This model revealed that each 1,000 cell/mm^3 increase in the baseline neutrophil count was associated with a 36% lower risk of subsequently acquiring HIV (OR=0.64; 95% CI=0.45–0.94; Table 2, Model 1). An association between the baseline values of other WBC components and HIV infection was not detected (Table 2, Model 1). In a backward elimination stepwise regression model, the only two PBCs that remained statistically significant were platelet and neutrophil counts (Table 2, Model 2). In the multivariate logistic regression models, when we replaced the baseline lymphocyte count with baseline counts for CD4+ or CD8+ T cells, only neutrophil and platelet counts were retained in the final models (data not shown), suggesting that among the PBCs analyzed, baseline levels of these two hematological parameters associate strongly with future risk of acquiring HIV in this study population.

Table 2: Association of PBC counts and WBC components with future risk of acquiring HIV.

	All Subjects			Excluding Outliers		
PBC	**OR**	**95% CI**	***P***	**OR**	**95% CI**	***P***
Model 1: Baseline (pre-seroconversion) cell counts at study entry						
RBC	1.11	0.27 – 4.54	0.887	1.19	0.27 – 5.31	0.817
Platelet	2.48	1.42 – 4.33	0.004	2.77	1.44 – 5.35	0.002
Neutrophils	0.64	0.45 – 0.94	0.021	0.62	0.42 – 0.93	0.022
Lymphocytes	0.61	0.28 – 1.33	0.217	0.58	0.25 – 1.32	0.194
Monocytes	1.28	0.89 – 1.89	0.212	1.18	0.78 – 1.81	0.429
Eosinophils	2.01	0.25 – 16.2	0.511	1.91	0.22 – 16.9	0.560
Basophils	0.89	0.74 – 1.07	0.209	0.91	0.75 – 1.06	0.345
Model 2: Final model from backward stepwise regression						
Platelet	2.15	1.22 – 3.80	0.008	2.40	1.31 – 4.40	0.004
Neutrophils	0.69	0.50 – 0.95	0.024	0.65	0.46 – 0.92	0.016
Model 3: Model 2 adjusted for PC1 – PC10 from EIGENSTRAT						
Platelet	2.73	1.44 – 5.17	0.002	2.90	1.48 – 5.69	0.002
Neutrophils	0.64	0.46 – 0.90	0.011	0.58	0.39 – 0.86	0.006

OR, odds ratio (an odds ratio that is >1 or <1 indicates that a higher value of the hematological parameter is associated with a higher or lower risk of acquiring HIV infection, respectively). ORs are per 10^6 RBCs, 10^5 platelets, 10^3 neutrophils or lymphocytes, 100 monocytes, and 10 basophils (all in cells/mm^3). Model 1, full unconditional multivariate logisitc regression model. Model 2, backward elimination stepwise regression model with a probability retention criterion of 0.1. Model 3, final model after adjustment for PC1 to PC10 from EIGENSTRAT.

Three lines of evidence demonstrated that the associations of baseline neutrophil or platelet counts with HIV risk were unlikely to be due to differences in population substructure of the HIV-infected versus HIV-negative HRW/CSW. First, none of the mean scores for the top 10 PCs for HIV-positive and HIV-negative groups differed significantly (Fig. 1B). Second, the associations for neutrophil and platelet counts observed in the final model of the stepwise regression for PBCs shown in Table 2 (model 2) remained unchanged after inclusion of the PCs as a covariate in this regression model (Table 2, model 3). Third, all the aforementioned associations remained unchanged upon excluding from the multivariate model subjects who were classified as outliers (Table 2, right).

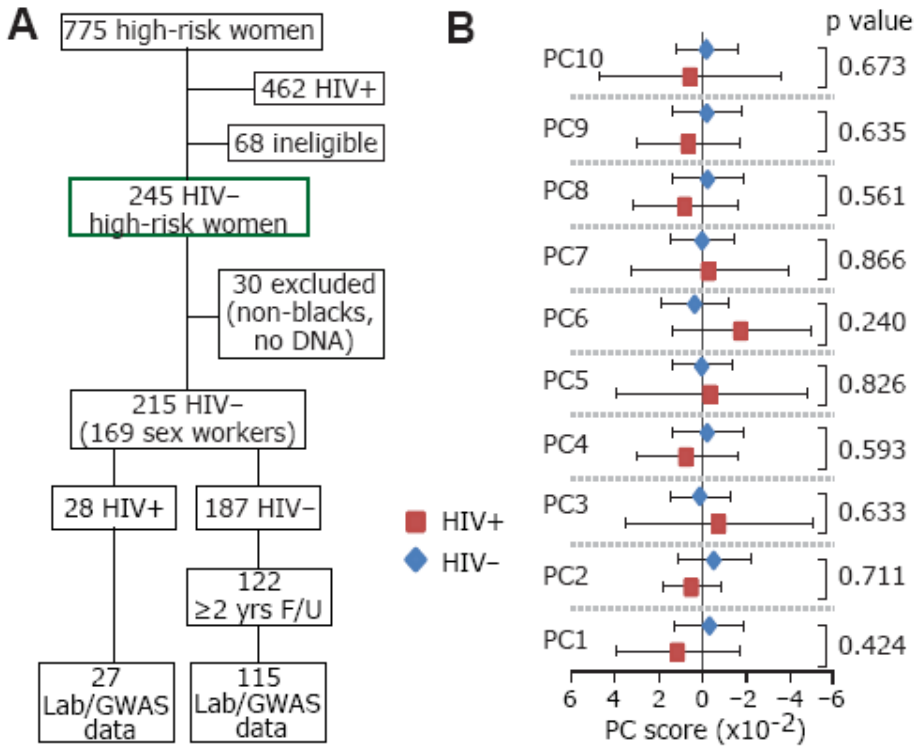


Figure 1: Study subjects and principal component analyses for population stratification. (A) Study subjects. CAPRISA 002 Acute Infection Study screened 775 high risk women who self-identified as commercial sex workers or who reported more than 3 sexual partners in the

prior 3 month [2]. 462 of these were HIV-positive subjects and 68 met exclusion criteria as described previously (e.g. pregnant, declined follow-up for two years) [2]. 245 subjects (green box) were enrolled into the Acute Infection study [2]. 30 were excluded for the present study because they were either not black or DNA was unavailable. Of the 215 HIV-negative women who were followed prospectively, 169 (78.6%) were self-reported sex workers and at the end of the two year follow up period 28 women had seroconverted and 122 remained HIV seronegative (exposed uninfected). 65 women who had less than 2 years of follow-up were excluded from the current analyses. PBC and Genome Wide Association study (GWAS) data was available on 27 HIV-seroconverting women and 115 HIV non-seroconverting women. (B) Evaluation of population stratification in the study groups. Red squares and blue diamond's indicate the mean PC score for the first 10 principal components (ordered from left to right) for HIV-infected and -uninfected subjects, respectively. Error bars represent the 95% confidence interval. Numbers at the right side are significant values obtained using Student's T test.

Neutrophil count threshold for increased HIV risk

Although the aforementioned findings suggested that lower neutrophil counts were associated with increasing HIV risk, thereafter to determine whether this risk was distributed proportionately across a range of neutropenic values or was more accurately represented as a threshold effect. Normal ranges of human blood differential counts vary depending on age, gender, population group, and other factors, and there are inter-laboratory differences in reference intervals. Nonetheless, 2,500 neutrophils/mm^3 is widely used as the lower limit of normal [6], and a neutrophil count of <1,500 cells/mm^3 has become the commonly accepted

definition of neutropenia [5]. Taking into account the general guidelines, we determined the threshold of baseline neutrophil counts below which future risk of HIV increased (Fig. 2A). Approximately 31% of the HRW/CSW with a baseline neutrophil count of <1500 or 1,500-2,500 cells/mm^3 subsequently seroconverted whereas only ~15% of those with >2,500 neutrophils/mm^3 did so (Fig. 2A). By logistic regression analyses, in addition to platelet counts, an initial neutrophil count of <2,500 cells/mm^3 associated with a ~ 3-fold greater future risk of subsequently acquiring HIV infection before or after adjusting for population admixture (Fig. 2B). Thus, in our study population, an initial neutrophil count of 2,500 cells/mm^3 reflected a threshold for altered HIV risk.

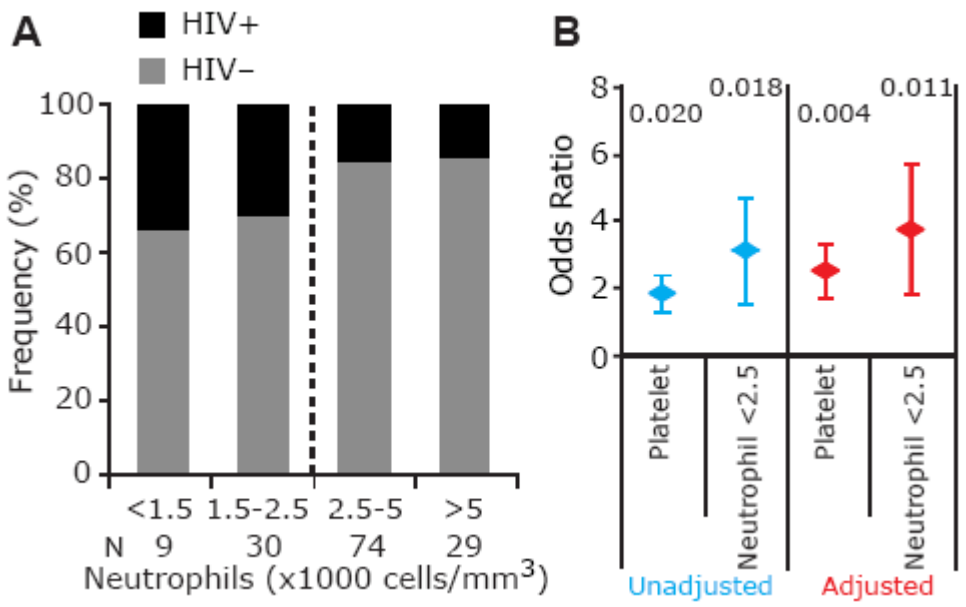


Figure 2: Association of the initial/baseline neutrophil and platelet counts with risk of acquiring HIV infection. (A) Proportion of HIV-positive and HIV-negative subjects in the indicated categories of baseline pre-seroconversion neutrophil counts. (B) Multivariate logistic regression analyses for the association of platelet counts and low neutrophil counts (defined as <2,500 neutrophils/mm^3) before (blue) and after (red) adjustment for potential

population stratification. Adjustment was done by including the ten principal components as covariates in the logistic regression model. Numbers at the top are the significance values.

Genome-wide association for neutrophil and platelet counts

Because of the observed associations of neutrophil and platelet counts with HIV risk, we used a GWAS approach to identify polymorphisms that associate with variability in these two traits. GWAS analyses for neutrophil and platelet counts as quantitative outcomes revealed only one genetic marker, rs2814778 on chromosome 1 that associated significantly with neutrophil counts, and it represents the T-46C polymorphism in the promoter region of the gene for Duffy Antigen Receptor for Chemokines (*DARC*; $P=1.4x10^{-8}$; Fig. 3A). Even after removal of subjects classified as outliers and adjusting for the top ten PCs, the association of the *DARC* polymorphism with neutrophil counts remained highly significant ($p=7.9x10^{-11}$). This *DARC* polymorphism explained 26% (R^2 by univariate model) of variability in neutrophil counts. An association for platelet counts that met the genome-wide level for statistical significance was not detected (data not shown).

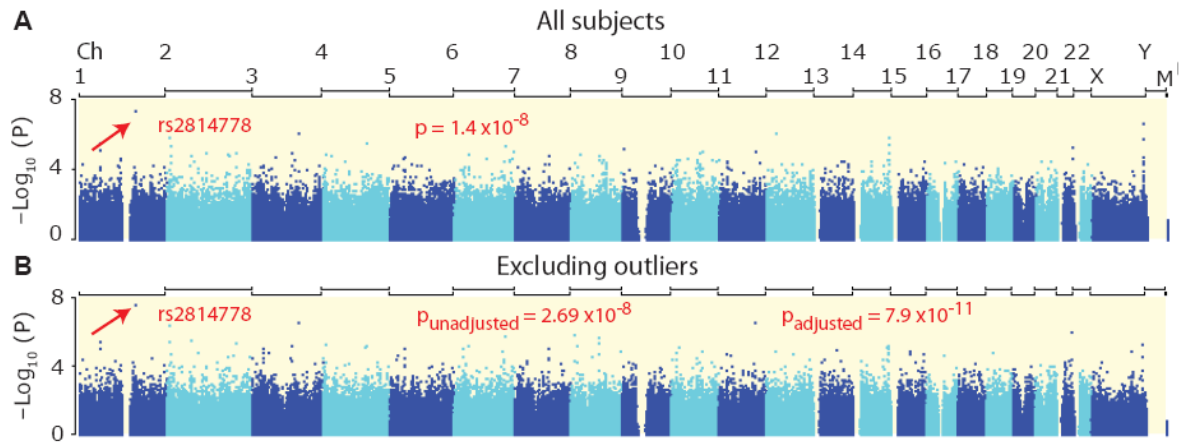


Figure 3: GWAS for the traits of neutrophil and platelet counts. Plots indicate the $-\log_{10}$ P value of the association statistic plotted by the chromosomal location of the marker (x-axis).

Chromosome numbers are shown at the top. Red arrows indicate the marker with strongest association that was statistically significant at the genome-wide significance threshold level of $5.8x10^{-8}$. Results are shown as the significance value based on the Wald statistic for all subjects (A) and as significance values obtained using multivariate linear regression analyses after removing 4 subjects who were classified as outliers. (B). In panel B, $P_{unadjusted}$ and $P_{adjusted}$ indicate the significance values obtained using the unadjusted and adjusted (for the top ten principal components) models, respectively.

DARC −46C/C, neutrophil count and risk of HIV acquisition

An association of *DARC -46C/C* genotype with low neutrophil and WBC counts in persons of African ancestry has been described previously [7-13]. This genotype is also associated with selective loss of DARC expression on RBC (DARC-negative or Duffy-null trait on RBC) but not endothelial cells, and resistance to *P. vivax* malaria [14]. However, because a low neutrophil count also associated with increased HIV susceptibility in our study population, and He *et al*, previously reported that *DARC −46C/C* associated with increased risk of acquiring HIV infection among African-Americans [15], we surmised that HIV risk might differ according to whether DARC-negative (*DARC −46C/C*) subjects had a low versus high initial neutrophil count.

Consistent with the GWAS findings, DARC-negative subjects had significantly lower baseline neutrophil counts than DARC-positive individuals (Table 3). Although the leukopenia associated with the *DARC −46C/C* is attributable mainly to low neutrophil counts [12, 13], subjects bearing this genotype also had lower monocyte counts (Table 3).

Table 3: Association of PBC counts with *DARC* genotype.

PBC (cells/mm^3)*	DARC+			DARC−			*P*
	N	Mean	SE	N	Mean	SE	
RBC (x10^6)	50	4.32	0.05	92	4.37	0.038	0.453
Platelet (x10^5)	50	3.22	0.11	92	3.02	0.88	0.179
WBC (x10^3)	50	7.99	0.27	92	5.90	0.19	<0.001
Neutrophils (x10^3)	50	5.01	0.21	92	3.03	0.17	<0.001
Lymphocytes (x10^3)	50	2.25	0.09	92	2.23	0.07	0.807
Monocytes (x10^2)	50	0.46	0.02	92	0.36	0.01	<0.001
Eosinophils (x10^2)	50	0.20	0.02	92	0.26	0.03	0.386
Basophils (x10)	50	0.06	0.01	92	0.05	0.003	0.273

*Mean values of initial pre-seroconversion hematological parameters in HRW/CSWs possessing the *DARC −46C/C* genotype (DARC− on RBC) compared to those lacking this genotype (DARC+ on RBC).

Predictably, the cumulative distributions of neutrophil counts in DARC-negative (*DARC −46C/C*) and DARC-positive (i.e., those with *DARC −46 T/T* or *DARC −46 T/C* genotypes) subjects were strikingly different (Fig. 2A). Remarkably, none of the 50 DARC-positive subjects had a baseline neutrophil count <2,500 (Fig. 4A). This low level of neutrophil counts was found only in DARC-negative subjects, and those DARC-negative subjects with a low neutrophil count comprised nearly 27% of the entire cohort (Fig. 4B). Notably, 2,500 neutrophils/mm^3 is the same neutrophil threshold that associated with altered risk of acquiring HIV infection (compare Fig. 4B and Fig. 2A). Predictably then, compared with all other subjects, those DARC-negative subjects with a baseline neutrophil count of <2,500 cells/mm^3 had a 2.6-fold higher risk of acquiring HIV (OR=2.61, 95% CI=1.09-6.24;

P=0.028), resulting in an overrepresentation of those with the 'DARC-negative-low baseline neutrophil' genotype-phenotype relationship among subjects who seroconverted, contrasting with an underrepresentation of this genotype-phenotype correlate among those who remained HIV-negative (Fig. 4C). Additionally, those with DARC-negative-associated low baseline neutrophil counts seroconverted nearly 2.4 times faster than all other subjects (Fig. 4D), suggesting that this DARC-related phenotype is associated with both an increased risk and accelerated rate of acquiring HIV infection.

Ideally, the comparator group for DARC-negative subjects with <2,500 neutrohils/mm^3 is DARC-positive subjects with <2,500 neutrohils/mm^3. However, no such individuals were present in our study population. Thus, we used alternative statistical approaches in which the entire spectrum of baseline neutrophil counts was used to determine the association of the trait of *DARC −46C/C*-associated low baseline neutrophil counts with HIV risk. These analyses indicated that either before or after accounting for the baseline platelet count and population admixture, this trait associated with an approximately 4- (OR=4.09; 95% CI=1.46-11.4; P=0.007) or 3- (relative hazard=3.37; 95% CI=1.45-7.82; P=0.005) fold increased risk and rate of acquiring HIV infection, respectively.

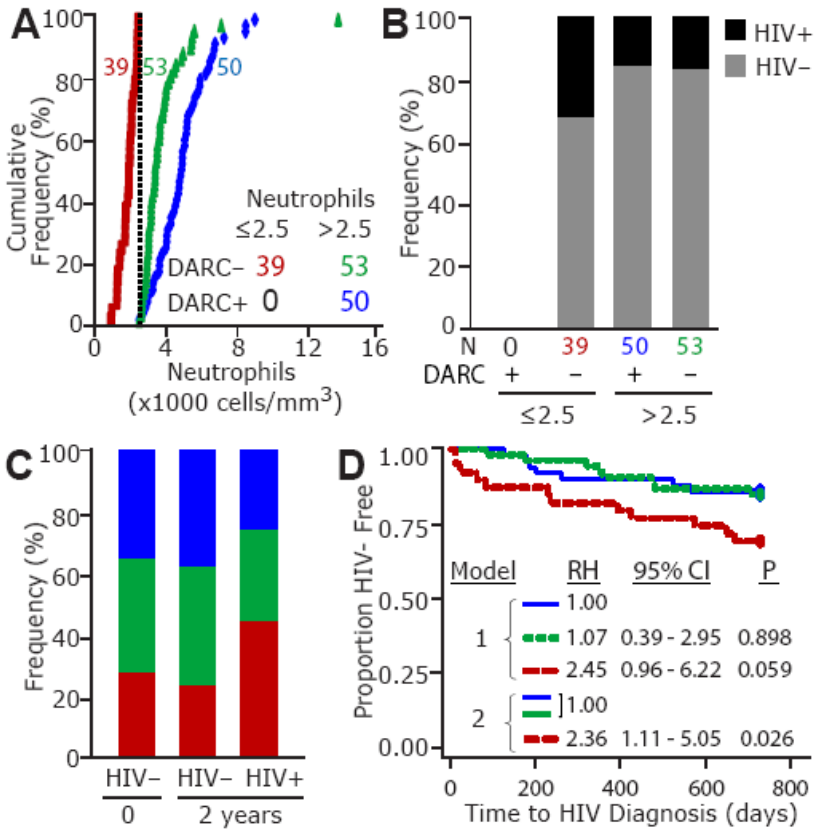


Figure 4: Association of *DARC −46C/C*-associated low neutrophil counts with risk and rate of acquiring HIV infection. (A) Cumulative frequency distribution of neutrophil counts based on whether subjects were DARC-negative or DARC-positive and had a baseline neutrophil count of ≤ or >2,500 cells/mm^3. The number of subjects in each of the four possible *DARC* genotype-neutrophil groups is shown and color-coded to match the frequency plots. (B) Proportion of HIV-positive and HIV-negative subjects according to *DARC* genotype and baseline neutrophil counts of ≤ or > 2,500 cells/mm^3. (C) Prevalence of DARC-negative-low neutrophil (red), DARC-negative-high neutrophil (green), and DARC-positive subjects (blue) phenotypes at enrollment (time 0 day), and 2 years after enrollment. (D) Kaplan-Meier plots for time to HIV diagnosis from enrollment into the cohort for the same 3 color coded groups shown in panel A.. RH, relative hazard, CI, confidence interval and P, significance values derived by Cox proportional hazard models. In model 1, the reference (RH=1) for the Cox

models are DARC-positive subjects. In model 2, comparison of persons with DARC-negative-low baseline neutrophil versus all other subjects (RH=1).

DISCUSSION

The epidemiological characteristics of the cohort of high-risk HIV-negative women/commercial sex workers (HRW/CSW) from South Africa that we evaluated made it uniquely suited for investigating whether pre-infection differences in hematological parameters associated with the future risk of acquiring HIV infection [2]. There are three main findings of this study.

Firstly, baseline high platelet and low WBC counts, attributable mainly to a low neutrophil count, especially <2,500 neutrophils/mm^3, associated with increased HIV susceptibility. HRW/CSW with this baseline neutrophil profile, comprising 27% of the cohort at enrollment, had a nearly three-fold increased risk of acquiring HIV compared with those having higher initial neutrophil counts. These associations persisted after adjustments for potential population admixture in the study sample. These findings are consistent with the observation made in high-risk Caucasian men that a high neutrophil count is associated with HIV resistance [1].

Secondly, GWAS was used to identify genetic markers that associate with variable neutrophil and platelet counts, and these analyses demonstrated that an African-specific polymorphism that associates with the Duffy-null or DARC-negative trait and *P. vivax* malaria resistance

[14] is a strong predictor of neutrophil counts. A statistically significant association for platelet counts was not observed.

Thirdly, remarkably, DARC-negative (*DARC -46C/C*) subjects with <2,500 neutrophils/mm^3 had a ~2.5-fold increased risk and rate of acquiring HIV relative to all other study participants. It was not possible to determine whether DARC-positive subjects with this low level of neutrophil counts also had an increased HIV risk because such subjects were not represented in the cohort. Because this comparator group (DARC-positive subjects with <2500 neutrophils/mm^3) was not represented in our study population, we employed a complementary statistical approach to determine the relationship of the trait of *DARC*-null-associated low neutrophil counts with HIV risk. These analyses demonstrated that this trait associated with a ~2.5-fold higher risk and 3-fold accelerated rate of acquiring HIV.

The association of the Duffy-null state with low neutrophil counts in the African population we evaluated is consistent with the findings from several other cohort-based studies [7, 9, 10, 12, 13]. The most conclusive of these are the analyses by Reich et al [13] who used an exhaustive admixture mapping approach to demonstrate that over and above its strong association with African ancestry, *DARC -46C/C* may be a causal variant for the leukopenia/neutropenia observed in an African American cohort. Although the leukopenia associated with *DARC −46C/C* is explained mainly by reduced neutrophil levels, we found that HRW/CSWs with this genotype also had lower monocyte counts. The latter relationship of the Duffy-null trait with low monocyte counts is in keeping with other reports [8, 11, 13], and interestingly, a high correlation between neutrophil and monocyte cell counts, but not among the levels of other leukocyte subsets, has been observed [6].

Despite extensive data suggesting that over and above its association with African ancestry, the *DARC −46C/C* may be strongly associated with ethnic leukopenia/neutropenia [7-13], the precise mechanisms of this association are unknown. Many chemokine-related functions have been proposed for erythroid DARC, ranging from its serving as a sink or reservoir for chemokines, such as IL-8, and maintenance of chemokine plasma concentrations [reviewed in [14]]. A close relationship between erythrocyte DARC and chemokine levels is also underscored by a GWAS showing that a polymorphism (Asp42Gly) in the *DARC* coding region is a major regulator of the *DARC*-mediated erythrocyte binding of MCP-1, IL-8 and RANTES and in turn, the circulating levels of these proteins [16]. Additionally, mice overexpressing DARC on blood vessel endothelium display enhanced IL-8-induced neutrophil extravasation, further underscoring a close relationship between DARC expression and neutrophil homeostasis [17]. Thus the neutropenia associated with the DARC-negative trait may relate to DARC's impact upon levels of chemokines that influence neutrophil trafficking [e.g. IL-8 [18]] during basal/resting conditions, and possibly following inflammation [11], However, other mechanisms underlying ethnic neutropenia have also been suggested [3, 4].

The association of platelet counts with HIV risk is consistent with multiple lines of evidence suggesting that platelets provide a key interface between coagulation and inflammation/immunity [19, 20], which in turn may play an important role in HIV risk; notably these pathways have also been implicated in HIV disease outcome [21]. Similarly, burgeoning data implicate neutrophils as the first defenders against infections, including viral infections, and in playing a critical role in instructing the adaptive immune responses relevant

to HIV pathogenesis [reviewed in [22]]. Additionally, neutrophils are a major source of α-defensins 1–4 which may play a role in the control of HIV replication [23]. Finally, akin to the strong association between increased levels of the inflammatory marker C-reactive protein (CRP) with enhanced morbidity/mortality in HIV-infected persons [21, 24], the elevated CRP levels found among HIV-negative healthy subjects with a low neutrophil count [8] may correlate with increased HIV risk. Thus, a low neutrophil count associated with the DARC-negative state may alter the inflammatory or innate immune response at the time of virus exposure, promoting HIV acquisition.

Of note, reports of neutropenia occurring in the setting of acute HIV seroconversion are uncommon [25, 26]. Nonetheless, to evaluate the possibility that the cause of low neutrophil counts in those patients who seroconverted soon after study enrollment was secondary to acute HIV seroconversion, we examined the differences in patients' post- and pre-seroconversion neutrophil counts. We found that regardless of when patients seroconverted after enrollment into the study there was no correlation between the time to HIV diagnosis and differences in the pre- and post-seronversion neutrophil counts (data not shown). Hence, the associations we detected even among those who seroconverted soon after enrollment were unlikely to be confounded by effects of viral infection on neutrophil counts.

These results suggest that in the high-risk subjects we evaluated there is a close triangulation among a pre-infection low neutrophil count, Duffy-null state and HIV risk. However, because of the close association of the Duffy-null state with low neutrophil counts, it is not possible to ascribe an independent effect of low neutrophil counts or the DARC-null state on HIV risk, and hence these two host factors may as act either independently or as a conjoint genetic-

cellular determinant. In addition, because at this juncture the full repertoire of the genetic variants that influence leukocyte levels remains incomplete, the WBC count or specific leukocyte levels (e.g. neutrophil counts) represent a surrogate for the effects of these variants. With this in mind, we conjecture that the associations of *DARC −46C/C* with HIV risk observed in this study may reflect the sum of two tiered effects: first, those related to the cellular (e.g., low neutrophil/monocyte counts) and inflammatory (e.g., chemokine binding) phenotypes associated with this genotype; and second, gene-gene interactions involving *DARC* and other genes that regulate leukocyte levels. However, as previously reported by He *et al*, [15], because HIV suppressive chemokines [e.g., RANTES; [14]] and the virus itself [14, 15, 27] also bind to DARC and DARC-bound HIV can be transferred in *trans* to CD4+ T cells [15], several concurrent mechanisms may underlie the observed association of the Duffy-null state and increased HIV susceptibility.

He *et al.* showed that *DARC −46C/C* is associated with an increased HIV acquisition risk among African Americans [15]. The present results suggest that the increased HIV risk associated with the DARC-negative state is mainly in those who also have a low neutrophil count, comprising 27% of the study population. The prior associations of He et al were not observed by others [9, 28-30]. Nonetheless, as noted in their riposte [31], He et al explained that contrary to the contention of others [9, 28-30, 32], their reported association was not due to population stratification, and suggested that the disparate associations of *DARC* genotype with HIV-AIDS susceptibility may be explained partly by differences in cohort characteristics and endpoints evaluated.

The importance of being mindful of cohort characteristics, especially when evaluating the associations for *DARC* genotype, is highlighted by the present findings, as we demonstrate that a positive association is not dependent on differences in the absolute distribution of subjects with *DARC −46C/C* in cases versus controls, but rather on the relative distribution of subjects with *DARC −46C/C*-associated low neutrophil counts. However, the relative representation of subjects with *DARC −46C/C*-associated low versus high neutrophil counts and DARC-positive individuals in HIV+ vs the comparative HIV− study groups may shift substantially depending on the epidemiologic characteristics of the HIV+ and HIV− subjects selected (e.g., when comparing unreferenced HIV-negative subjects with seroprevalent or seroconverting patients selected on the basis of specific clinical/laboratory characteristics). This may explain why some studies did not discern an association of *DARC −46C/C* with HIV susceptibility [28-30]. Similar considerations apply to the evaluation of the association of *DARC −46C/C* with HIV disease course because subsequent to the initial demonstration of this association by He *et al* [15], Kulkarni *et al* found that the influence of the Duffy-null state on clinical outcome was evident mainly in Duffy-null subjects who at HIV diagnosis during the early stages of their infection also had low WBC counts [10].

Although the cohort of HRW/CSW was highly suited to address the questions posed, our study has limitations. The sample size was small, and we could not account for the effect of intercurrent sexually transmitted infections upon HIV acquisition risk. Notwithstanding these limitations, as reported previously [2, 33] and noted above, the cohort we evaluated has many strengths that tend to offset these limitations. The study design was not a cross-sectional comparison of HIV-negative versus HIV-positive subjects; such a comparison would preclude the ability to assess the associations of the pre-infection trait of *DARC -46C/C*-associated low neutrophil counts or PBC levels with future risk of acquiring HIV. Instead, the

study participants were derived from a cohort of high-risk HIV-negative women, including commercial sex workers, who were enrolled when they were HIV-negative and then, prospectively and longitudinally evaluated in an intensive manner.

Additionally, the study participants were recruited into the cohort after screening a large pool (n=775) of high-risk women with a much higher prevalence of HIV infection (~60%) than the prevalence rates of HIV among other South African women (range: 13.3-33.3%) [2]. In light of this, the HIV-negative women who enrolled into this study are likely to be enriched for host genetic or other factors that provide relative protection against HIV acquisition, including those factors that associate with a protective high neutrophil and low platelet profile. These genetic-epidemiologic features may (i) attenuate the future risk of acquiring HIV infection, and (ii) preclude the identification of host genetic or other factors that associate with an increased risk of acquiring HIV, unless their effect sizes are large. Hence, this cohort favors the identification of those host factors with strong effects on HIV susceptibility despite its small sample size.

The epidemiologic considerations noted above have relevance for the evaluation of vaccine trials as they are typically conducted in subjects at high risk for acquiring HIV, similar to those studied herein. However, at the end of the two year observation period, subjects who remained HIV-negative were less likely to bear unfavorable host factors (e.g. DARC-negative associated low neutrophil phenotype), and were further enriched for favorable factors (e.g. DARC-positive genotype). Hence, when evaluating a vaccine in subjects enriched for protective genetic-cellular factors, misallocation of favorable versus unfavorable genetic-cellular factors in trial arms might result in spurious estimates of a vaccine's efficacy

because it might be difficult to differentiate between protective effects of the HIV vaccine versus those of protective genetic-cellular factors.

In summary, we extend the observation made in high-risk Caucasian men that a high neutrophil count is associated with HIV resistance [1], suggesting that the intrinsic cellular environment, as defined by quantitative differences in hematologic indices, influences the risk of acquiring HIV infection. In a cohort of Black HRW/CSW, we found that low neutrophil and high platelet counts associated with an increased risk of acquiring HIV. Using a GWAS approach we confirm that the Duffy-null state associates strongly with a low neutrophil count, and demonstrate that among Duffy-null subjects those who also have a low neutrophil count have an increased risk and rate of acquiring HIV infection. Hence, we suggest that the trait of Duffy-null-associated with low neutrophil counts is associated with HIV susceptibility. Given the high prevalence of this unfavorable genotype-hematological determinant among persons of African ancestry, this association may contribute substantially to the HIV epidemic in Africa.

METHODS

Study cohort

We evaluated a cohort of HRW/CSW from Durban, KwaZulu-Natal, South Africa who were part of the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 acute infection study [2]. The characteristics of the study subjects, as described previously [2], are summarized in Fig. 1A.

Explanatory variables

We used baseline peripheral blood cell (PBC) counts as predictors of subsequent risk of HIV acquisition, where baseline refers to values obtained when all subjects were seronegative for HIV at entry into the cohort. Genetic data were obtained from a genome-wide association study (GWAS; see below).

Genetic and Statistical analyses

Associations of the baseline values of the major PBCs [i.e., red blood cells (RBC), platelet and WBC] with subsequent risk of acquiring HIV infection were determined using multivariate unconditional logistic regression models. In subsequent multivariate logistic

regression models, we replaced the baseline WBC count with the baseline values of its major components i.e., neutrophil, lymphocyte, monocyte, eosinophil and basophils, or replaced the baseline lymphocyte counts in the latter model with baseline counts of CD4+ and CD8+ T cells. We tested the statistical significance of these associations at an alpha error rate of 0.05.

To account for population admixture, we identified 3160 autosomal ancestry-informative markers (AIMs) for persons of African ancestry from two reports [34, 35], and were also contained within the Illumina Human 1M-Duo chip. This set of markers included the *DARC T-46C* (rs2814778) polymorphism and 86 other markers that were significantly correlated with it ($P<0.05$). Removal of these 87 markers gave a set of 3073 AIMs (list available on request). We compared the values of the top ten principal components (PCs) derived by EIGENSTRAT analysis [36] between HIV-infected and -uninfected subjects. We used two additional measures to correct for potential population admixture, specifically adjustment of the multivariate models using the scores for the first ten PCs and exclusion of subjects classified as outliers (n=4) on the basis of the top ten PCs. An outlier was defined as a subject with at least one PC that was beyond six standard deviations from the population mean for that PC.

As the multivariate analyses showed associations between baseline counts for both platelets and neutrophils and future risk of acquiring HIV, we conducted a GWAS using these two hematological parameters as quantitative traits. All 142 samples satisfied the criterion of a call rate exceeding 0.97. We restricted the data to polymorphisms exhibiting a minor allele frequency of at least 0.05 and a Hardy-Weinberg equilibrium (HWE) significance value of >0.001 in the HIV-uninfected subjects. Of the 1,144,696 markers genotyped using the

Human Duo-1M Illumina chip according to the Infinium HD protocol (Illumina, San Diego, CA), there were 95 markers with at least one missing genotype, 1,292 markers were in Hardy-Weinberg disequilibrium at a significance level of 0.001 and 268,498 markers had minor allele frequency of <0.05. Thus, the markers included in this study were 874,956. Then using PLINK software [37], we determined the association of each polymorphism with the quantitative trait making use of the asymptotic Wald test. We also conducted the GWAS analyses using multivariate linear regression models (the linear command in PLINK) by adjusting for the top ten PCs. The statistical significance of the GWAS associations after Bonferroni correction was $p<5x10^{-8}$. Time to HIV identification was evaluated by Kaplan-Meier survival plots and Cox proportional hazard modeling. Additional statistical analyses were conducted using Stata 7.0 (Stata Corp, College Station, TX).

Association of DARC-negative-associated low neutrophil counts with HIV risk

Our results pointed towards the possibility of a trait that is determined by possession of the *DARC -46 C/C* genotype, i.e., a low absolute neutrophil count. The DARC-negative-associated with low neutrophil trait is designated here as the D-N trait that may be associated with an increased risk of acquiring HIV. While it is impossible to delineate the independent and interactive influence of these two parameters on HIV susceptibility, we attempted to quantify the degree to which an individual might possess the D-N trait. For this we first used univariate logistic regression analysis to estimate the probability of possessing the *DARC -46C/C* genotype for a given neutrophil count. A high value for this probability can also be interpreted as the probability of possessing the latent D-N trait.

We therefore examined if a high value for this probability (p) associates with an increased risk of HIV acquisition. For this, we first determined an optimum cut-off value above which *p* will be associated with the risk of acquiring HIV. Using receiver operating characteristic curve (Figure 5) we found that the best cut-off for this probability was 0.8630 above which there was a higher likelihood of acquiring HIV. To assess the association of this dichotomized latent trait probability with HIV susceptibility we conducted two sets of analyses: first, we determined the outcome of the risk of acquiring HIV (using logistic regression analyses) and second, we determined the outcome of the time to infection (using Cox proportional hazards modeling). For each of these analyses, we used four models with the following specifications: Model 1: dichotomized *p* alone as the predictor; Model 2: dichotomized *p* with scores for top 10 principal components (based on EIGENSTRAT procedure for population admixture) as covariates; Model 3: model specified in model 2 with platelet count as a covariate; and Model 4: model specified in model 3 replicated on 5000 bootstrap samples. We observed (Table 4) that a high probability for the D-N trait was associated with a significantly increased risk of both the risk of HIV and the time to infection in all four models.

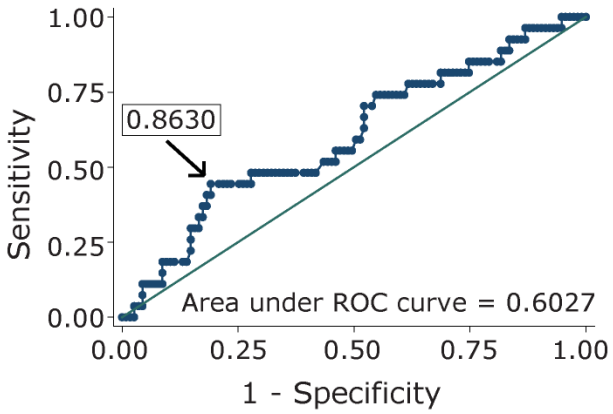


Figure 5: ROC curve for determining the cut-off at which the probability of possessing the DARC-negative-low neutrophil count trait is associated with an increased risk of acquiring HIV.

Table 4: Association of the D-N trait with susceptibility to HIV

Model	Description	Risk of HIV acquisition			Time to seroconversion		
		OR	95% CI	P	RH	95% CI	P
1	Trait prob>0.863033	2.88	1.20 – 6.93	0.018	2.57	1.20 – 5.50	0.0015
2	1 adjusted for PCs	2.92	1.15 – 7.38	0.024	2.65	1.19 – 5.91	0.017
3	2 adjusted for platelets	4.09	1.46 – 11.4	0.007	3.37	1.45 – 7.82	0.005
4	Bootstrapping for model 3	3.60	1.30 – 9.98	0.014	2.98	1.32 – 6.75	0.009

These results strongly support the notions that: i) a combination of low neutrophil counts along with possession of the *DARC -46C/C* genotype influences the risk of acquiring HIV; and ii) this risk is further increased in the context of increased platelet counts.

Acknowledgements

We thank the study participants, CAPRISA clinical and laboratory staff for providing specimens. Special acknowledgements to the following members of the CAPRISA Acute Infection Study team: C. Williamson, L. Morris, C. Gray and W. Hide. The CAPRISA 002 study was supported by the NIAID, NIH (grant U19 AI 51794). Veron Ramsuran was a recipient of a CAPRISA training fellowship and a Fogarty AITRP fellowship (TWO-0023) and a KwaZulu-Natal Research Institute for TB and HIV (K-RITH) travel award. Thumbi Ndung'u is supported by the Hasso Plattner Foundation and holds the South African DST/NRF Research Chair in Systems Biology of HIV/AIDS. This work was also supported by the VA HIV/AIDS Center of the South Texas Veterans Health Care System, NIH (R37046326), and the Doris Duke Distinguished Clinical Scientist Award to S.K.A. S.K.A. is also supported by a VA MERIT award and the Burroughs Wellcome Clinical Scientist Award in Translational Research.

References

1. Detels R, Liu Z, Hennessey K, et al. Resistance to HIV-1 infection. Multicenter AIDS Cohort Study. J Acquir Immune Defic Syndr **1994** Dec;7(12):1263-9.
2. van Loggerenberg F, Mlisana K, Williamson C, et al. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One **2008**;3(4):e1954.
3. D'Angelo G. Ethnic and genetic causes of neutropenia: clinical and therapeutic implications. Lab Hematol **2009**;15(3):25-9.
4. Haddy TB, Rana SR, Castro O. Benign ethnic neutropenia: what is a normal absolute neutrophil count? J Lab Clin Med **1999** Jan;133(1):15-22.
5. Hsieh MM, Everhart JE, Byrd-Holt DD, Tisdale JF, Rodgers GP. Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. Ann Intern Med **2007** Apr 3;146(7):486-92.
6. Bain BJ, England JM. Normal haematological values: sex difference in neutrophil count. Br Med J **1975** Feb 8;1(5953):306-9.
7. Afenyi-Annan A, Ashley-Koch A, Telen MJ. Duffy (Fy), DARC, and neutropenia among women from the United States, Europe and the Caribbean. Br J Haematol **2009** Apr;145(2):266-7.
8. Grann VR, Ziv E, Joseph CK, et al. Duffy (Fy), DARC, and neutropenia among women from the United States, Europe and the Caribbean. Br J Haematol **2008** Oct;143(2):288-93.
9. Julg B, Reddy S, van der Stok M, et al. Lack of Duffy antigen receptor for chemokines: no influence on HIV disease progression in an African treatment-naive population. Cell Host Microbe **2009** May 8;5(5):413-5; author reply 8-9.
10. Kulkarni H, Marconi VC, He W, et al. The Duffy-null state is associated with a survival advantage in leukopenic HIV-infected persons of African ancestry. Blood **2009** Jul 20;114(13):2570-1.
11. Mayr FB, Spiel AO, Leitner JM, et al. Duffy antigen modifies the chemokine response in human endotoxemia. Crit Care Med **2008** Jan;36(1):159-65.
12. Nalls MA, Wilson JG, Patterson NJ, et al. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. Am J Hum Genet **2008** Jan;82(1):81-7.
13. Reich D, Nalls MA, Kao WH, et al. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS Genet **2009** Jan;5(1):e1000360.
14. Rot A, Horuk R. The Duffy Antigen Receptor for Chemokines. In: Inc. E, ed. Methods in Enzymology Vol. 461, **2009**:191-206.
15. He W, Neil S, Kulkarni H, et al. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. Cell Host Microbe **2008** Jul 17;4(1):52-62.
16. Schnabel RB, Baumert J, Barbalic M, et al. Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. Blood **2009** Jul 1;115(26):5289-99.
17. Pruenster M, Mudde L, Bombosi P, et al. The Duffy antigen receptor for chemokines transports chemokines and supports their promigratory activity. Nat Immunol **2009** Jan;10(1):101-8.
18. Wengner AM, Pitchford SC, Furze RC, Rankin SM. The coordinated action of G-CSF and ELR + CXC chemokines in neutrophil mobilization during acute inflammation. Blood **2008** Jan 1;111(1):42-9.
19. Delvaeye M, Conway EM. Coagulation and innate immune responses: can we view them separately? Blood **2009** Sep 17;114(12):2367-74.

20. Torre D, Pugliese A. Platelets and HIV-1 infection: old and new aspects. Curr HIV Res **2008** Sep;6(5):411-8.
21. Kuller LH, Tracy R, Belloso W, et al. Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med **2008** Oct 21;5(10):e203.
22. Nathan C. Neutrophils and immunity: challenges and opportunities. Nat Rev Immunol **2006** Mar;6(3):173-82.
23. Klotman ME, Chang TL. Defensins in innate antiviral immunity. Nat Rev Immunol **2006** Jun;6(6):447-56.
24. Lau B, Sharrett AR, Kingsley LA, et al. C-reactive protein is a marker for human immunodeficiency virus disease progression. Arch Intern Med **2006** Jan 9;166(1):64-70.
25. Colson P, Foucault C, Mokhtari M, Tamalet C. Severe transient neutropenia associated with acute human immunodeficiency virus type 1 infection. Eur J Intern Med **2005** Apr;16(2):120-2.
26. Kinloch-de Loes S, de Saussure P, Saurat JH, Stalder H, Hirschel B, Perrin LH. Symptomatic primary infection due to human immunodeficiency virus type 1: review of 31 cases. Clin Infect Dis **1993** Jul;17(1):59-65.
27. Lachgar A, Jaureguiberry G, Le Buenac H, et al. Binding of HIV-1 to RBCs involves the Duffy antigen receptors for chemokines (DARC). Biomed Pharmacother **1998**;52(10):436-9.
28. Horne KC, Li X, Jacobson LP, et al. Duffy antigen polymorphisms do not alter progression of HIV in African Americans in the MACS cohort. Cell Host Microbe **2009** May 8;5(5):415-7; author reply 8-9.
29. Walley NM, Julg B, Dickson SP, et al. The Duffy antigen receptor for chemokines null promoter variant does not influence HIV-1 acquisition or disease progression. Cell Host Microbe **2009** May 8;5(5):408-10; author reply 18-9.
30. Winkler CA, An P, Johnson R, Nelson GW, Kirk G. Expression of Duffy antigen receptor for chemokines (DARC) has no effect on HIV-1 acquisition or progression to AIDS in African Americans. Cell Host Microbe **2009** May 8;5(5):411-3; author reply 8-9.
31. He W, Marconi V, Castiblanco J, et al. Response: Association of Duffy Antigen Genotypes with HIV-AIDS Susceptibility. Cell Host Microbe **2009** 21 May 2009;5(5):418-9.
32. An P, Winkler CA. Host genes associated with HIV/AIDS: advances in gene discovery. Trends Genet Mar;26(3):119-31.
33. Naicker DD, Werner L, Kormuth E, et al. Interleukin-10 promoter polymorphisms influence HIV-1 susceptibility and primary HIV-1 pathogenesis. J Infect Dis **2009** Aug 1;200(3):448-52.
34. Baye TM, Wilke RA, Olivier M. Genomic and geographic distribution of private SNPs and pathways in human populations. Per Med **2009** Nov 1;6(6):623-41.
35. Reich D, Patterson N. Will admixture mapping work to find disease genes? Philos Trans R Soc Lond B Biol Sci **2005** Aug 29;360(1460):1605-7.
36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet **2006** Aug;38(8):904-9.
37. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **2007** Sep;81(3):559-75.

CHAPTER 5 - PAPER TWO

Unprecedented Genetic Diversity of a Duplicated Chemokine Gene-rich Locus on chromosome 17q12 and its Association with HIV-AIDS susceptibility

In preparation for submission to international journal.

Unprecedented Genetic Diversity of a Duplicated Chemokine Gene-rich Locus on chromosome 17q12 and its Association with HIV-AIDS susceptibility

Veron Ramsuran[1,2,3,4,5]*, Weijing He[1,2]*‡, Hemant Kulkarni[1,2]*‡, Koleka Mlisana[3], Matthew J. Dolan[6-8], Lise Werner[3], Robert A. Clark[1,2]‡, Salim S. Abdool Karim[3], Thumbi Ndung'u[3,4]*†, Sunil K. Ahuja[1,2,9]†‡

[1]Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, San Antonio, TX 78229
[2]Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229
[3]Centre for the AIDS Program of Research in South Africa (CAPRISA), and [4]Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu Natal, Durban, 4013, South Africa
[5]Discipline of Genetics, School of Biochemistry, Genetics, Microbiology and Plant Pathology, University of KwaZulu-Natal, Pietermaritzburg, 3209, South Africa.
[6]Infectious Disease Clinical Research Program (IDCRP), Uniformed Services University, Bethesda, Maryland, 20814
[7]Infectious Diseases Service, and Henry M. Jackson Foundation, Wilford Hall United States Air Force Medical Center, Lackland AFB, TX 78236
[8]San Antonio Military Medical Medical Center (SAMMC), Ft. Sam Houston and Lackland AFB, TX 78234
[9]Departments of Microbiology and Immunology, and Biochemistry, University of Texas Health Science Center, San Antonio, TX 78229

* Equal contributions. ‡Data from the CEPH cohort was obtained from previous work and included here to further support the result obtained here. †To whom correspondence should be addressed: ahujas@uthscsa.edu (S.K.A.) or ndungu@ukzn.ac.za (T.N)

Author Contributions

Conception and design of the study	- VR, HK, WH, SKA, TN
Performed Lab work	- VR
Analysis and interpretation of data	- VR, HK, WH, SKA
Collection and assembly of data	- VR, WH
Drafting of the article	- VR, WH, SKA
Statistical expertise	- HK, LW
Critical revision and approval	- VR, KM, MJD, RAC, SSAK, SKA, TN

Abstract

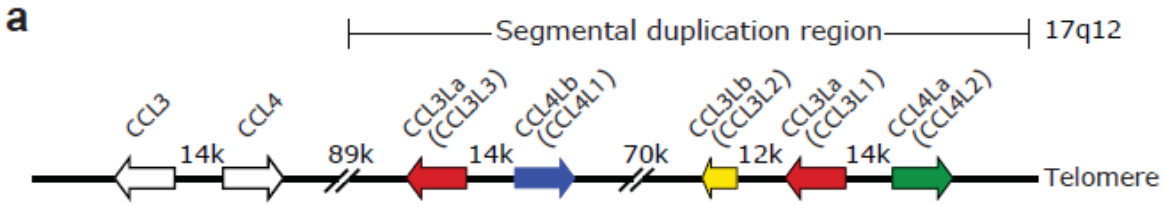
Segmental duplications are selectively enriched for host defense genes such as pro-inflammatory and anti-HIV-1 chemokines. However, the extent of copy number variation (CNV) of distinct genes within a segmental duplication and its impact on disease susceptibility is largely unknown. Populations with a higher content of genes that transcribe secreted chemokines (*CCL3L1*, *CCL4L2*) also have a higher content of genes predicted to encode non-secreted chemokines (*CCL3L2*, *CCL4L1*), raising the possibility that the latter genes may have evolved to offset the pro-inflammatory state associated with a higher dose of genes encoding classical chemokines. Thus, CNV of chemokine genes generates an unprecedented degree of inter-individual and population-specific genomic architectural complexity. Failure to account for this complexity may preclude full characterization of the human genome and obscure the true impact of CNV on disease susceptibility. Illustrating this, in a well-characterized cohort of South African women at high risk for HIV-1, we demonstrate that it is the combinatorial content of *CCL3L*- and *CCL4L*-related genes rather than the gene dose of any one of these genes that affects both HIV susceptibility and disease progression.

INTRODUCTION

Copy number variations (CNV) comprise a substantial proportion of the human genome (22, 30, 31, 42, 52). Although genomic architectural complexity in the form of smaller CNV within larger ones have been described (2, 40), the full extent of this complexity and its impacts on disease susceptibility is unknown. Indeed, although progress has been made in evaluating the contributions of CNV to disease susceptibility (10, 27, 31, 44, 52), defining the phenotypic impact of distinct CNVs within a segmental duplication or copy number variation regions (CNVR), remains a critical but largely unmet challenge (11). This is an important challenge because segmental duplications are enriched for host defense genes (4) and a seemingly similar segmental duplication or CNVR may harbor a distinct repertoire of smaller CNVs that may be population-specific. Consequently, inter-subject differences in the repertoire of these smaller CNVs within a CNVR may result in variable phenotypes or disease associations. Thus, the primary hypothesis of this study is that there is extensive inter-individual and inter-population heterogeneity in the genetic content of a segmental duplication or CNVR and failure to account for this heterogeneity may obscure CNV-disease associations. However, there is a lack of informative disease and CNV model systems to probe the extent to which this genomic architectural complexity at the level of both individuals and populations may influence the design, evaluation and interpretation of CNV-disease association studies.

To address this gap in knowledge, here this study focused on HIV-AIDS as a disease model because in vitro, in vivo and genetic association studies strongly implicate the chemokine-chemokine receptor system, including chemokine genes present within a segmental

duplication, with HIV-AIDS susceptibility. For example, the Δ32/Δ32 genotype of the major HIV coreceptor CC chemokine receptor 5 (CCR5) associates with loss of CCR5 surface expression and striking protection against HIV acquisition (24). The ligands of CCR5 (e.g. CCL3, CCL4) display HIV suppressive effects, and a cluster of chemokine genes with homology to CCL3 and CCL4 designated as CCL3-like (*CCL3L*) and *CCL4L*-like (*CCL4L*) co-localize to a region on 17q12 that has undergone segmental duplications (34, 35, 47) (**Fig. 1a**). There are three *CCL3L*-related genes that are subject to CNV and are designated as *CCL3L*, *CCL3L2* and *CCL3L3*, and the two *CCL4L*-related genes subject to CNV are designated as *CCL4L1* and *CCL4L2* (**Fig. 1a**) (34). Notably, *CCL3L* and *CCL3L3* share the identical coding sequence, and among the CCR5 ligands are the most potent with respect to downregulation of surface CCR5 expression and HIV-suppressive effects (34, 39).



b

Nomenclature used		NCBI (Gene ID)
CCL3L	CCL3La	CCL3L1 (6349)
		CCL3L3 (414062)
	CCL3Lb	CCL3L2 (390788)
CCL4L	CCL4La	CCL4L2 (388372)
	CCL4Lb	CCL4L1 (9560)

Figure 1: Chromosome 17q12 segmental duplication, chemokine gene copy number variation nomenclature and correlation between copy number of *CCL3L* and*CCL4L*. (**a**) Schematic representation of CCL3,CCL4,*CCL3L* and*CCL4L* genes. Arrows indicate the orientation of each gene. Shown on top is the distance between the indicated genes. (**b**) indicates the previous literature used. (Note the difference between the Shostakovich-Koretskaya *et al*, 2009 figure which has the *CCL3L*3 gene swapped with *CCL3L*) unpublished.

Several lines of evidence underscore the biological plausibility that chemokine gene CNV influences HIV-AIDS susceptibility: (i) High copy number (CN) of the *CCL3L*-containing segmental duplication correlates with higher mRNA/protein expression levels of *CCL3L*-related genes, lower proportions of HIV target cells (CD4+CCR5+ T cells) and increased leukocyte chemotaxis (16, 47, 49); (ii) Genetic epidemiologic studies across multiple and varied cohorts (e.g. children vs adults; vertical vs horizontal transmission) showed that a low CN of the *CCL3L*-containing segmental duplication associated with an enhanced risk of acquiring HIV, higher HIV plasma viral load (VL), faster rates of disease progression, reduced HIV-specific immune responses, and lower cell-mediated immune responses (1, 13, 16, 21, 25, 32, 38, 43, 45, 46) and reviewed in Ref. (20)); and, (iii) There is an evolutionary correlation to this association whereupon in the non-human rhesus macaque model of AIDS, a low CN of the *CCL3L*-containing segmental duplication also associates with an accelerated disease course following experimental infection with SIV(12).

However, the *CCL3L*-containing segmental duplication also contains *CCL4L* genes, raising the specter that these genes might also contribute to the ultimate phenotypic impact of this segmental duplication on HIV-AIDS susceptibility. Consistent with this possibility, a recent study demonstrated that it was the combinatorial content of both *CCL3L* and *CCL4L* gene that together influenced HIV-AIDS susceptibility (46), and a separate study showed that the relative content of *CCL4L* genes influenced HIV susceptibility (9).

Hence, to test our hypothesis, we determined whether failure to account for the combinatorial content of chemokine genes obscures their association with HIV-AIDS in a well-characterized cohort of South African women. We chose this cohort because its epidemiological characteristics were ideal to evaluate this association as it included female sex workers (FSW) who were participants of an acute HIV infection cohort, acquiring HIV during prospective follow-up (50).

Materials and Methods

Unique cohort of high-risk women

We evaluated a cohort of women from Durban, KwaZulu-Natal, South Africa, an area which has a very high prevalence of HIV-1 infection (23). These women are at significantly increased risk of acquiring HIV infection because of their high-risk sexual behavioral characteristics (50). Seven hundred and seventy five self-identifying female sex workers (FSWs) or women with more than 3 sexual partners in the past 3 months prior to this study, were screened for entry into a prospective cohort (50). Of these, those who were already HIV-infected (60%) and those who did not meet the (9%) were excluded (e.g., pregnancy) (50). Thirty more HIV-negative women were excluded from genetic analysis on the basis of lack of DNA or being self-reported non-Blacks. The final study group for this genetic study

comprised 215 HIV-negative Black high-risk women, of whom 169 (78.6%) were self-reported FSWs. Reluctance to self-identify as being a sex worker or undertaking transactional sex was considered to be prevalent in the remaining 46 high-risk women. Under the auspices of the Centre for the AIDS Programme of Research in South Africa (CAPRISA), these subjects were prospectively and intensively evaluated over a two-year period to accrue a group of subjects that would be representative of individuals who remain seronegative during the two-year follow-up, despite heavy exposure to HIV. During this two-year period, there were 28 seroincident cases, 122 women remained seronegative, and 65 women did not complete the two-year follow-up and therefore excluded from analyses. Mainly for analyses of evaluating the association between chemokine CNV and viral load and $CD4^+$ T cell loss, we also included a separate group of 32 seroincident cases who were part of other prospective CAPRISA studies on acute HIV infection. The laboratory characteristics (viral load and rate of $CD4^+$ T cell decline) of these 32 subjects did not differ from the 28 seroincident cases from the high-risk cohort. The key characteristics of the HIV- (n=116) and HIV+ (n=59) individuals evaluated in the present study are summarized in a detailed description provided below in Table 1.

Table 1: Characteristics of the study subjects

Characteristic	HIV infected	HIV-uninfected
Whole cohort	**63**	**118**
Site of recruitment		
Durban (N, %)	46 (73.0%)	118 (100.0%)
Vulindlela (N, %)	17 (27.0%)	0 (0%)
Source of data		
Original CAPRISA study (N, %)	25 (39.7%)	118 (100%)
Other CAPRISA studies (N, %)	38 (60.3%)	0 (0%)
Mean age (mean ±SD years)	27.7 (±8.8)	31.1 (±8.7)
Mean age at first sexual encounter (mean ±SD years)	17.4 (±1.9)	17.1 (±2.3)
Mean frequency of coital acts per month (mean ±SD months)*	7.6 (±8.5)	9.8 (±6.2)
Use of any contraceptive method (N, %)	44 (69.8%)	96 (81.4%)
Marital status (N, %)		
Divorced	-	2 (1.7%)
Many partners	20 (31.8%)	64 (54.2%)
Married	3 (4.8%)	6 (5.1%)
Refused to reveal	-	1 (0.9%)
Separated	-	2 (1.7%)
Single and no steady partner	3 (4.8%)	2 (1.7%)
One stable partner	37 (58.6%)	40 (33.9%)
Widowed	-	1 (0.9%)
Education (N, %)**		
0	1 (1.6%)	1 (0.9%)
1	-	1 (0.9%)
2	-	1 (0.9%)
3	1 (1.6%)	1 (0.9%)
4	-	3 (2.5%)
5	-	5 (4.2%)
6	1 (1.6%)	5 (4.2%)
7	2 (3.2%)	9 (7.6%)
8	4 (6.4%)	15 (12.7%)
9	5 (7.9%)	11 (9.3%)
10	11 (17.5%)	13 (11.0%)
11	20 (31.8%)	24 (20.3%)
12	18 (28.6%)	29 (24.6%)
Presence of HBcAb at baseline (N, %)	30 (47.6)	69 (59.0%)
Practices anal sex (N, %)	13 (20.6%)	40 (33.9%)
Female Sex Workers (FSW)†	**19**	**96**
Duration of sex work (mean ±SD months)	45.9 (±48.5)	78.6 (±87.2)
Number of days in a week for which engaged in sex work (mean ±SD days)	3.4 (±1.9)	2.7 (±1.3)
Mean age at which started sex work (mean ±SD years)	27.3 (±9.9)	28.5 (±7.9)

*Number of study subjects for the HIV+ group was 58 and HIV- was 112. **Education level from grade 1 to grade 12, 0: no school education. †The proportion of individuals that are FSW is 76% for the HIV+ group (n=25) and 81.4% for HIV- group (n=118).

Genotyping methods

Nomenclatures of CCL3L and CCL4L chemokine genes are as described in reference 46 and illustrated in fig. 1.. DNA was isolated from EDTA-treated whole blood by using a Qiagen kit. The copy number of *CCL3L* (previously designated as the *CCL3L*-containing segmental duplication and includes *CCL3L/2/3*) was genotyped as described previously (17). Similar Taqman based Real-time PCR assays were developed to genotype the copy number of the other chemokine genes. Briefly, the human housekeeping *beta-globin* (*BGB*) gene was used as an internal control as a gene with two copies per diploid genome (pdg). The A431 cell line was used as a standard for *CCL3La*, *CCL4L* and *CCL4La* with each gene possessing two copies dpg ((47) and data not shown). The K562 cell line was used as a standard for *CCL3Lb* and *CCL4Lb* with one and two copies pdg, respectively (data not shown). *CCL4La* and *CCL4Lb* were run in a duplex assay whereas the others were run in singleplex. Each sample was run in triplicate in three separate 384-well plates. Rounded average numbers were used for analysis, and quality control procedure was applied as described previously (17). Primer and probe sequences are available upon request.

Statistical Methods

In this study, we examined the association of the overall *CCL3L* and *CCL4L* gene copy number and their components (*CCL3La*, *CCL3Lb*, *CCL4La*, and *CCL4Lb*) with three end-points: risk of HIV acquisition, plasma viral load and rates of CD4+ T cell count decline. To

assess the association with the risk of HIV acquisition we used unconditional multivariate logistic regression analyses that included the risk-behavior attributes as covariates.

To examine the association of the CNV with plasma viral load we used two outcomes: serial measurements of plasma viral load and steady-state viral load. The latter was defined as the mean viral load estimated within the interval 24-28 weeks post-seroconversion. To analyze the association of the genotypes with the first outcome (serial viral load measurements) we used generalized estimating equations (GEE) methods for panel data. We first estimated the overall viral load profile using non-linear GEE and then spline-smoothed the curves using knots every three months. We also used a linear GEE model which permitted the prediction of the viral load on the day of serconversion (the intercept from the GEE model). To evaluate the association between the steady state viral load and the gene copy number we used least-squares linear regression. For all the analyses on viral load measurement the viral load was log-transformed to the base 10. To analyze the association of the genotypes with the outcome of rates of CD4 cell count decline, we used linear GEE models estimated separately for subjects belonging to different genotypic groups. The rates of CD4 decline (slopes estimated from the GEE models) for different genotypic groups were compared using Student's t test.

Upon summarizing the correlation patterns across the components of the *CCL3L* and *CCL4L* genes we conducted factor analysis using the method of principal components. For this, we first extracted significant factors that were defined as those with an eigenvalue exceeding unity and then applying a varimax rotation to the extracted orthogonal factors. In the next step, we generated factor scores for each extracted factor for each individual based on the results obtained from the previous step. We then determined the association of these factor

scores with plasma viral load and rates of CD4 decline using linear regression and linear GEE analyses as described above.

Factor analysis: Factor analysis provides an unbiased exploratory means to understand complex correlation patterns among several *variables* and to distill these complex patterns into a parsimonious mathematical solution called as *factors*. Highly correlated variables load onto a common factor and highly uncorrelated variables segregate onto different factors. We used a method of factor analysis referred to as principal components method to parse out the complex correlation pattern among the components of the duplicated genes on chromosome 17q, i.e., *CCL3La, CCL3Lb, CCL4La* and *CCL4Lb* . For deriving the factor solution we used a criterion of a minimum eigenvalue of 1.

The principal components method yielded the following two factor solution:

(principal factors; 2 factors retained)

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	1.56722	0.36042	0.5933	0.5933
2	1.20680	1.22662	0.4568	1.0501
3	-0.01982	0.09272	-0.0075	1.0426
4	-0.11254	.	-0.0426	1.0000

Factor Loadings

Variable	1	2	Uniqueness
CCL3La	0.83292	0.41390	0.13492
CCL3Lb	0.41024	0.16649	0.80398
CCL4La	0.83740	-0.42533	0.11785
CCL4Lb	-0.06256	0.90932	0.16923

The factor loadings show that there were only two retained factors based on the four variables (genes) – in other words the correlation pattern of the components of the duplicated segment could be summarized using two factors. Since principal components analysis assumes that the factors are orthogonal (i.e., the factors are themselves uncorrelated), we next used the varimax rotation to arrive at an optimized factor solution as follows:

Rotated Factor Loadings

Variable	1	2	Uniqueness
CCL3La	0.89933	0.23725	0.13492
CCL3Lb	0.43541	0.08025	0.80398
CCL4La	0.73430	-0.58562	0.11785
CCL4Lb	0.12229	0.90323	0.16923

The rotated factor loadings suggest that *CCL3La* and the *CCL4La* gene copy numbers loaded heavily on the first factor while the *CCL4Lb* gene copy number loaded onto the second factor. The *CCL4La* gene copy number also strongly negatively loaded onto the second factor. Interestingly, the *CCL3Lb* gene copy number only mildly loaded onto the first factor and minimally loaded onto the second factor.

The uniqueness of each variable, as shown by the rotated factor loadings, is a measure of the degree of correlation of the indicated gene copy numbers with the factors that were mathematically extracted. It can be seen that the uniqueness of *CCL3La*, *CCL4La* and *CCL4Lb* is low indicating that these gene copy numbers are strongly correlated with the two factors. However, the *CCL3Lb* is highly unique and only minimally loads onto either factor.

Therefore, it can be surmised that the *CCL3Lb* gene copy number provides unique information that is distinct from that provided by the other three variables.

In order to translate the results of factor analysis into clinically and genetically meaningful information, we generated factor scores for each individual. In theory, since the variables load onto factors it is possible to estimate the degree to which an individual might load onto that factor given that individual's data on the contributing variables. In our case, based on the results of the factor analysis, we generated the factor scores using the coefficient shown below:

Scoring Coefficients

Variable	1	2
CCL3La	0.57264	0.32764
CCL3Lb	0.05369	0.01721
CCL4La	0.46032	-0.51838
CCL4Lb	0.12902	0.53658

To illustrate this point, assuming that the standardized (z-value/score) results of the Taqman PCR assays for the four components were 1.5, 0.04, 1.2, and -0.5 for *CCL3La*, *CCL3Lb*, *CCL4La*, and *CCL4Lb*, respectively, then that individual's score for the first factor will be (based on the coefficients shown in the table above): (0.57x1.5) + (0.05x0.04) + (0.46x1.2) + (0.12x-0.5) = 1.35. The same individual's score on the second factor will be (0.33x1.5) + (0.01x0.04) + (-0.52x1.2) + (0.54x-0.5) = -0.40. Thus, this individual will have a high score for the first factor and a negative score for the second factor. Of interest, as can be observed

from the scoring coefficients shown in the table above, the *CCL3Lb* gene copy number contributed minimally to the factor scores.

Having generated the scores for each individual, we proceeded to study the association of the factor scores to steady state plasma viral load and the rates of CD4 decline. To assess the association between the factor scores and steady state viral load we regressed the factor scores onto the log-transformed steady state viral load using least-squares linear regression. To assess the influence of the factor scores on the rates of CD4 decline, we first divided the study subjects into three groups based on the tertiles of the factor scores and then estimated, using GEE models, the weekly rate of CD4 cell decline. We also created a composite factor score ratio by taking the ratio of the first factor score and second factor score. Conceptually, this ratio represents the expansion or shrinkage of the first factor score relative to the second factor score. However since both the factor scores can be distributed over the 2-dimensional real number space, we first converted these scores into a positive number by addition of a constant of 2.5. We then took the ratio of the positive factor scores as an indicator of the relative factor scores (Factor 1/Factor 2). Lastly, we examined the association of the factor score ratios to the steady-state viral load and rates of CD4 cell decline using the same analytical strategy as described above.

Lastly, since genetic associations can be confounded due to population substructure, we had genotyped the study subjects for an additional 3150 ancestry informative markers (AIM). Details of the 3150 markers are below. On the marker data, we used the EIGENSTRAT method to extract the first 10 significant principal components (PC) and compared whether the average scores for the ten PCs were differentially distributed across the HIV-infected and

HIV-uninfected subjects. We accounted for the potential population substructure in the multivariate analyses in two ways: first, we adjusted the multivariate models for the ten PCs by including the PC scores as covariates; and second, we identified outliers from the PC structure (defined as a subject with at least one PC score having a relative deviate exceeding 6). We then ran the multivariate models by excluding the outlier subjects from analyses. All the statistical analyses were conducted using Stata 10.0 software package (College Station, TX).

RESULTS

Chemokine genes subject to CNV

As described previously (20, 46), because *CCL3L* and *CCL3L3* have the identical sequences in their coding regions (34) they were designated as *CCL3La*, *CCL3L2* was designated as *CCL3Lb*, *CCL4L2* as *CCL4La*, and *CCL4L1* as *CCL4Lb* (**Figure 1 a,b**). The sequences of the second and third coding exons of the two *CCL3La* genes (*CCL3L* and *CCL3L3*) and *CCL3Lb* are identical, since *CCL3Lb* lacked the first coding exon that is present in the *CCL3La* genes, it has been classified as a pseudogene (**Figure 1**, (34)). We reevaluated this assumption. Using bioinformatic and other approaches we identified novel 5' exons that spliced onto exons 2 and 3 of *CCL3Lb*. The alternatively spliced transcripts of *CCL3Lb* are predicted to encode a non-secreted protein or an atypical chemokine (data not shown). *CCL3Lb* is subject to CNV, and *CCL3L*b transcripts are present only in those individuals who display possession

this gene. These findings underscore that *CCL3Lb* is unlikely to be a pseudogene and its gene products may encode non-classical chemokines.

By analogy to the findings observed for *CCL3L*-related genes, the transcripts of one of the *CCL4L*-related genes (*CCL4Lb*) are also predicted to encode non-classical chemokines (9). This is because of a fixed mutation in the intron-exon splice junction between intron 2 and exon 3 of *CCL4Lb* results in the usage of several new acceptor splice sites and generates several *CCL4L*b mRNAs that are non-secreted isoforms distinct (9). By contrast, *CCL4L*a is predicted to encode a secreted chemokine. Thus, the 'a' and 'b' components of *CCL3L* and *CCL4L* genes encode for secreted classical and non-classical chemokines, respectively.

We developed the following assays to quantify copy number of the following chemokine genes: a) total number of all *CCL3L*-related genes (*CCL3L*a plus *CCL3L*b) and *CCL4L*-related (*CCL4L*a plus *CCL4L*b) genes; and b) individual gene components of *CCL3L* (*CCL3L*a, *CCL3L*b) and *CCL4L* (*CCL4L*a, *CCL4L*b). Using strategies described previously (20), we demonstrate that the six assays we used to assess chemokine CNV were both valid and precise (F**igure 2**).

We developed CNV-specific probes and evaluated the validity of the assays in the following ways. First, since we used different assays for the total copy number and a and b components of both *CCL3L* and *CCL4L*, we argued that if all the assays are valid then the sum of the copy number estimates for a and b components should highly correlate with the assay that measured the total copy number for that gene. Second, we argued that if the assays are

precise then all the assays should provide CN estimates that cluster near integer values. In support of the first argument, we observed (**Figure 2a**) that for both *CCL3L* and *CCL4L* the sum of the estimates for a and b components was highly correlated with the estimate obtained from the assays for the total copies. Analysis of covariance demonstrated that these two variables explained 90-95% variability of each other. Rounding the copy estimates to integer values also showed very high agreement between these variables - weighted Cohen's kappa values were ~93%. Supporting the second argument, we observed that the raw estimates of copy number for each of the six assays showed profound clustering near the integer values (**Figure 2b**). Together, these findings suggested that the six assays we used were both valid and precise.

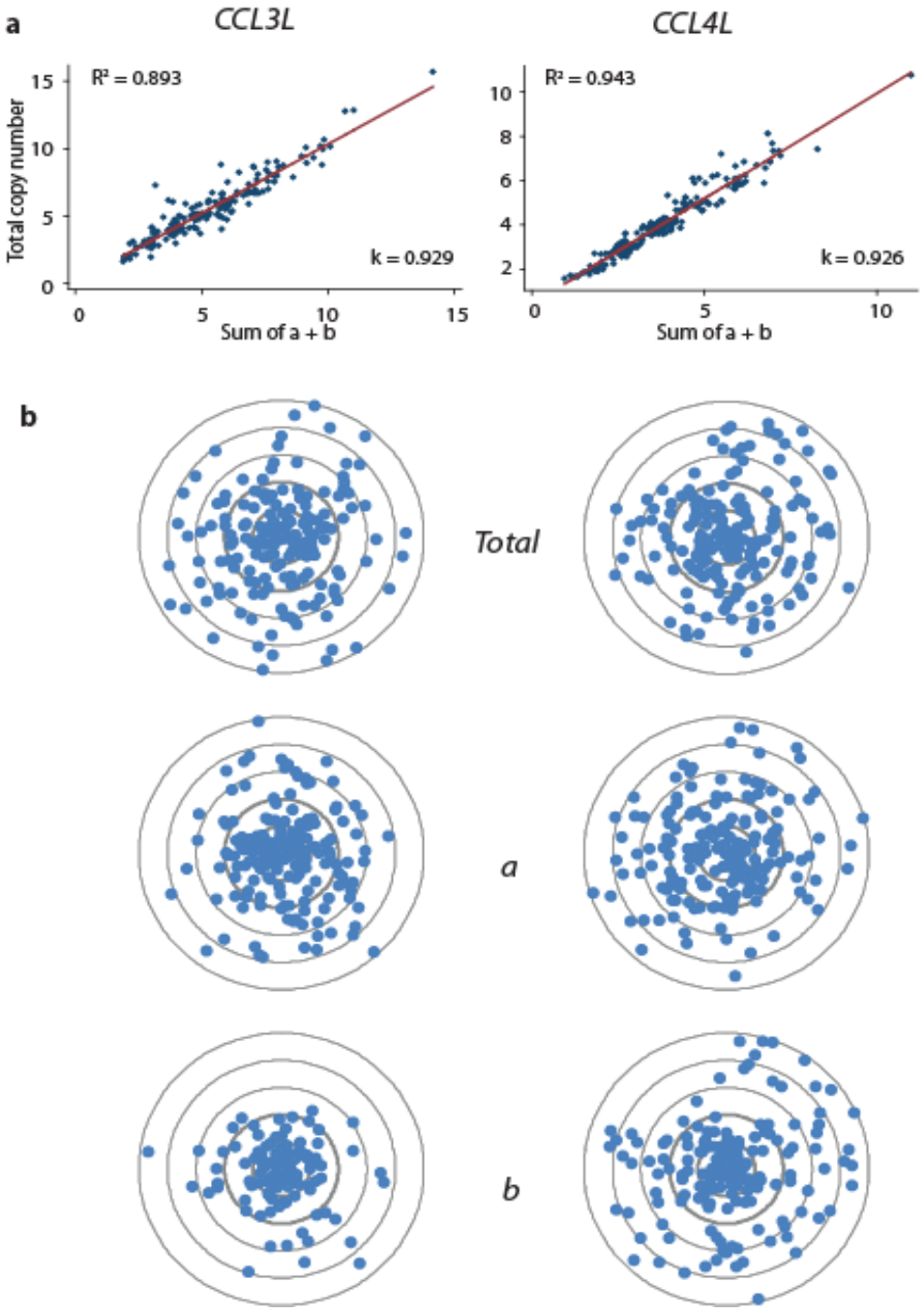


Figure 2. Validity and accuracy of the various CNV assays used in this study. (**a**) Panels show a scatter plot for the estimates of the total *CCL3L-CCL4L* copies obtained by two methods: on the y-axis is the assay that estimated the total number while on the x-axis is the sum of the two assays that reported the number of copies for the a and b components of the *CCL3L*(left) and *CCL4L* (right). Red line is the least squares line. R^2, variability in the total copy number that is explained by the sum of the two estimates for the components a and b; k, weighted Cohen's kappa for the two estimates when the estimated copy number was discretized into integers by rounding. (**b**) Target diagrams demonstrating the clustering of the copy number estimates from the integers. For each plot, the center of the target diagram represents center and the concentric centers indicate the distance from the integer. The concentric circles are placed 0.1 units apart.

Chemokine CNV in a South African cohort

Analyses of the HGDP-CEPH panel revealed that CNV of chemokine genes resulted in extensive inter-individual and inter-population genetic diversity of chemokine gene content (data not shown). How might the greater diversity in African populations' impact on the evaluation of the association of chemokine CNV with HIV-AIDS susceptibility? Despite having a much narrower range of chemokine genetic diversity than Africans, a recent study found that the relative content of specific *CCL3L* and *CCL4L* genes influenced HIV-AIDS susceptibility in a European population (46). In order to evaluate whether a similar paradigm is also present in an African population, we evaluated women mainly from Durban, KwaZulu-Natal, South Africa, an area which has a very high prevalence of HIV-1 infection (23).

The clinical characteristics of the study subjects are well-suited to determine the association of host genotype with HIV-AIDS risk. Foremost, because of their high-risk sexual behavioral characteristics these women that included female sex workers are at significantly increased risk of acquiring HIV infection (50). Under the auspices of the Centre for the AIDS Programme of Research in South Africa (CAPRISA), these subjects were prospectively and intensively evaluated over a two-year period to accrue a group of subjects that would be representative of individuals who remain seronegative during the two-year follow-up, despite heavy exposure to HIV and a separate group that seroconverted during this time-period (**Figure 3a**). During this two-year period, there were 28 seroincident cases, 122 women

remained seronegative (**Figure 3a**). We also included a separate group of 44 seroincident cases who were part of other prospective CAPRISA studies on acute HIV infection (**Figure 3a**).

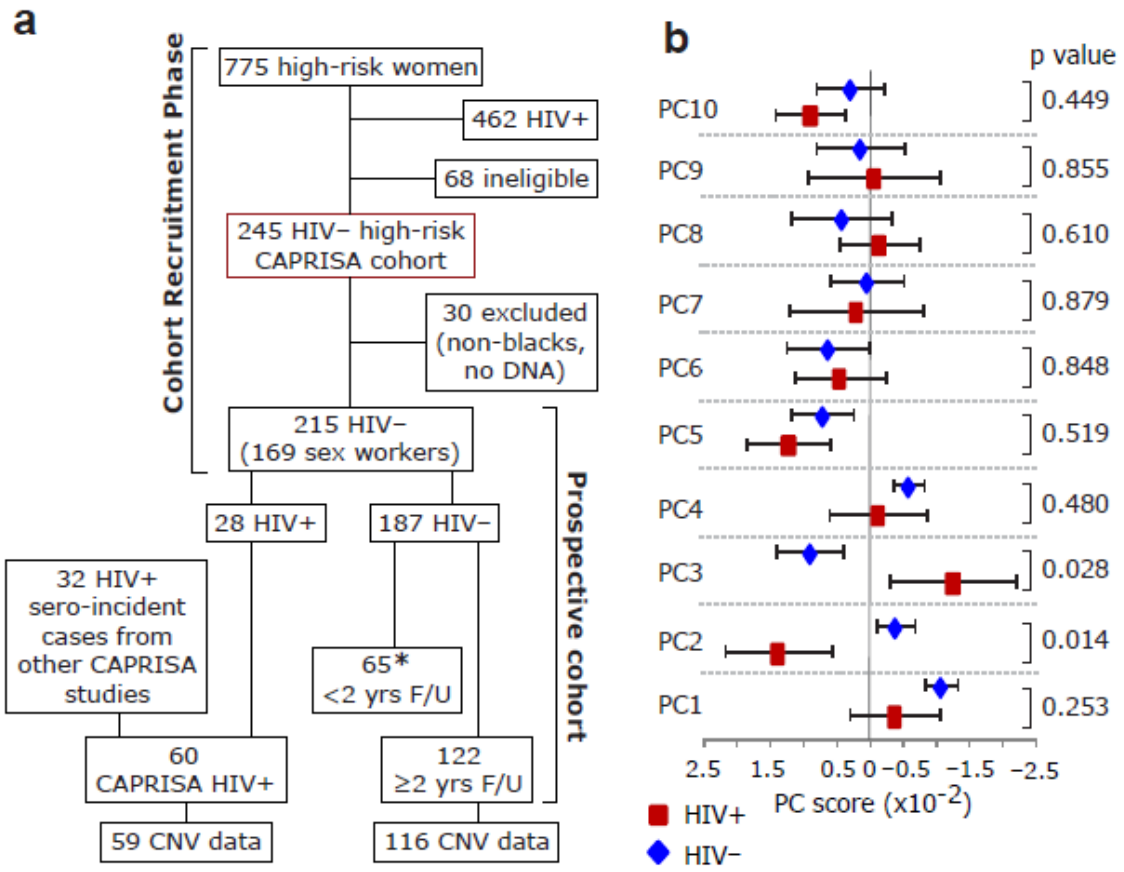


Figure 3: Study subjects and principal component analyses for population stratification. (**a**) Study subjects. CAPRISA 002 Acute Infection Study screened 775 high risk women who self-identified as commercial sex workers or who reported more than 3 sexual partners in the prior 3 month (50). (**b**) Evaluation of population stratification in the study groups. Red squares and blue diamond's indicate the mean PC score for the first 10 principal components (ordered from left to right) for HIV-infected and -uninfected subjects, respectively. Error bars represent the 95% confidence interval. Numbers at the right hand side represent significant values obtained using the Student's T test.

On the basis of 3,150 ancestry informative markers (AIMs), we computed the top ten EIGNSTRAT principal components (PC).We identified 10 subjects (5.7%) who had at least

one PC exceeding 6 relative deviates and excluded them from further analyses. With the exception of PC2, the scores for all the other PCs were comparable between HIV-infected and -uninfected subjects (**Figure 3 b**). These results suggested that there were minimal differences in the degree of admixture between HIV-positive and –negative subjects.

Although the correlation between CN of *CCL3L*-related and *CCL4L*-related genes in the CAPRISA study population was high (R^2 = 0.73; **Figure 4 a**), the R^2 estimates also implied that 27% of the variability in *CCL3L* CN was not accounted for by *CCL4L* CN. The CN of *CCL3L* genes ranged between 2 to 16 copies, with a median of five copies [inter-qaurtile range (IQR) = 4-7) copies), whereas the range was smaller for *CCL4L* genes (median =4, IQR=3-5 copies; **Figure 4 a-d**). These findings along with prior results (46, 47) and those in the HGDP panel emphasized the notion that there are quantitatively more copies of *CCL3L* than *CCL4L* genes and that the CN of *CCL3L*-related genes is not a proxy for the CN of *CCL4L*-related genes. The median copy number of *CCL3L* or *CCL4L* genes was similar in the HIV-positive and HIV-negative subjects (**Figure 4 c, d**).

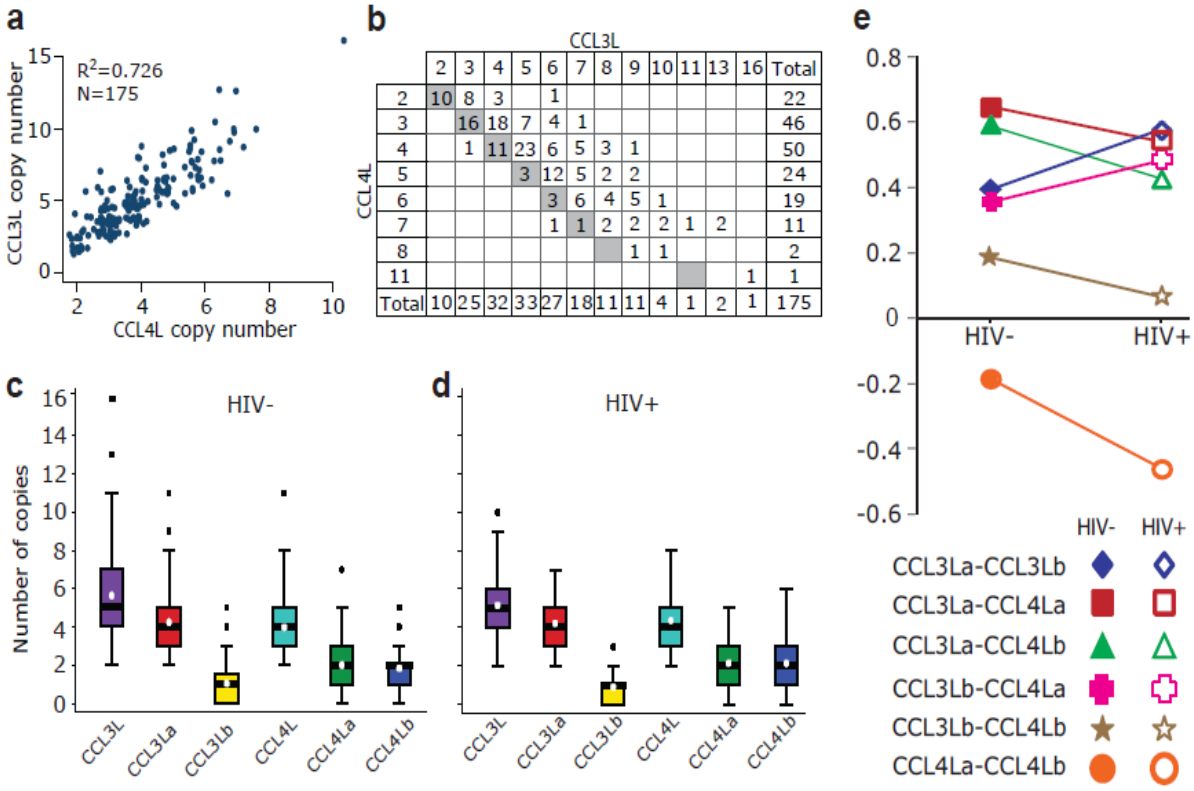


Figure 4: Correlation, distribution and mean/median copy numbers of the *CCL3L*-related and *CCL4L*-related genes. (**a**) Correlation of the *CCL3L* and *CCL4L* genes. (**b**) Distribution of the *CCL3L* and *CCL4L* copies, the *CCL3L* gene has higher copies than the *CCL4L* genes. The median and mean copy number of *CCL3L* or *CCL4L* genes for HIV negative individuals (**c**) and HIV positive individuals (**d**). Correlation patterns for *CCL3L*-related and *CCL4L*-related genes (**e**) of note there is a significant inverse correlation between *CCL4La* and *CCL4Lb.*

Consistent with the results in the HGDP-CEPH panel, examination of the different components of *CCL3L* and *CCL4L* genes revealed that they each had distinct distribution patterns. The gene dose of *CCL3La* (median=4 copies) was on average four times greater than *CCL3Lb* (median=1 copy), whereas the median CN for *CCL4La* and *CCL4Lb* was two (**Figure 4 c,d**). Pair-wise evaluation of the chemokine gene CN revealed that there was a very strong positive correlation between *CCL3La* and *CCL3Lb* ($R = 0.44$, $P < 1x10^{-4}$), *CCL3La* and *CCL4La* ($R = 0.60$, $P < 1x10^{-4}$), as well as between *CCL3La* and *CCL4Lb* ($R = 0.54$, $P < 1x10^{-4}$; **Figure 4e**). In contrast there was a significantly inverse correlation between *CCL4La* and *CCL4Lb* ($R = -0.27$, $P < 1x10^{-4}$; **Figure 4e**). The inverse correlation between copies of

CCL4La and *CCL4Lb* was not a non-specific finding because a similar statistically significant inverse correlation between the CN of *CCL4L*a and *CCL4L*b was observed in the European but not other populations represented in the HGDP-CEPH panel (20) as well as among children from the Ukraine (46). These findings suggested that the genetic structure of the 17q12 chemokine gene-rich region is highly population-specific.

Taken together, the differential distribution pattern of the *CCL3L* and *CCL4L* genes, the complex correlation patterns of the CNs and extensive repertoire of *CCL3L* and *CCL4L* gene combinations in the CAPRISA cohort (i) precluded the evaluation of the association of distinct chemokine gene combinations with HIV risk, and (ii) provided further impetus to test the hypothesis that the overall balance between the genetic content of *CCL3L*- and *CCL4L*-related genes rather than the CN of a single gene component in the *CCL3L* gene-containing segmental duplication may influence HIV susceptibility among subjects represented in the CEPH cohort.

Non-genetic factors that may influence HIV risk

Before undertaking the genetic association analyses, we first determined which non-genetic cohort characteristics were significantly associated with the HIV serostatus. Our analyses, using multivariate logistic regression analyses, revealed that among the parameters evaluated (age at entry into cohort or first sexual encounter, anal sex, contraceptive use, and education level), only a younger age associated with an increased risk of acquiring HIV (P=7.1 x 10^{-6}) whereas predictably, barrier contraceptive use was associated with over 80% reduced risk of

acquiring HIV infection P = 0.0003). We therefore included these two covariates in the statistical models used to determine the association of the CN of *CCL3L*- and *CCL4L*-related genes with HIV risk.

CCL3L and *CCL4L* genes and HIV risk

To determine the association of *CCL3L* and *CCL4L* CN with HIV risk, we conducted a series of bivariate and multivariate logistic regression models. We conducted these analyses using the entire range of copy numbers *CCL3L* and *CCL4L* or after classifying the total *CCL3L* and *CCL4L* CN into high versus low relative to the median copy number present in the study population (**Table 2**). To test our hypothesis, in these regression models the *CCL3L* and *CCL4L* genes were evaluated either alone or in combination.

When examined separately, CN of *CCL3L*- or *CCL4L*-related genes did not associate strongly with HIV risk (**Table 2,** models 1 and 2, respectively). However, when evaluated together in the context of a multivariate logistic regression model, CN of *CCL3L*-related genes associated with a significantly reduced risk of acquiring HIV: each additional copy of a *CCL3L* –related gene associated with a 60% lower risk of acquiring HIV (OR=0.40; 95% CI=0.27-0.59; P<0.001; **Table 2**; model 3). By contrast, each additional CN of *CCL4L*-related genes associated with a nearly 4-fold increased risk of acquiring HIV (OR=3.91; 95% CI=2.21-6.91; P<0.001; **Table 2**, model 3). These association persisted even when age and contraceptive use were included as covariates (**Table 2**, model 4), or when the top ten PCs from EIGENSTRAT were used as additional covariates (**Table 2**, Model 5). These data

suggested that a high CN of *CCL3L*- and *CCL4L*-related genes had opposing associations with HIV risk, and that the genotypic combination of high *CCL3L* and low *CCL4L* gene copies may afford the greatest protection against acquiring HIV and conversely, low *CCL3L* and high *CCL4L* copies may impart the highest risk of acquiring HIV.

Table 2. Association of *CCL3L* and *CCL4L* copy numbers with the risk of HIV acquisition

Model	n	Covariate	OR	95% CI	P
1	175	*CCL3L*	0.91	0.79 - 1.05	0.210
2	175	*CCL4L*	1.18	0.96 - 1.44	0.116
3	175	*CCL3L*	0.40	0.27 - 0.59	<0.001
		CCL4L	3.91	2.21 - 6.91	<0.001
4	175	*CCL3L*	0.43	0.28 - 0.66	<0.001
		CCL4L	3.32	1.77 - 6.23	<0.001
		Age	0.90	0.56 - 0.94	<0.001
		Contraceptive use	0.29	0.11 - 0.72	0.008
5*	175	*CCL3L*	0.38	0.24 - 0.62	<0.001
		CCL4L	3.57	1.80 - 7.09	<0.001
		Age	0.89	0.84 - 0.93	<0.001
		Contraceptive use	0.29	0.11 - 0.81	0.018

*, these models are adjusted for PCs derived from 3150 ancestry informative markers; OR, Odd ratio. P: multivariate logistic regression p values.

To affirm the aforementioned findings, on the basis of whether a subject possessed a CN that was ≥ or < than the median CN of *CCL3L* (5 copies) or *CCL4L* (4 copies) subjects were classified as those possessing a high vs low *CCL3L* copy number ($CCL3L^{high}$ vs $CCL3L^{low}$), or high versus low *CCL4L* copy number ($CCL4L^{high}$ vs $CCL4L^{low}$). Again, we observed that

when examined individually, the associations of *CCL3L*^{high} vs *CCL3L*^{low} or *CCL4L*^{high} vs *CCL4L*^{low} with HIV risk were not statistically significant (**Table 3**; models 1 and 2). However, when examined in unison, compared to those classified as *CCL3L*^{high}, those classified as *CCL3L*^{low} had a nearly 5-fold higher risk for acquiring HIV infection (OR=4.96; 95% CI= 1.70 - 14.5; P = 0.003; **Table 3**, model 3). By contrast, compared to those classified as *CCL4L*^{high}, those classified as *CCL4L*^{low} had a nearly 75% lower risk of acquiring HIV infection (OR=0.25; 95% CI=0.08 - 0.73; P = 0.012; **Table 3**, model 3). These association persisted even when age and contraceptive use were included as covariates (**Table 3**, model 4), or when the top ten PCs from EIGENSTRAT were used as additional covariates (**Table 3**, Model 5).

Table 3. Association of low *CCL3L* and *CCL4L* copy numbers with the risk of HIV acquisition

Model	n	Covariate	OR	95% CI	P
1	175	Low *CCL3L*	1.78	0.94 - 3.38	0.076
2	175	Low *CCL4L*	0.81	0.42 - 1.55	0.528
3	175	Low *CCL3L*	4.96	1.70 - 14.5	0.003
		Low *CCL4L*	0.25	0.08 - 0.73	0.012
4	175	Low *CCL3L*	6.94	1.87 - 25.7	0.004
		Low *CCL4L*	0.21	0.06 - 0.77	0.019
		Age	0.89	0.85 - 0.93	<0.001
		Contraceptive use	0.20	0.08 - 0.49	<0.001
5*	175	Low *CCL3L*	7.57	1.81 - 31.7	0.006
		Low *CCL4L*	0.24	0.06 - 0.99	0.049
		Age	0.88	0.83 - 0.92	<0.001
		Contraceptive use	0.18	0.07 - 0.48	<0.001

*, these models are adjusted for PCs derived from 3150 ancestry informative markers. OR, Odd ratio. P: multivariate logistic regression p values.

Thus, regardless of whether CN was evaluated as a continuous variable (**Table 2**) or dichotomized as high vs low (**Table 3**), the results pointed to the combinations of 'high *CCL3L*-low *CCL4L*' (*CCL3L*^{high}-*CCL4L*^{low}) and 'low *CCL3L*-high *CCL4L*' (*CCL3L*^{low}-*CCL4L*^{high}) associating with decreased vs increased HIV risk, respectively. These data predicted that the former and latter *CCL3L-CCL4L* genotypic groups should be overrepresented and underrepresented in HIV-negative compared with HIV-positive subjects in this study population.

Consistent with this prediction, and demonstrating the importance of accounting for the CN of both *CCL3L* and *CCL4L* genes together, we observed the following: i) *CCL3L*low and *CCL4L*high were overrepresented in the HIV-positive subjects compared with HIV-negative subjects, albeit these associations did not achieve statistical significance (**Figure 5a,b**), and ii) the distribution of *CCL3L-CCL4L* genotypic combinations were significantly different in HIV-infected versus HIV-uninfected subjects ($p = 0.008$, **Figure 5c**). Specifically, the combination of *CCL3L*^{high}-*CCL4L*^{low} was not observed in the HIV-infected subjects while the combination of *CCL3L*^{low}-*CCL4L*^{high} was more prevalent in HIV-infected than HIV-uninfected subjects (**Figure 5c**).

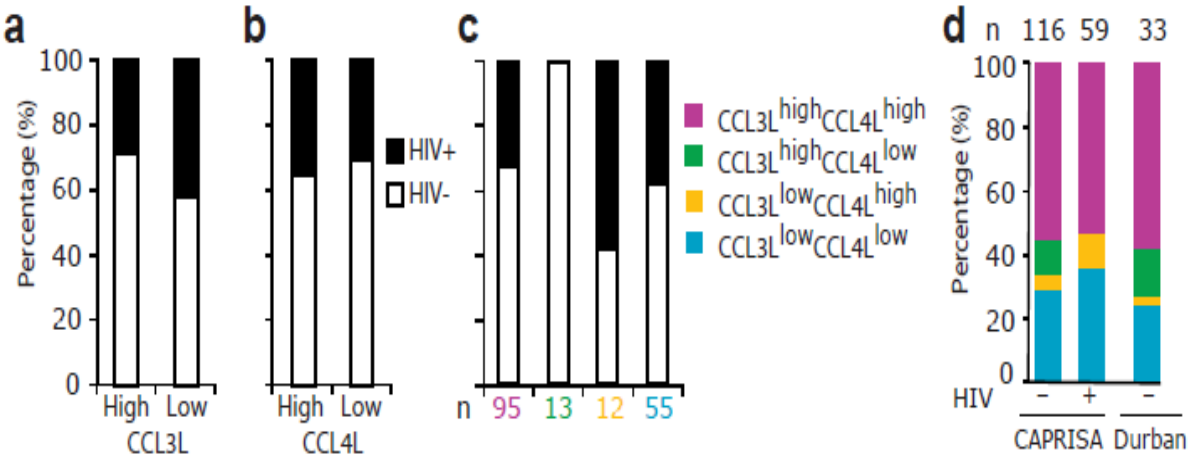


Figure 5: Copy number of both *CCL3L* and *CCL4L* genes together represented as high and low relative to the average race specific copy number i.e. 5 for *CCL3L* and 4 for *CCL4L* using HIV status as a comparison. (**a**) *CCL3L* High/low comparison, black bars represent HIV infected, (**b**) *CCL4L* High/Low comparison. (**c**) Examining the *CCL3L-CCL4L* genes together grouped according to their High/Low status. (**d**) Comparison of the CAPRISA HIV infected and uninfected subjects to a cohort of uninfected low HIV risk control samples from Durban.

This differential distribution of *CCL3L-CCL4L* genotypic combinations in the study population could be interpreted in two ways. First, $CCL3L^{high}$-$CCL4L^{low}$ is selectively enriched in the high-risk HIV-negative group or alternatively, $CCL3L^{low}$-$CCL4L^{high}$ is selectively over-represented in the HIV-positive study subjects. To distinguish between these two possibilities, we compared the prevalence of the *CCL3L-CCL4L* genotypic groups present in the CAPRISA study population with those in a group of low-risk HIV-negative subjects from Durban, South Africa. These analyses demonstrated that the distribution pattern of *CCL3L-CCL4L* genotypic groups in the high-risk HIV-negative subjects was very similar to those in the low-risk HIV-negative subjects (**Figure 5d**). These findings demonstrated that the HIV-positive subjects are over- and under-represented for the $CCL3L^{low}$-$CCL4L^{high}$ and $CCL3L^{high}$-$CCL4L^{low}$ genotypic groups, respectively (**Figure 5d**).

Individual *CCL3L-CCL4L* genes and HIV risk

Quantification of CN of *CCL3L* and *CCL4L* captures the sum of its gene components, raising the possibility that only one of the gene components, or the combination of some of these *CCL3L* (*CCL3La* or *CCL3Lb*) or *CCL4L* (*CCL4La* or *CCL4Lb*) genes associate with HIV risk. However, Table 4 shows that a high gene dose of both *CCL3La* and *CCL3Lb* associated with a lower risk of acquiring HIV, whereas a high CN of both *CCL4La* and *CCL4b* associated with a higher risk of acquiring HIV (**Table 4**). These results indicated that the direction of the associations for each of the individual components of *CCL3L*- and *CCL4L*-related genes were similar and consistent.

Table 4: Association of *CCL3La, CCL3Lb, CCL4La* and *CCL4Lb* copy numbers with the risk of HIV acquisition

Model	Covariate	OR	95% CI	P
1	*CCL3La*	0.16	0.07 – 0.37	<0.001
	CCL3Lb	0.64	0.40 – 1.01	0.058
	CCL4La	7.48	3.14 – 17.81	<0.001
	CCL4Lb	10.30	3.89 – 27.27	<0.001
2	*CCL3La*	0.08	0.02 – 0.28	<0.001
	CCL3Lb	0.58	0.32 – 1.03	0.064
	CCL4La	11.40	3.65 – 35.62	<0.001
	CCL4Lb	22.14	5.29 – 92.77	<0.001
	Age	0.90	0.86 – 0.95	<0.001
	Contraceptive use	0.27	0.11 – 0.70	0.007

However, given the distinct distribution pattern (**Figure 4c-e**) and complex correlation structure of these genes (**Figure 4e**), we evaluated the following two possibilities. First, because HIV-negative subjects classified as *CCL3L*high are overrepresented with subjects who are also classified as *CCL4L*low (**Figure 5c**), subjects classified as *CCL4L*low may be enriched for a specific *CCL3L* component (*CCL3L*a or *CCL3L*b). Consistent with this, although the CN of *CCL3L*b increased with increasing copy numbers of *CCL3L* and *CCL4L*, only 10% and ~25% of persons with 2 and 3 copies of *CCL3La* also possessed *CCL3Lb*, respectively, whereas 40% and 60% of subjects with 2 and 3 copies of *CCL4L* also possessed *CCL3L*b (**Figure 6a**). The influence of *CCL3Lb* on HIV risk was highlighted by the observation that HIV-positive subjects with a low copy number of *CCL3L* (<5) or *CCL4L* (<4) had a low proportion of subjects who also possessed *CCL3L*b compared to HIV-negative subjects (**Figure 6b,c**). Hence, *CCL3L*high-*CCL4L*low signifies subjects who are enriched for both *CCL3L*a and *CCL3L*b gene copies, and a low copy number for each of these genes associates with an increased risk of acquiring HIV in the study population.

A similar set of analyses was run on the *CCL4L*b gene as the *CCL3L*b gene figure 28d,e,f. The data for the *CCL4L*b yielded no significant difference between the *CCL3L* and *CCL4L* genes (**Figure 6d**) as compared to the *CCL3L*b data. Furthermore no significant findings were noticed when the HIV negative and HIV positive probability of possessing *CCL4L*b (**Figure 6e,f**).

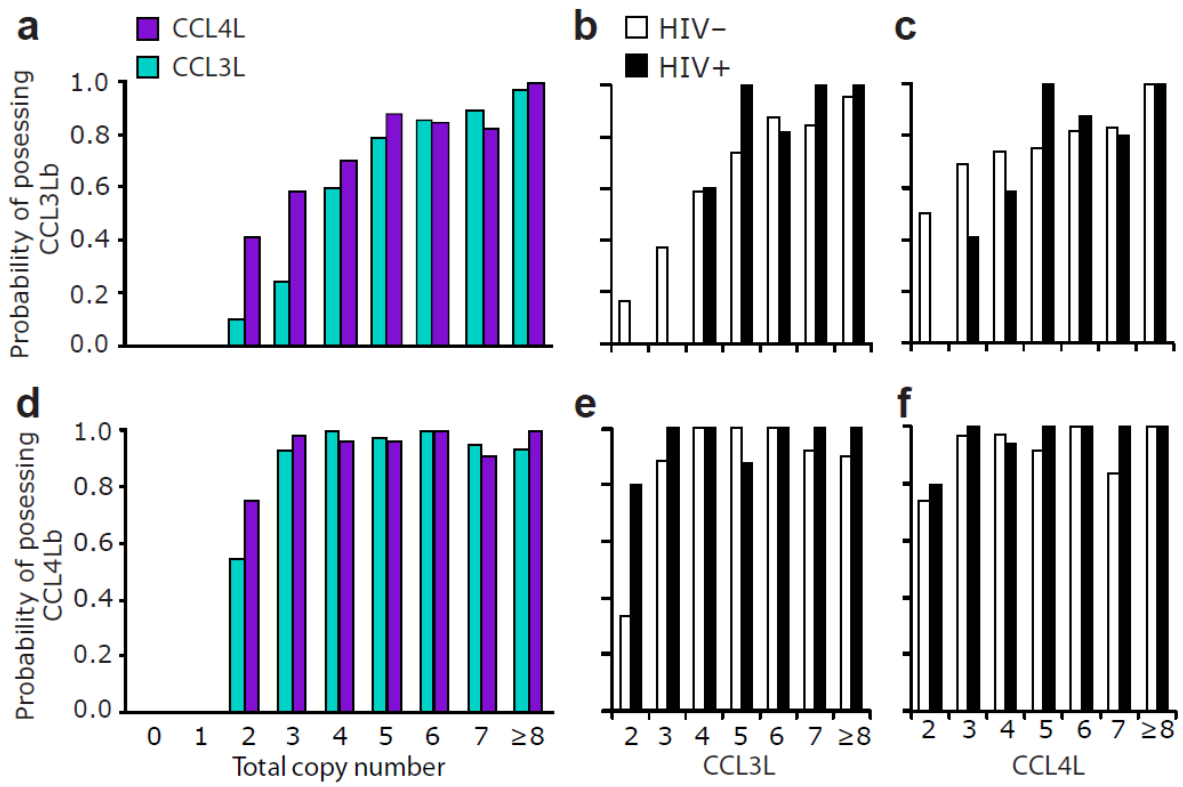


Figure 6: Proportion of subjects possessing either *CCL3L*b or *CCL4L*b relative to HIV status and total GCN. (**a**) Probability of possessing *CCL3L*b gene in relation to the total *CCL3L*and *CCL4L* genes, after stratifying by HIV serostatus a further comparison was performed for the *CCL3L* (**b**) and *CCL4L* (**c**). A similar set of analyses was performed on the proportion of subjects possessing *CCL4L*b gene (**d, e, f**).

Taken together, in the CAPRISA cohort, among subjects classified as $CCL3L^{high}$-$CCL4L^{low}$ those who remained HIV-negative were relatively more enriched with *CCL3La* and *CCL3Lb* gene copies than HIV-positive subjects whereas among subjects classified as $CCL3L^{low}$-$CCL4L^{high}$, those who were HIV-positive were relatively more under- and over-represented with *CCL3Lb* and *CCL4Lb*, respectively, than HIV-negative subjects. Collectively, the findings presented thus far suggested a model wherein the combinatorial content or balance between chemokine gene components associated with HIV susceptibility.

Chemokine CNV and HIV disease phenotypes

We next determined the associations of chemokine CNV with viral load and $CD4^+$ T cell loss, two highly correlated parameters that are strong predictors of AIDS progression rates (33). On the basis of the aforementioned findings, we sought to test the hypothesis that it was balance between the doses of genes conferring protective (e.g. *CCL3La* and *CCL4La*) versus detrimental (e.g., *CCL4Lb*) effects that impact on these laboratory predictors of AIDS risk. The most protective *CCL3L-CCL4L* genotypic group $CCL3L^{high}$-$CCL4L^{low}$ was not present among the HIV-positive subjects we evaluated, limiting the possibility to assess the effects of the full range of the copy numbers of *CCL3L*- and *CCL4L*-related genes on restriction of viral replication and CD4 T cell loss, and it also precluded our ability to apply the same strategies used to evaluate the association of chemokine CNV with HIV risk towards possible associations with disease progression.

We used principal components analyses to develop a greater understanding of the observed pattern of correlations between the CN of the chemokine genes (**Figure 7a**). By factor analyses, we found that the four chemokine components loaded differentially onto a two-factor solution (**Figure 7a**): *CCL3La* and *CCL4La* loaded heavily onto the first factor (Factor 1), while the second factor (Factor 2) was represented by *CCL4Lb* loading very highly onto it (**Figure 7a**). *CCL3Lb* loaded partially onto the first factor and minimally onto the second factor. The degree of uniqueness of *CCL3La*, *CCL3Lb*, *CCL4La* and *CCL4Lb* was 0.13, 0.80, 0.12 and 0.17 indicates that with regards to copy number distribution, *CCL3Lb* the copy number was uniquely different from the other three genes. Thus, factor 1 largely reflects *CCL3La* and *CCL4La*, and factor 2, largely reflects *CCL4Lb*.

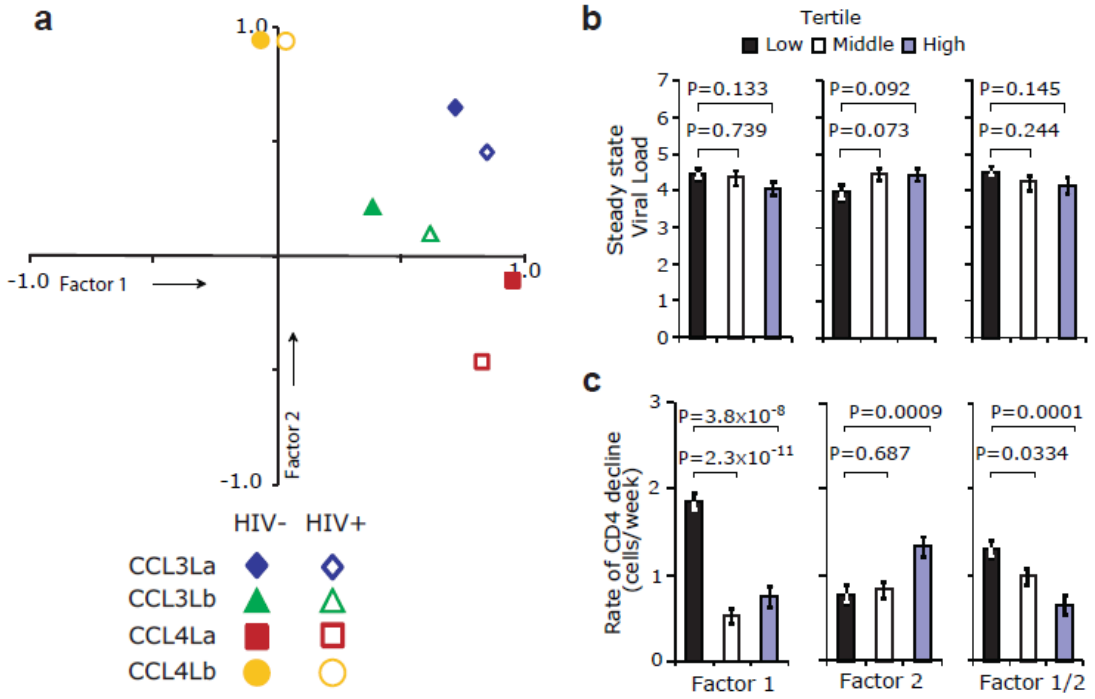


Figure 7: Factor analysis of the components of *CCL3L* and *CCL4L* and the influence of combinatorial chemokine gene content on steady-state viral load and rate of CD4 cell decline. (**a**) Results of principal components analyses. Two orthogonal factors (represented by abscissa and ordinate) were retained based on an eigenvalue >1. The plot represents the varimax-rotated loadings for each of the components on the two factors. Considering these loading patterns, the first factor correlated strongly with *CCL3La* and *CCL4La*, while the second factor correlated strongly with *CCL4Lb*. Solid shapes represent HIV negative subjects while the hollow shapes represent HIV infected subjects. (**b**) Association of the factor scores with the steady-state viral load. Each study subject was scored based on the possession of the copy numbers of all four components (*CCL3La*, *CCL3Lb*, *CCL4La*, and *CCL4Lb*), and their respective loadings shown in panel b. The graph on the left, middle, and right show the associations for factor 1, factor 2 and ratio of the scores for factors 1 and 2 with steady-state viral load, respectively. (**c**) Association for the tertiles of factor 1 and 2, and the ratio of factor 1 and 2 (Factor 1/2) with rates of $CD4^+$ T cell decline. The vertical bars indicate the point estimates while the error bars indicate the 95% confidence intervals rates of $CD4^+$ T cell decline.

To assess whether biological interpretations made from the correlation patterns and that the principal components analyses were justifiable, we conducted the following analyses. Using the results from principal components analyses, we generated standardized scores for both the

factors for each individual. To test the hypothesis that it is the balance between the gene content of different chemokine genes that influences HIV disease, we determined the phenotypic effects on HIV disease associated with the tertiles of the ratio of factor 1 and factor 2 scores (Factor 1/2). Consistent with this, higher estimates of the Factor 1/2 score ratio, which reflects a relative enrichment for Factor 1 and/or depletion of Factor 2 scores were strongly correlated with a decrease in the steady-state viral load (P=0.0056; **Figure 7b**, right).

Concordant associations were detected when we determined the association of the tertiles of the factor score ratios with the rate of decline of $CD4^+$ T cell counts. A low tertile of Factor 1 score was associated with the fastest rates of $CD4^+$ T cell declines (**Figure 7c**, left). Although an association for factor 2 scores was not observed (**Figure 7c**, middle), the lowest tertile of Factor 1/2 score ratio was associated with the fastest rates of $CD4^+$ T cell declines compared to the other two tertiles (**Figure 7c**, right).

DISCUSSION

Evaluation of a well-characterized cohort of South African women that included commercial sex workers demonstrated that it is the combinatorial content of *CCL3L*- and *CCL4L*-related genes rather than the gene dose of any one of these genes that affects both HIV susceptibility and disease progression. Specifically, our results point to an important role for the balance between the copy number of protective (*CCL3La* and *CCL3L*b) versus detrimental (e.g.,

CCL4Lb) chemokine genes in influencing HIV susceptibility, viral replication and CD4+ T cell loss. Thus, our results highlight that the CNV of chemokine genes are an excellent model system to (i) evaluate the extent of genomic architectural complexity in the form of smaller CNV within a segmental duplication or CNVR, (ii) gain further insights into the evolutionary history of CNV of immune response genes, (iii) investigate the impact of this genomic complexity on susceptibility to human diseases especially infectious, autoimmune or inflammatory diseases, and (iv) advance the concept that failure to account for the combinatorial genomic content of distinct CNVs, including their population-specific distributions, may both confound association studies and pose as a challenge for the full characterization of human genome sequence.

The extensive combinatorial content of chemokine genes is unprecedented and is a consequence of complex CNV of *CCL3L* and *CCL4L* genes that can be broadly categorized into those that encode classical secreted chemokines, i.e., *CCL3La* (*CCL3L1* and *CCL3L2*) and *CCL4La* (*CCL4L*2), versus non-classical chemokines, i.e., *CCL3Lb* (*CCL3L2*) and *CCL4Lb* (*CCL4L1*); *CCL3L* and *CCL3L3* are classified as *CCL3La* because they both are predicted to encode identical proteins. Among the 948 unrelated subjects represented in the HGDP-CEPH panel, these chemokine gene components generate >150 unique CNV combinations. However, these combinations are distributed non-randomly with continent-of-origin explaining a large proportion of this variability. There is a greater variability in African and the least variability in European is reflected by the observation that ~65% and ~20% of the >150 chemokine gene combinations were present in African and European populations, respectively.

Remarkably, evaluation of the top five most common chemokine gene combinations in Europeans and Africans according to the dose of *CCL3La*, *CCL3Lb*, *CCL4La*, and *CCL4Lb* genes (e.g., [2, 0, 2, 0]) demonstrates a contrasting pattern. Approximately, 50% of Europeans have a diploid genotype for the classical chemokine genes with two copies of *CCL3La* and *CCL4La*, but lack the non-classical chemokine genes *CCL3Lb* and *CCL4Lb* [2, 0, 2, 0]. Remarkably, the next most common chemokine genotype representing ~15% of Europeans is the hemizygous state of *CCL3La* and *CCL4La* genes [1, 0, 1, 0]. By contrast, none of the chemokine gene combinations represented >5% of the Africans, and inspection of the individual sub-populations collectively classified as Africans in the HGDP-CEPH panel revealed that nearly each person within a sub-population had a unique chemokine CNV genotype. The diploid [2, 0, 2, 0] or hemizygous [1, 0, 1, 0] genetic state was rare among Africans, and the top four chemokine genome combinations in this population have nearly the same frequency and were [3, 0, 3, 0], [4, 0, 2, 2], [4, 1, 2, 2] and [4,2,2,2]. Thus, if we assume the diploid chemokine gene state of [2, 0, 2, 0] as an equipoise, CNV in Europeans and Africans are in opposing directions from this state, favoring a reduction of the non classical and an increase of the classical chemokines.

From an evolutionary perspective, whether the greater content of chemokine genes in Africa arose as a means to combat ancestral infectious diseases remains unknown but plausible for the following reasons. Infectious diseases are thought to have been a major driver of genetic diversity, and in sub-Saharan Africa the most common genetic diseases have evolved as a consequence of selection to reduce susceptibility to infection (e.g. hemoglobinopathies). Members of the chemokine system play a key role in mediating immune responses against infectious diseases, including viruses (29, 37). Notably, chemokine gene duplications are observed in humans and non-human primates and in both species an association of

chemokine gene copy number with lentiviral infection has been evident (12, 18, 20). Viruses, in turn, have evolved mechanisms to mimic, inhibit or usurp the immune effects of the chemokine system (18, 36), including hijacking of the chemokine system, whereby certain viruses such as HIV utilize chemokine receptors for their cell entry (5).

Thus, duplication of potent immune response chemokine genes may be reflective of eons of co-evolution of primate and viral genomes (18). This is underscored by what appears to be two evolutionary trade-offs: first, increasing dosages of genes that transcribe classical secreted chemokines (*CCL3La* and *CCL3Lb*) are offset by increasing dosages of genes that transcribe mRNA isoforms that are not classical chemokines, and second, the observed inverse correlation between the copy number of *CCL4L*a (secreted chemokine) and *CCL4L*b (non-secreted chemokine) suggests that in some populations there might be an evolutionary pressure to maintain a balance between classical and non-classical *CCL4L* genes. Because high gene doses of functional chemokines would have imparted a persistent pro-inflammatory state, it is conceivable that the presence of these non-classical chemokine genes may have evolved to counter-balance the pro-inflammatory effects associated with possession of a high gene content of secreted classical chemokines. Activated CD8+ T cells from subjects who are null for the gene encoding *CCL4L*a (the classical secreted *CCL4L* variant), but possess *CCL4Lb* gene (encoding a non-classical chemokine) express lower amounts CCL4, a CCR5-specific chemokine (9). By analogy, in preliminary studies, we have found that possession of *CCL3Lb* is associated with reduced expression of the classical chemokine *CCL3L* following stimulation of peripheral blood mononuclear cells with the microbial product lipopolysaccharide (unpublished data).

As a result of the higher chemokine gene content in African populations, the cohort of HIV-uninfected and –infected South African women afforded the unique opportunity to determine whether the combinatorial content of chemokine genes versus the dose of a single chemokine gene (e.g. *CCL3L*) influenced HIV susceptibility. We identify concordant associations for duplicated chemokine genes across three phenotypic endpoints: HIV susceptibility, viral load setpoint, and rate of $CD4^+$ T cell decline. Specifically, we find that these phenotypic endpoints are influenced not only by the quantitative content of each chemokine gene *per se*, but by the combinatorial content or balance of these genes.

These findings suggest that restricting CNV-phenotype analysis solely to the *CCL3L*-segmental duplication alone, although providing an aggregate estimate of all *CCL3L*-related genes, may in some instances provide an incomplete assessment of the phenotypic impact of chemokine CNVs for three main reasons. First, our findings demonstrate that there is a complex population-specific correlation pattern between the CNV of different chemokine genes and hence as reflected in this study, failure to account for the effects of both *CCL3L* and *CCL4L* genes obscures the associations of this locus with HIV-AIDS susceptibility. Second, a role for the CNV of *CCL4L* independent of *CCL3L* in HIV-AIDS has been identified in cohorts of European ancestry (9, 46): in a cohort of adults from Spain demonstrated that subjects lacking *CCL4La* but possessing *CCL4Lb* had an increased risk of acquiring HIV infection (9). A separate study, demonstrated that in Ukranian children exposed perinatally to HIV the impact on HIV-AIDS susceptibility of the chemokine gene-rich locus on 17q12 was dependent on the balance between the doses of genes conferring protective (*CCL3L*a and *CCL4L*a) versus detrimental (*CCL4L*b) effects; the frequency of subjects with *CCL3Lb* was too low to determine an independent effect of this gene (46). Concordantly, we find that a similar balance also impacts on HIV susceptibility and

laboratory parameters predictive of AIDS risk in the CAPRISA cohort. Third, the region on 17q12 that has undergone segmental duplication is large (~100 kb) and is thought to be replete with many breakpoints that can lead to duplicated segments that vary in chemokine gene content amongst individuals (7, 35, 42, 47).

On the basis of the known HIV-suppressive effects of *CCL3L*/3 (*CCL3La*) and our preliminary results suggesting that *CCL3Lb* gene product may impart anti-inflammatory properties (data not shown), we propose a model wherein *CCL3L*-related genes may influence HIV susceptibility through a dual mechanism: via the anti-HIV effects associated with *CCL3La* gene products (*CCL3L* and *CCL3L*3) and second, by dampening inflammation, a key component of HIV pathogenecity, via the *CCL3L*b gene product. Studies are underway to validate this model.

In conjunction with the results of previous studies (reviewed in (20)), including a recent report suggesting that HIV preferentially selects for subjects with a low gene dose of the *CCL3L*-containing segmental duplication, these findings support a strong link between chemokine gene CNV and HIV-AIDS susceptibility. Notwithstanding this, the associations of the *CCL3L*-containing segmental duplication with HIV-AIDS susceptibility have been contested by two groups (6, 49), and Field et al asserted that the methods we employed in this study to quantify *CCL3L*-related genes were inaccurate (14). He et al in their riposte to these contestations outlined methodological, genetic and epidemiologic reasons why inter-cohort differences in the associations of *CCL3L* CNV may have been observed (20). Highlighting the robustness and precision of the assays we used to quantify CN, we reaffirm that in each of the populations we studied, the sum of the CN estimates for the ‘a’ and ‘b’ components of

CCL3L and *CCL4L* were very highly correlated with the total CN for either *CCL3L*-related and *CCL4L*-related genes and furthermore the CN estimates clustered near integer values. He et al also noted that among the genetic factors that can obscure or confound the associations of chemokine CNV with HIV-AIDS is to conceptualize the *CCL3L*-containing segmental duplication as an invariant single CNV unit (20). They hypothesized that depending on the study population evaluated this may obfuscate both the genomic architectural complexity of the chemokine gene-rich 17q12 CNVR and the relative contributions of the doses of the individual chemokine genes on HIV risk and HIV disease susceptibility (20). The results of the present study affirm this hypothesis.

One imporatant question may be asked as to how might the complex patterns of chemokine gene content impact upon evaluation of vaccine efficacy. High-risk individuals such as those we studied herein represent good candidates for a vaccine trial. We found that the women who seroconverted during prospective follow-up were more likely to possess a low dose of *CCL3L* genes at time of enrollment into the study. Analyses of intravenous drug users, another high-risk group, suggested that HIV may preferentially select for subjects with a low dose of the *CCL3L*-containing segmental duplication (21). Hence, hypothetically enhanced susceptibility associated with a low copy number may overshadow or nullify the benefits of a partially protective vaccine. Furthermore, because communicability is related to the viral load of the transmitting partner (3, 8, 19, 41, 51), chemokine gene content, by influencing the degree of viremia and immune deficiency as reflected by CD4 cell declines, may also affect sexual transmission rate by vaccinated subjects who become infected with HIV subsequent to vaccination. Thus, as chemokine gene balance impacts on four parameters relevant to the evaluation of vaccine efficacy (acquisition, viral load, $CD4^+$ T cell decline, and transmission rate) (15, 48), accounting for combinatorial chemokine gene content should improve the

ability to assess true vaccine efficacy. This premise is also supported by our modeling studies, wherein we find that even low levels of unbalanced trial arm allocation of participants with protective or detrimental *CCL3L-CCR5* genotypes may confound the outcome of a vaccine trial (26). Further, the immunological phenotypes linked to $CCL3L^{high}$-$CCL4L^{low}$ and $CCL3L^{low}$-$CCL4L^{high}$ the most protective and detrimental genotypes against HIV acquisition detected in this study, may provide important information regarding the immune correlates of protection.

The discovery of the complex combinatorial inheritance pattern of chemokine genes has broad relevance to gaining a better appreciation of the phenotypic impact of CNVs on a wide variety of diseases, which is anticipated to be extensive (10, 28, 31, 42, 44, 52). We suggest that given the population-specific nature of the chemokine CNVs, and possibly other CNVs, this appreciation will require careful attention to the epidemiologic features of the study as well as fine-scale analyses of the phenotypic impacts of distinct CNVs, rather than assessment of any single component of a genomic region that has undergone segmental duplications. A similar premise was also advanced by the authors of a study that used high-resolution array-based comparative genomic hybridization to interrogate known CNV regions (40). Hence, despite the considerable excitement over the potential functional significance of CNVs, genotype-phenotype studies directed at uncovering the disease phenotypes of CNVs may need to account for this largely uncharacterized complexity, which may be a difficult undertaking using whole genome-sequencing approaches solely.

ACKNOWLEDGMENTS

This work was supported by the Veterans Administration (VA) Center on AIDS and HIV infection of the South Texas Veterans Health Care System, and a MERIT (R37046326) and other awards (AI043279 and MH069270) from the National Institutes of Health (NIH) to S.K.A. S.K.A. is a recipient of the Elizabeth Glaser Scientist Award and the Burroughs Wellcome Clinical Scientist Award in Translational Research. The CAPRISA 002 study was supported by the National Institute of Allergy and Infectious Diseases (NIAID), (NIH), U.S. Department of Health and Human Services (grant U19 AI 51794). S.M. and S.K.A. are also supported by a Merit Review grant awarded by the Department of Veterans Affairs. V.R. is a recipient of a CAPRISA training fellowship and a Fogarty AITRP fellowship (TWO-0023). T.N. is supported by the Hasso Plattner Foundation and holds the South African DST/NRF Research Chair in Systems Biology of HIV/AIDS. We thank the study participants, CAPRISA clinical and laboratory staff for providing specimens, and special thanks to Dr. Carolyn Williamson. We thank D. Hepp, A. Carrillo and K. Luk for technical help and U. Aluyen for excellent graphic work.

References

1. **Ahuja, S. K., H. Kulkarni, G. Catano, B. K. Agan, J. F. Camargo, W. He, R. J. O'Connell, V. C. Marconi, J. Delmar, J. Eron, R. A. Clark, S. Frost, J. Martin, S. S. Ahuja, S. G. Deeks, S. Little, D. Richman, F. M. Hecht, and M. J. Dolan.** 2008. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. Nat Med **14:**413-20.
2. **Alkan, C., J. M. Kidd, T. Marques-Bonet, G. Aksay, F. Antonacci, F. Hormozdiari, J. O. Kitzman, C. Baker, M. Malig, O. Mutlu, S. C. Sahinalp, R. A. Gibbs, and E. E. Eichler.** 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet **41:**1061-7.
3. **Anderson, R., and M. Hanson.** 2005. Potential public health impact of imperfect HIV type 1 vaccines. J Infect Dis **191 Suppl 1:**S85-96.
4. **Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler.** 2002. Recent segmental duplications in the human genome. Science **297:**1003-7.
5. **Berger, E. A., P. M. Murphy, and J. M. Farber.** 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol **17:**657-700.
6. **Bhattacharya, T., J. Stanton, E. Y. Kim, K. J. Kunstman, J. P. Phair, L. P. Jacobson, and S. M. Wolinsky.** 2009. CCL3L1 and HIV/AIDS susceptibility. Nat Med **15:**1112-5.
7. **Cardone, M. F., Z. Jiang, P. D'Addabbo, N. Archidiacono, M. Rocchi, E. E. Eichler, and M. Ventura.** 2008. Hominoid chromosomal rearrangements on 17q map to complex regions of segmental duplication. Genome Biol **9:**R28.
8. **Cohen, M. S., N. Hellmann, J. A. Levy, K. DeCock, and J. Lange.** 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. J Clin Invest **118:**1244-54.
9. **Colobran, R., P. Adreani, Y. Ashhab, A. Llano, J. A. Este, O. Dominguez, R. Pujol-Borrell, and M. Juan.** 2005. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. J Immunol **174:**5655-64.
10. **Conrad, D. F., and M. E. Hurles.** 2007. The population genetics of structural variation. Nat Genet **39:**S30-6.
11. **Cooper, G. M., D. A. Nickerson, and E. E. Eichler.** 2007. Mutational and selective effects on copy-number variants in the human genome. Nat Genet **39:**S22-9.
12. **Degenhardt, J. D., P. de Candia, A. Chabot, S. Schwartz, L. Henderson, B. Ling, M. Hunter, Z. Jiang, R. E. Palermo, M. Katze, E. E. Eichler, M. Ventura, J. Rogers, P. Marx, Y. Gilad, and C. D. Bustamante.** 2009. Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). PLoS Genet **5:**e1000346.
13. **Dolan, M. J., H. Kulkarni, J. F. Camargo, W. He, A. Smith, J. M. Anaya, T. Miura, F. M. Hecht, M. Mamtani, F. Pereyra, V. Marconi, A. Mangano, L. Sen, R. Bologna, R. A. Clark, S. A. Anderson, J. Delmar, R. J. O'Connell, A. Lloyd, J. Martin, S. S. Ahuja, B. K. Agan, B. D. Walker, S. G. Deeks, and S. K. Ahuja.** 2007. CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. Nat Immunol **8:**1324-36.
14. **Field, S. F., J. M. Howson, L. M. Maier, S. Walker, N. M. Walker, D. J. Smyth, J. A. Armour, D. G. Clayton, and J. A. Todd.** 2009. Experimental aspects of copy number variant assays at CCL3L1. Nat Med **15:**1115-7.
15. **Gilbert, P. B., V. G. DeGruttola, M. G. Hudgens, S. G. Self, S. M. Hammer, and L. Corey.** 2003. What constitutes efficacy for a human immunodeficiency virus vaccine that ameliorates viremia: issues involving surrogate end points in phase 3 trials. J Infect Dis **188:**179-93.
16. **Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A.**

Clark, S. A. Anderson, J. O'Connell R, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science **307:**1434-40.

17. **Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A. Clark, S. A. Anderson, R. J. O'Connell, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja.** 2005. The influence of CCL3L1 gene containing segmental duplications of HIV-1/AIDS susceptibility. Science **307:**1434-1440.
18. **Gornalusse, G., S. Mummidi, W. He, G. Silvestri, M. Bamshad, and S. K. Ahuja.** 2009. CCL3L Copy number variation and the co-evolution of primate and viral genomes. PLoS Genet **5:**e1000359.
19. **Gray, R. H., M. J. Wawer, R. Brookmeyer, N. K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. vanCott, and T. C. Quinn.** 2001. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. Lancet **357:**1149-53.
20. **He, W., H. Kulkarni, J. Castiblanco, C. Shimizu, U. Aluyen, R. Maldonado, A. Carrillo, M. Griffin, A. Lipsitt, L. Beachy, L. Shostakovich-Koretskaya, A. Mangano, L. Sen, R. J. Nibbs, C. T. Tiemessen, H. Bolivar, M. J. Bamshad, R. A. Clark, J. C. Burns, M. J. Dolan, and S. K. Ahuja.** 2009. Reply to: "Experimental aspects of copy number variant assays at CCL3L1". Nat Med **15:**1117-20.
21. **Huik, K., M. Sadam, T. Karki, R. Avi, T. Krispin, P. Paap, K. Ruutel, A. Uuskula, A. Talu, K. Abel-Ollo, and I. Lutsar.** 2010. CCL3L1 copy number is a strong genetic determinant of HIV seropositivity in Caucasian intravenous drug users. J Infect Dis **201:**730-9.
22. **Jakobsson, M., S. W. Scholz, P. Scheet, J. R. Gibbs, J. M. VanLiere, H. C. Fung, Z. A. Szpiech, J. H. Degnan, K. Wang, R. Guerreiro, J. M. Bras, J. C. Schymick, D. G. Hernandez, B. J. Traynor, J. Simon-Sanchez, M. Matarin, A. Britton, J. van de Leemput, I. Rafferty, M. Bucan, H. M. Cann, J. A. Hardy, N. A. Rosenberg, and A. B. Singleton.** 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature **451:**998-1003.
23. **Karim, S. S.** 2007. HIV incidence estimates are key to understanding the changing HIV epidemic in South Africa. S Afr Med J **97:**190.
24. **Kaslow, R. A., T. Dorak, and J. J. Tang.** 2005. Influence of host genetic variation on susceptibility to HIV type 1 infection. J Infect Dis **191 Suppl 1:**S68-77.
25. **Kuhn, L., D. B. Schramm, S. Donninger, S. Meddows-Taylor, A. H. Coovadia, G. G. Sherman, G. E. Gray, and C. T. Tiemessen.** 2007. African infants' CCL3 gene copies influence perinatal HIV transmission in the absence of maternal nevirapine. AIDS **21:**1753-61.
26. **Kulkarni, H., V. C. Marconi, B. K. Agan, C. McArthur, G. Crawford, R. A. Clark, M. J. Dolan, and S. K. Ahuja.** 2008. Role of CCL3L1-CCR5 genotypes in the epidemic spread of HIV-1 and evaluation of vaccine efficacy. PLoS One **3:**e3671.
27. **Lupski, J. R.** 2007. An evolution revolution provides further revelation. Bioessays **29:**1182-4.
28. **Lupski, J. R.** 2007. Genomic rearrangements and sporadic disease. Nat Genet **39:**S43-7.
29. **Mahalingam, S., K. Clark, K. I. Matthaei, and P. S. Foster.** 2001. Antiviral potential of chemokines. Bioessays **23:**428-35.
30. **Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher.** 2009. Finding the missing heritability of complex diseases. Nature **461:**747-53.
31. **McCarroll, S. A., and D. M. Altshuler.** 2007. Copy-number variation and association studies of human disease. Nat Genet **39:**S37-42.

32. **Meddows-Taylor, S., S. L. Donninger, M. Paximadis, D. B. Schramm, F. S. Anthony, G. E. Gray, L. Kuhn, and C. T. Tiemessen.** 2006. Reduced ability of newborns to produce CCL3 is associated with increased susceptibility to perinatal human immunodeficiency virus 1 transmission. J Gen Virol **87:**2055-65.
33. **Mellors, J. W., A. Munoz, J. V. Giorgi, J. B. Margolick, C. J. Tassoni, P. Gupta, L. A. Kingsley, J. A. Todd, A. J. Saah, R. Detels, J. P. Phair, and C. R. Rinaldo, Jr.** 1997. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med **126:**946-54.
34. **Menten, P., A. Wuyts, and J. Van Damme.** 2002. Macrophage inflammatory protein-1. Cytokine Growth Factor Rev **13:**455-81.
35. **Modi, W. S.** 2004. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. Genomics **83:**735-8.
36. **Murphy, P. M.** 2001. Viral exploitation and subversion of the immune system through chemokine mimicry. Nat Immunol **2:**116-22.
37. **Murphy, P. M., M. Baggiolini, I. F. Charo, C. A. Hebert, R. Horuk, K. Matsushima, L. H. Miller, J. J. Oppenheim, and C. A. Power.** 2000. International union of pharmacology. XXII. Nomenclature for chemokine receptors. Pharmacol Rev **52:**145-76.
38. **Nakajima, T., H. Ohtani, T. Naruse, H. Shibata, J. I. Mimaya, H. Terunuma, and A. Kimura.** 2007. Copy number variations of CCL3L1 and long-term prognosis of HIV-1 infection in asymptomatic HIV-infected Japanese with hemophilia. Immunogenetics **59:**793-8.
39. **Nibbs, R. J., J. Yang, N. R. Landau, J. H. Mao, and G. J. Graham.** 1999. LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity. J Biol Chem **274:**17478-83.
40. **Perry, G. H., A. Ben-Dor, A. Tsalenko, N. Sampas, L. Rodriguez-Revenga, C. W. Tran, A. Scheffer, I. Steinfeld, P. Tsang, N. A. Yamada, H. S. Park, J. I. Kim, J. S. Seo, Z. Yakhini, S. Laderman, L. Bruhn, and C. Lee.** 2008. The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet **82:**685-95.
41. **Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, and R. H. Gray.** 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med **342:**921-9.
42. **Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles.** 2006. Global variation in copy number in the human genome. Nature **444:**444-54.
43. **Sadam, M., T. Karki, K. Huik, R. Avi, and K. Ruutel.** 2008. Presented at the 15th Conference on Retroviruses and Opportunistic Infections.
44. **Sebat, J.** 2007. Major changes in our DNA lead to major changes in our thinking. Nat Genet **39:**S3-5.
45. **Shalekoff, S., S. Meddows-Taylor, D. B. Schramm, S. L. Donninger, G. E. Gray, G. G. Sherman, A. H. Coovadia, L. Kuhn, and C. T. Tiemessen.** 2008. Host CCL3L1 gene copy number in relation to HIV-1-specific CD4+ and CD8+ T-cell responses and viral load in South African women. J Acquir Immune Defic Syndr **48:**245-54.
46. **Shostakovich-Koretskaya, L., G. Catano, Z. A. Chykarenko, W. He, G. Gornalusse, S. Mummidi, R. Sanchez, M. J. Dolan, S. S. Ahuja, R. A. Clark, H. Kulkarni, and S. K. Ahuja.** 2009. Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. AIDS **23:**679-88.
47. **Townson, J. R., L. F. Barcellos, and R. J. Nibbs.** 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. Eur J Immunol **32:**3016-26.

48. **Trials, A. V.** 2001, posting date. Considerations for Phase III Trial Design and Endpoints N http://www.niaid.nih.gov/vrc/pdf/p3trialsend.pdf. [Online.]
49. **Urban, T. J., A. C. Weintrob, J. Fellay, S. Colombo, K. V. Shianna, C. Gumbs, M. Rotger, K. Pelak, K. K. Dang, R. Detels, J. J. Martinson, S. J. O'Brien, N. L. Letvin, A. J. McMichael, B. F. Haynes, M. Carrington, A. Telenti, N. L. Michael, and D. B. Goldstein.** 2009. CCL3L1 and HIV/AIDS susceptibility. Nat Med **15:**1110-2.
50. **van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. Abdool Karim, A. Grobler, N. Barnabas, I. Iriogbe, and S. S. Abdool Karim.** 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One **3:**e1954.
51. **Wawer, M. J., R. H. Gray, N. K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, F. Nalugoda, F. Wabwire-Mangen, M. P. Meehan, and T. C. Quinn.** 2005. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis **191:**1403-9.
52. **Zhang, F., W. Gu, M. E. Hurles, and J. R. Lupski.** 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet **10:**451-81.

CHAPTER 6 - PAPER THREE

Determining the impact of epigenetics on HIV using a cohort of South African Black Women

Submitted to international journal for publication

Gornalusse, G., A. A. Gaitan, S. Mummidi, A. M. Murillo, **V. Ramsuran**, et al. 2011. DNA methylation regulates T-cell expression levels of the major HIV co-receptor CCR5 and associates with disease phenotypes**. Nature Immunology. Submitted for publication.**

Determining the impact of epigenetics on HIV using a cohort of South African Black Women

Veron Ramsuran[1,2,3,4]*, German Gornalusse[3,4]*, Hemant Kulkarni[3,4], Weijing He[3,4], Koleka Mlisana[2], Lise Werner[2], Robert A. Clark[3,4], Salim S. Abdool Karim[2], Sunil K. Ahuja[3,4†], Thumbi Ndung'u[1,2†]

[1]HIV Pathogenesis Programme (HPP), Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, 4013, South Africa

[2]Centre for the AIDS Program of Research in South Africa (CAPRISA), Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, 4013, South Africa

[3]Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, San Antonio, TX 78229

[4]Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229

*Equal contributions. †To whom correspondence should be addressed: ahujas@uthscsa.edu (S.K.A.) or ndungu@ukzn.ac.za (T.N.)

Author Contributions

Conception and design of the study	- VR, GG, SKA, TN
Performed Lab work	- VR, GG
Analysis and interpretation of data	- VR, GG, HK, SKA
Collection and assembly of data	- VR, WH
Drafting of the article	- VR, GG, SKA
Statistical expertise	- VR, GG, HK, LW
Critical revision and approval	- VR, GG, KM, RAC, SSAK, SKA, TN

ABSTRACT

Studies have shown that genetic factors account for the heterogeneity in the risk of acquiring HIV-1 and in the rate of disease progression among HIV infected persons. Epigenetic factors have been examined in great detail in the cancer field but studies in the HIV field have been lacking. Epigenetic factors are thought to account for some of variation noticed in HIV. We initiated studies to determine the association between the DNA methylation status in the regulatory regions of CCR5 in HIV-uninfected and HIV–infected subjects. The study group consisted of Black South African women, including commercial sex workers. The assays used were; firstly, DNA modification using bisulfite conversion; secondly, a PCR assay used to amplifly the CCR5 promoter and intron regions; Thirdly, pyrosequencing assays were performed to assess DNA methylation at CpG dinucleotides. Statistical analyses were conducted using Stata 7.0 (Stata Corp, College Station, TX). We demonstrate that a strong negative correlation between the methylation levels within the CCR5 promoter and intron regions, and with the CCR5 expression levels, and further show that the DNA methylation status is lower in HIV+ subjects compared to HIV- subjects. These studies reinforce the idea that epigenetic factors are likely to influence HIV-AIDS pathogenesis.

Introduction

Epigenetics, literally meaning above the genome, is a relatively new field of epigenetics and epigenetic epidemiology (12) has recently gained much attention with the International Human Epigenome Project (HEP) being launched (1, 42). However, despite the attention that has surrounded this field there have been limited studies examining the direct association between host epigenetics and infectious diseases. Greater attention has been placed on two major disorders: cancer and mental retardation syndromes (reviewed in (12)).

DNA methylation is a useful marker for assessing the epigenetic state of a locus, since this state is retained during DNA isolation, and can be measured by PCR-based techniques. This can be described as a one-dimensional measurement of a multidimensional state (20, 52) such as combinations of DNA methylation and transcription factor binding (49), gene expression and DNA methylation (35), DNA methylation and histone modifications (37, 49).

Low levels of DNA methylation are associated with active chromatin and higher gene activity whereas high levels of methylation are associated with condensed chromatin and gene silencing, i.e., CpG methylation is mainly a "repressive signature". DNA methylation patterns vary between cell types and individuals (13, 40), potentially stemming from environmental exposure (51), stochastic methylation events (14) or heritability(25), and influence development of disease (e.g. cancer, autoimmune diseases), and also time of disease onset (2, 5, 6, 9, 10, 12, 15-17, 21, 24, 27, 30, 36, 38, 39, 43-47, 50).

Some pioneering studies have demonstrated that in the HIV-1 field , DNA methylation can be employed to favor viral latency or immune evasion. For example, it has been shown that HIV-1 infection may directly modify the host epigenome by increasing the DNMT1 expression levels (probably by Tat-mediated Erk inactivation) which in turn resulted in *de novo* increase of global methylation content (11) as well as an increase in CpG methylation of certain genes such as interferon gamma (34).

Although still controversial, it appears that the HIV life cycle is also regulated by DNA methylation (19, 26, 29). This epigenetic point elicits two effects on HIV-1 biology: it represses transcription and it maintains latency. In fact, many endogenous retroviral elements (e.g. H, K and W families (28, 31)) as well as infective retroviruses (e.g. Rous sarcoma virus (22)) exploit 5' LTR CpG methylation to accrue transcriptional silencing. HIV integration sites in activated and resting CD4+ T-cells also show a preferential targeting of the active chromatin compartments (i.e. CpG-island gene-rich domains), as shown by 454 sequencing (3).

Most studies on epigenetic regulation of genes have been conducted either in mouse models or *in vitro* and are circumscribed mostly to genes that are hallmarks of Th_1-Th_2-Th_{17} differentiation/polarization pathways. However, there is a lack of *in vivo* studies on epigenetic regulation of HIV-1 host-restriction genetic factors.

Why is there a need to explain HIV-1/AIDS pathogenesis in light of epigenetics? After ~25 years of extensive research, explaining and predicting HIV-associated CD4 cell loss, and the

pathogenesis of HIV disease still is a puzzle. Several lines of evidence from human and non-human primate studies indicate that viral entry-dependent mechanisms cannot account for the full extent of CD4+ T cell loss observed during HIV infection. First, viral load (VL), a parameter of the extent of viral entry and replication is insufficient in explaining fully time to AIDS or rates of CD4+ T cell declines. Dolan *et al.* found that ~9% and ~4% of variability in time to AIDS and CD4+ slope was explained by steady-state VL, respectively (7). Likewise, ~6% and ~3% of variability in time to AIDS and CD4+ slope, respectively, was explained by polymorphisms in *CCR5* and *CCL3L*. Second, during chronic untreated HIV infection, relatively few HIV infected CD4+ cells can be detected, even in subjects with a high VL (23). Third, SIV infection of several non-human primate species does not cause AIDS despite high levels of viral replication (41). Together, these three observations suggest that host factors might have a comparable or even greater impact on disease progression than does the extent of viral replication, such that indirect, viral-entry independent mechanisms (those not attributable to HIV replication) might contribute significantly to HIV-AIDS pathogenesis.

Epigenetics has the ability to alter gene expression, as demonstrated in other disease models. A change in the expression of the CCR5 gene is critical for HIV susceptibility/protection as demonstrated by many previous studies. Measuring DNA methylation could help explain the changes in the CCR5 gene expression, since methylated DNA does not allow the transcription machinery to bind, therefore resulting in lack of gene expression. Furthermore, using the CAPRISA cohort we are provided with the unique opportunity of measuring the DNA methylation levels in HIV negative individuals and following them over time to the point of HIV seroconversion, and thereafter followed up regularly during the course of HIV

disease progression. To our knowledge there have been no studies exploring the relationship between host epigenetics and HIV/AIDS.

Materials and Methods

Sodium Bisulphite Mutagenesis.

Genomic DNA was bisulphite-modified and PCR amplified and individual clones were sequenced according to the original procedure described by Clark *et al.* (18). Concentration of NaHSO3 was increased to 5 M and time reduced to 5 hrs to increase kinetics of reaction and decrease template degradation. Alternatively, bisulphite modification was performed using a commercially available kit (EZ DNA methylation Direct, Zymo Research), following manufacturer's instructions.

PCR Amplification of *CCR5* promoter 2, intron 2 and IL-2

For CCR5 Pr2 methylation analysis, a 173bp fragment encompassing four CpG sites, located in CCR5 Pr2 at -776, -768, -702 and -681 positions was amplified. For quantification of CCR5 intron2 methylation, a 296-bp amplicon containing 6 CpGs in intron2: 2169, 2187, 2258, 2311, 2342 and 2365 was designed. Primer sequences are listed in table 4. The PCR

reaction mix consisted of PCR buffer, 3 mM $MgCl_2$, 200µM of each dNTP, 0.2µM of forward primer, 0.2 µM 5' biotinylated reverse primer, 2.5U *Taq* polymerase and 2-4 µL of bisulphite treated DNA. PCR was performed under "hot-start" conditions to reduce non-specific amplification. For amplification of Pr2, PCR cycling conditions were: initial stage of 80°C for 4 minutes and 94°C for 2 minutes, followed by 45 cycles of (94°C for 10 sec, 55°C for 30 sec and 72°C for 30sec) and a last stage of 72°C for 10 min. For amplification of int2 fragment, similar conditions were followed but annealing temperature was 60°C. Successful PCR amplification was confirmed by gel electrophoresis. Five microliters of PCR reaction mixtures were mixed with 1µl of xylene cyanol sample loading buffer and loaded onto a 3% w/v agarose gel.

Pyrosequencing

Pyrosequencing was employed as strategy to quantify methylation levels in the genomic DNA and was performed by a commercial firm (Epigendx, MA). PCR products were then pyrosequenced with the PSQ 96HS system (Biotage AB, Uppsala, Sweden), according to the established method (8), primer sequences available on request.

Results

CCR5 expression is crucial for HIV resistance or susceptibility; therefore the factors that affect the expression of CCR5 are significant. Hence, we examined the CCR5 expression using flow cytometry CAPRISA cohort. We first calculated that number of cells expressing CCR5, thereafter we determined the number of CCR5 molecules on a cell. The CAPRISA cohort replicates previous findings (unpublished data) which show that these two factors were correlated with each other.

Methylation negatively correlates with CCR5 expression

To examine the impact DNA methylation has on CCR5 protein expression we conducted an experiment that measures the percentage methylation per CCR5 methylation site. The results show that there is a strong negative correlation between CCR5 expression and CCR5 DNA methylation sites (Table 1). The correlation matrix displays the Pearson's correlation coefficient and the p-values are shown in parenthesis for each measurement, significant values ($p<0.05$) are in bold. The percentage of cells CCR5 positive among CD4 positive cells shows a significant negative correlation with the CCR5 promoter 2 and IL-2 CpG positions (Table 1). Unactivated cells (HLA-DR-) showed greater significant negative correlations with the methylation positions than the activated cells (HLA-DR+). Both %CCR5+ HLA-DR- CD4+ and Mean CCR5+ HLA-DR- CD4+ showed a significant negative correlation for most of the selected CpG positions (Table 1).

Table 1: Correlation matrix of CCR5 expression and methylation positions from CCR5 intron 2, promoter 2 and IL-2.

	CCR5 Intron 2						CCR5 Promoter 2				IL-2	
	Pos #4	**Pos #5**	**Pos #6**	**Pos #7**	**Pos #8**	**Pos #9**	**Pos 1**	**Pos 2**	**Pos 3**	**Pos 4**	**Pos #5**	**Pos #4**
% R5+ in CD4+	-0.074 (0.5126)	-0.127 (0.2609)	-0.186 (0.0988)	**-0.223 (0.0467)**	-0.203 (0.0711)	-0.099 (0.3828)	**-0.275 (0.0135)**	**-0.278 (0.0126)**	**-0.279 (0.0121)**	**-0.292 (0.0085)**	**-0.259 (0.0205)**	**-0.305 (0.006)**
Geo Mean R5+ CD4+	-0.083 (0.4639)	-0.12 (0.2883)	-0.084 (0.4609)	-0.037 (0.7439)	-0.166 (0.141)	-0.125 (0.2703)	-0.035 (0.7588)	-0.02 (0.8606)	0.014 (0.9015)	-0.018 (0.8738)	-0.07 (0.538)	-0.092 (0.4186)
Mean R5+ CD4+	-0.168 (0.1354)	-0.033 (0.7683)	-0.142 (0.2074)	-0.174 (0.123)	-0.181 (0.1076)	-0.203 (0.0714)	-0.176 (0.1179)	-0.189 (0.0929)	-0.138 (0.2212)	-0.145 (0.2007)	-0.133 (0.2391)	-0.136 (0.2279)
%R5+ HLA-DR-CD4+	-0.166 (0.1406)	-0.118 (0.2979)	**-0.253 (0.0238)**	**-0.244 (0.029)**	**-0.317 (0.0042)**	-0.181 (0.1086)	**-0.342 (0.0019)**	**-0.356 (0.0012)**	**-0.348 (0.0016)**	**-0.355 (0.0012)**	**-0.32 (0.0038)**	**-0.352 (0.0014)**
Geo Mean R5+ HLA-DR-CD4+	-0.049 (0.667)	-0.005 (0.9652)	-0.123 (0.2768)	-0.047 (0.6772)	-0.202 (0.0729)	-0.153 (0.1741)	-0.084 (0.4575)	-0.088 (0.4368)	-0.036 (0.7542)	-0.064 (0.5757)	-0.115 (0.3088)	-0.13 (0.2491)
Mean R5+ HLA-DR-CD4+	**-0.214 (0.0564)**	-0.097 (0.3906)	**-0.237 (0.0342)**	-0.186 (0.0992)	**-0.267 (0.0165)**	**-0.24 (0.0317)**	**-0.251 (0.0249)**	**-0.241 (0.0316)**	-0.184 (0.1019)	**-0.244 (0.0294)**	**-0.229 (0.0411)**	**-0.226 (0.0436)**
%R5+ HLA-DR+ CD4+	0.017 (0.8797)	-0.033 (0.773)	-0.034 (0.7669)	**-0.221 (0.0486)**	-0.045 (0.6892)	-0.036 (0.7483)	-0.106 (0.3478)	-0.099 (0.3824)	-0.125 (0.2677)	-0.108 (0.3407)	-0.097 (0.3919)	-0.143 (0.2072)
Geo Mean R5+ HLA-DR+ CD4+	-0.084 (0.458)	-0.067 (0.5557)	-0.089 (0.4304)	-0.055 (0.6259)	-0.156 (0.1674)	-0.152 (0.1788)	-0.042 (0.7102)	-0.046 (0.6861)	-0.008 (0.9439)	-0.008 (0.9459)	-0.062 (0.5847)	-0.079 (0.4841)
Mean R5+ HLA-DR+ CD4+	-0.172 (0.1262)	-0.057 (0.6143)	-0.075 (0.5111)	-0.142 (0.2077)	-0.142 (0.2099)	-0.176 (0.1178)	-0.132 (0.2415)	-0.155 (0.1685)	-0.126 (0.2659)	-0.081 (0.4748)	-0.08 (0.4796)	-0.085 (0.4528)

*****p-values are shown parenthesis under the respective correlation coefficient. Significant ($p < 0.05$) correlations are shown in bold.

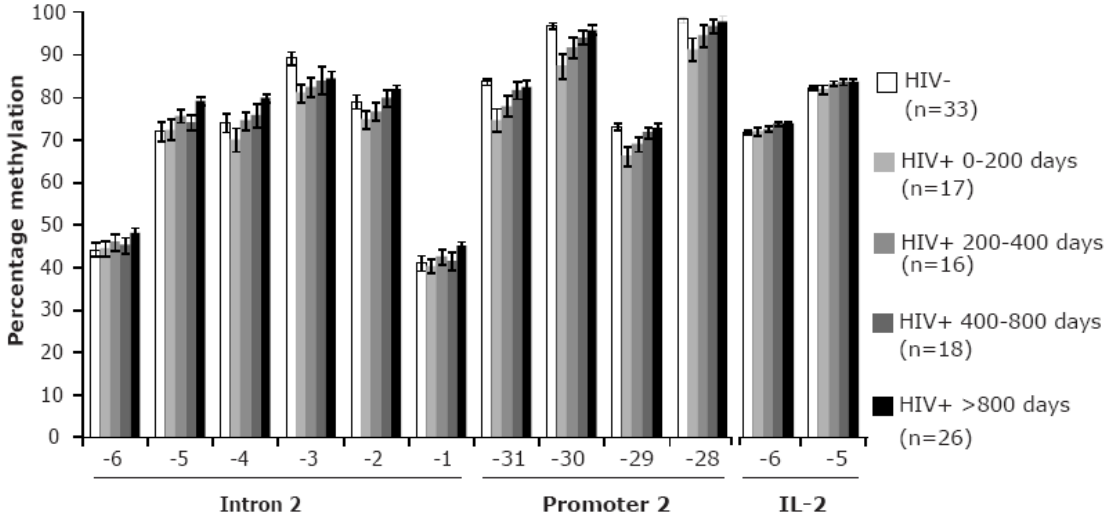
DNA methylation and HIV acquisition

To calculate if there is sample bias,for methylation quantification we compared DNA extracted from blood and compared it to DNA extracted from PBMCs. The results show that there is a significant difference between DNA extracted from whole blood and PBMCs. Bearing this in mind all the comparisons made in this study is from DNA extracted from whole blood.

To determine the impact of DNA methylation on HIV risk of infection, we measured the selected CCR5 promoter2, CCR5 intron2 and IL-2 CpG positions. To account for changes over the course of disease progression, we grouped the HIV infected samples according to days post HIV infection i.e. 0-200, 200-400, 400-800 and >800 days (Figure 1). A comparison of HIV uninfected versus each HIV infected group for the selected CCR5 promoter2, CCR5 intron2 and IL-2 CpG positions was performed. The data shows that there is a significant difference between HIV uninfected and the first HIV infected group (0-200 days) for all CCR5 promoter 2 CpG positions and position 3 of intron 2. Thereafter the percentage methylation at subsequent timepoints showed a steady increase over time. The increase in methylation levels leads to a loss of significance noticed in the later HIV+ groups (400-800 and >800) (Figure 1). Overall, the data shows there is a decrease in the methylation levels from the HIV uninfected group to the HIV infected group within 200 days after infection. Thereafter the methylation levels increases over the course of disease progression.

Also of particular interest, all significant positions show that in the HIV negative individuals DNA methylation is higher than the HIV positive individuals. This could possibly indicate a

protective mechanism found in the HIV negative individuals or it could also show that HIV infection is associated with demethylation of sites as a strategy to increase expression of a key entry coreceptor. It is thought that an increase in DNA methylation leads to a CCR5 gene structure and thus decreases the amount of protein being expressed. This decrease in CCR5 protein expression results in fewer sites available for HIV to bind onto and hence results in a decrease likelihood of HIV acquisition.



	0-200	200-400	400-800	>800
Intron2 Pos -6	0.9158	0.5053	0.6917	0.0611
Intron2 Pos -5	0.9288	0.3167	0.5247	**0.0131**
Intron2 Pos -4	0.2928	0.9108	0.6404	**0.0375**
Intron2 Pos -3	**0.0017**	**0.0095**	0.0795	**0.0396**
Intron2 Pos -2	0.1127	0.3918	0.7475	0.1465
Intron2 Pos -1	0.7905	0.6362	0.8738	0.0826
Promoter2 Pos -31	**0.0001**	**0.0055**	0.2525	0.4458
Promoter2 Pos -30	**0.0001**	**0.0155**	0.0959	0.507
Promoter2 Pos -29	**0.0007**	**0.0107**	0.3324	0.8801
Promoter2 Pos -28	**0.002**	0.07	0.325	0.7825
IL2 Pos -6	0.8461	0.3566	**0.0217**	**0.0112**
IL2 Pos -5	0.7518	0.2312	0.1116	0.0539

P-values of HIV- samples versus HIV+ samples at various timepoints.

Figure 1: CCR5 promoter2, intron2 and IL-2 DNA methylation levels compared against HIV uninfected and infected samples. The HIV infected samples are grouped according to days post infection. Table shows the p-values of HIV negative samples against each HIV positive group.

The HIV negative individuals are then compared to the HIV positive group, thereafter weexamine the methylation patterns for each individual before sero-conversion and compare this to post sero-conversion. The unique CAPRISA cohort was excellently designed for us to test this; since individuals were recruited during the HIV negative phase and followed up till seroconversion and even followed up after sero-conversion.

Figure 2 shows an individual whom we have pre- and post-seroconversion timepoints, the lighter blue bars indicate the HIV negative timepoint, which is calculated to be 116 days before date of infection (calculation for date of infection is explained in the Material and Methods section). The second timepoint is 138 days after date of infection, while the third timepoint is approximately a year (382 days) after infection. There is a significant drop in the DNA methylation levels from pre-infection to post-infection. Most of the selected positions show a significant decrease between the pre and post infection time points. A further important note in this result is after infection the methylation levels remain at about the same level or decreases slightly. This is interesting, because when we examine the individuals viral load and CD4 count it can be assumed this person is progressing faster to AIDS due to a high viral load and low CD4 count (below 400 cells/mm^3) the source of DNA. This therefore implies the methylation levels are tracking the most commonly used markers of disease progression i.e. viral load and CD4 count. The DNA extracted from these samples for the three time points were from whole blood.

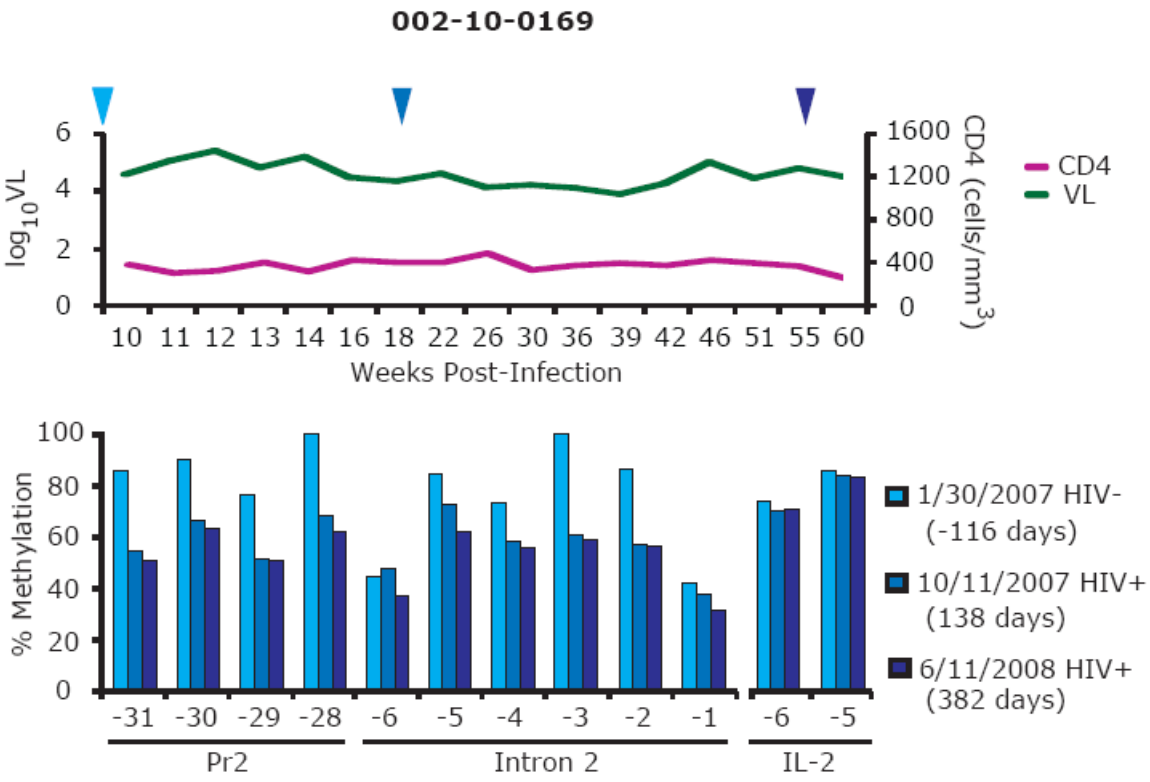


Figure 2: The comparison of the pre and post infection methylation levels from the same individual. The log VL and CD4 count is plotted against weeks post infection. The date drawn for the sample used in the methylation experiments is indicated on the graph.

Examining an individual with a higher CD4 count (>800) and a lower viral load showed that after infection the individuals methylation levels increased over the course of disease progression (Figure 3). This data provides further evidence to the hypothesis that methylation levels are tracking the markers for disease progression CD4 count and viral load. The first timepoint (10/2/2007) was taken 59 days after infection, while the second (12/12/2007) and third (8/26/2008) timepoints were taken 130 and 388 days after infection respectively.

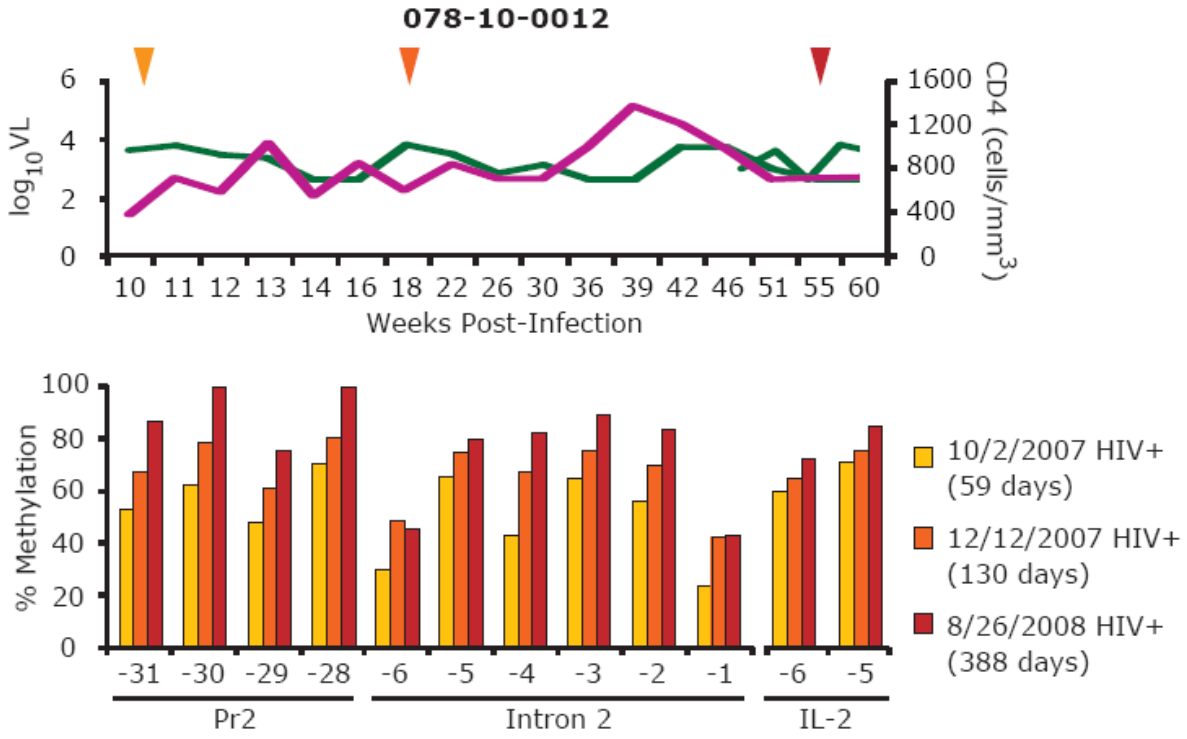


Figure 3: An individual after HIV infection with methylation levels for serial timepoints. The CD4 counts and viral loads are plotted against weeks post infection. This individual has a higher CD4 count (>800 cells/mm^3) and a lower viral load. The methylation levels are measured at specific days after infection and these are indicated on the graph.

Comparison of methylation levels between mucosal and peripheral blood samples.

As shown by many studies the site of HIV transmission i.e. mucosal sites, have differences in the CCR5 expression when compared to peripheral blood (32). Therefore, examined how the methylation levels might play a role in this change in expression. In the CAPRISA HIV negative group vulvovaginal swabs were collected and DNA isolated and methylation levels measured. When we compare the average vaginal and PBMC methylation levels for 10 individuals (figure 4), we notice that the vaginal methylation levels are significantly higher

than the levels found in PBMCs. The CCR5 promoter 2 positions show the most drastic differences in the methylation levels.

Taking a representative individual, when we compared the vaginal sample to that of the matched PBMC sample, there are striking differences between the two (figure 4b). The methylation levels in the vaginal samples are significantly higher than those in the PBMCs. Since these women are female sex workers the route of transmission is probably via the vagina. Higher levels of methylation in the vagina results in lower levels of CCR5 expression, and could therefore indicate a protective mechanism for HIV entry.

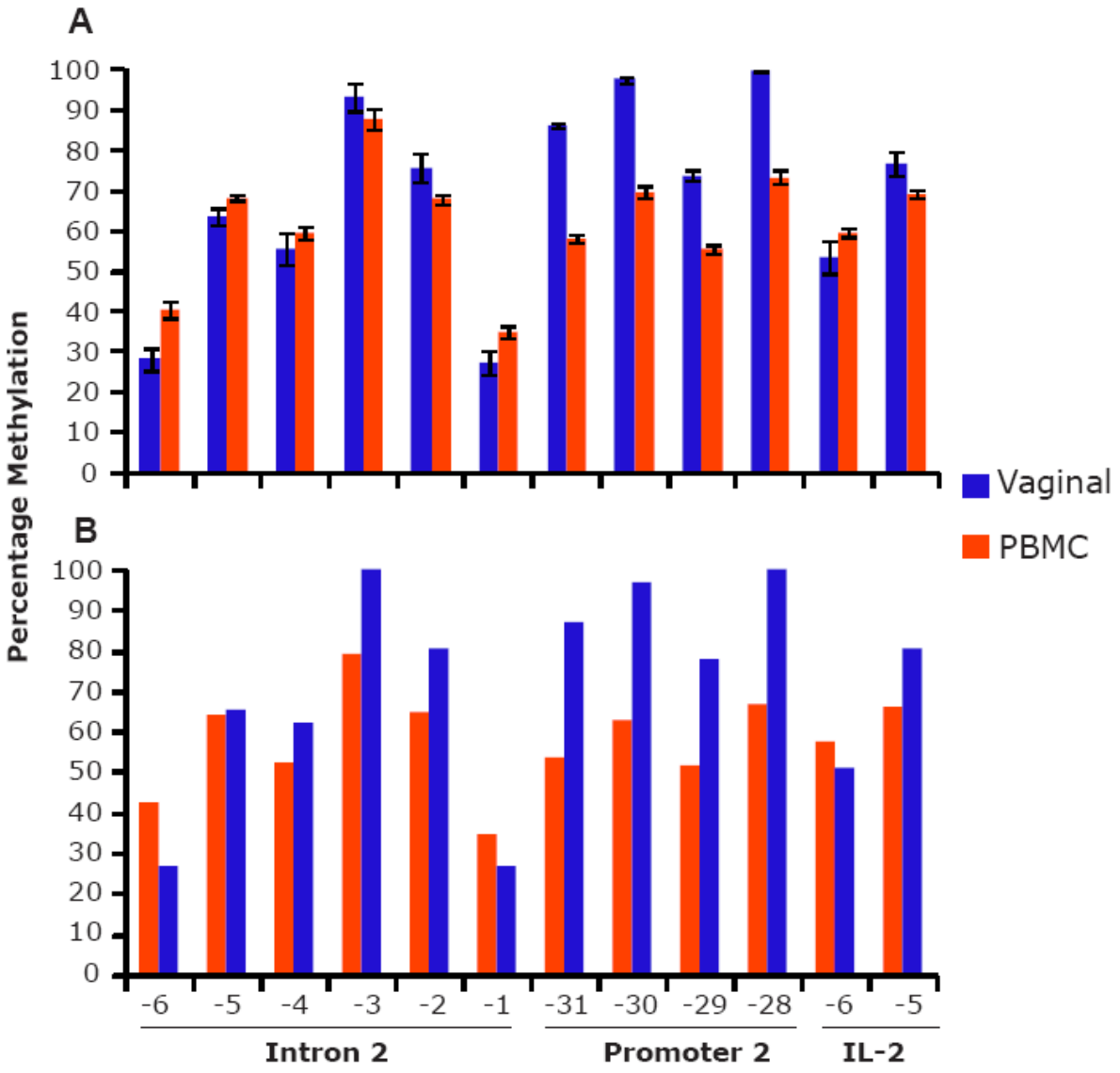


Figure 4: DNA methylation patterns from mucosal tissue compared to that from peripheral blood cells**. (A)** Average DNA methylation levels for vaginal and PBMC samples for 10 individuals, **(B)** representative individual whose DNA methylation levels from the vagina are compared to the DNA methylation levels from the matched PBMC sample.

Discussion

The levels of CCR5 expression have been shown to be extremely important for HIV infection, individuals with lower levels of CCR5 expression have decreased likelihood of acquiring the virus due to the reduced number of coreceptors sites available. In this study we measured the CCR5 expression and to get a better understanding as to what factors effect the CCR5 expression we also measured the DNA methylation and determined how these two factors correlate with each other.

The results show a negative correlation between the CCR5 promoter 2 gene region as well as the CCR5 intron 2 gene region with the CCR5 expression. Upon selecting some sites to measure the DNA methylation levels, it was noticed that these sites correlated negatively with the CCR5 expression i.e. as the methylation levels increased the expression of the CCR5 gene decreased, therefore implying that the methylation levels played a role in the CCR5 expression. The negative correlation on gene expression is particularly important because factors that affect the DNA methylation will directly affect the CCR5 expression and hence influence HIV susceptibility. Important drug discoveries that promote the methylation of a gene could be extremely useful in decreasing the expression of certain genes like CCR5, which assist the virus getting into the host cell. Drugs such as hydralazine, and procainamide are known to inhibit DNMTs, resulting in hypomethylation. By analogy, the recently FDA approved drug decitabine, or 5-azadC for the treatment of myelodysplasic syndrome also induces DNA demethylation by selectively blocking DNMT, in an irreversible manner.

A difference in the DNA methylation levels between the HIV negative and positive individuals indicates that there is a protective mechanism in play found in the HIV negative individuals or active demethylation subsequent to infection. It can be assumed that since the DNA methylation levels are higher in the HIV negative subjects, this leads to a decrease in the expression of CCR5 (as shown in the above experiment and unpublished data) and hence leaves fewer coreceptor binding sites available for HIV-1 entry into the cell. The HIV positive subjects have lower DNA methylation therefore indicating that there is more CCR5 gene expression on the cell surface and thus a greater chance for HIV to bind, hence leading to greater infections. In this study we also decided to follow one individual from the HIV negative state till the HIV positive state and compare the methylation levels in each of these states. Interestingly, the DNA methylation levels in the HIV negative state was much higher that post HIV infection, once again showing that within the same individual there is protection in the HIV negative state and this protection is lost and hence the individual serconverters. Following an individual during the course of HIV-1 progression and measuring the methylation at different points eliminates the patient to patient variation in the methylation levels as explained above. Also the CAPRISA cohort is perfectly suited to measure both uninfected and infected timepoints. This is first study that addresses the question of HIV acquisition and DNA methylation.

Upon comparing different sources of DNA the results indicated there are significant differences in the methylation levels between the samples extracted from whole blood to that of PBMC. DNA extracted from PBMCs show a decrease in DNA methylation while the sample from whole blood are elevated which could be due to the presence of other cell types that increase the methylation content. Therefore that comparison of DNA extracted from different sample sources should be analyzed with much caution.

HIV disease progression is the stage after infection, this stage is extremely important as there are approximately 35 million people infected with HIV around the world. In this study we examined the effect of DNA methylation on disease progression. Using the selected positions mentioned above for the CCR5 promoter 2 and intron 2 regions we measured the methylation levels for two time periods after HIV infection. The result indicated that there was no significant difference in the DNA methylation after HIV infectionindependent of when the samples were drawnThe data indicates that that samples that was drawn one year after infection and compared to a sample two years after infection, for the same individual, there is no significant change in the methylation levels.

The site of infection has also shown to be extremely important for HIV infection, as the mucosal sites show a distinct depletion of CD4+ memory T cells. Previous studies have shown this mucosal CD4+ memory T cells depletion using the simian immunodeficiency virus (SIV), and also shown in the HIV (4) infection model (33, 48). This therefore implying that these mucosal CD4+ T cells are the first line of target cells from HIV due to their increased levels in CCR5 surface expression (32). Examining the levels of DNA methylation for the CCR5 gene, we noticed the mucosal samples had a higher level of DNA methylation compared to DNA extracted from PBMC's. Therefore indicating that at the site of infection, the vagina, there is lower expression of CCR5 therefore leading to a decreased likelihood of acquiring HIV. Our results show in the highly exposed HIV negative individuals examined ,there is higher DNA methylation in the mucosal sites implying lower expression of CCR5 offers protection against HIV-1 infection. Since the individuals are female sex workers the route of viral entry is more likely to be via the mucosal route.

We have some limitations in our study; firstly we were unable to measure the CCR5 expression in the mucosal samples due to the limited number of cells. Secondly, we measured the CCR5 DNA methylation for all cell populations from the mucosal area. This might create bias as other cell type like epithelial cells could inflate the methylation levels. Thirdly, the sample size was relatively small and we could not account for the effect of intercurrent sexually transmitted infections upon HIV acquisition risk. Notwithstanding these limitations, the cohort we evaluated has strengths that tend to offset these limitations. The study design was not a cross-sectional comparison of HIV-negative versus HIV-positive subjects. The study participants were derived from a cohort of high-risk HIV-negative women, including commercial sex workers, who were enrolled when they were HIV-negative and then, prospectively and longitudinally evaluated in an intensive manner.

In summary, we show that there is a significant negative correlation associated with the DNA methylation levels in the CCR5 regulatory regions with the CCR5 expression on the cell surface. Therefore implying, a increase in DNA methylation levels in the CCR5 regulatory regions will result in a decrease on the expression of CCR5 and vice versa. Hence effecting the HIV aquistion. Our results also demonstrate that within the CCR5 regulatory regions there are significant differences in the DNA methylations levels between the HIV- uninfected and HIV–infected samples, furthermore the results also indicate that after HIV-1 infection the DNA methylation levels tend to increase over the course of disease progression. Individuals with a high viral load and low CD4 count decrease the DNA methylation levels, whereas individuals with a low viral load and high CD4 count are able to increase the DNA methylation levels. An importance in examining samples at the site of infection can be

highlighted by the difference in methylation levels noticed between mucosal and peripheral blood samples. Future work would involve exploring samples at the site of infection by measuring the level of expression and stratifying cells into different classes to avoid inflated methylation levels.

Acknowledgements

We thank the study participants, CAPRISA clinical and laboratory staff for providing specimens. Special acknowledgements to the following members of the CAPRISA Acute Infection Study team: C. Williamson, L. Morris, C. Gray and W. Hide. The CAPRISA 002 study was supported by the NIAID, NIH (grant U19 AI 51794). Veron Ramsuran was a recipient of a CAPRISA training fellowship and a Fogarty AITRP fellowship (TWO-0023) and a KwaZulu-Natal Research Institute for TB and HIV (K-RITH) travel award. Thumbi Ndung'u is supported by the Hasso Plattner Foundation and holds the South African DST/NRF Research Chair in Systems Biology of HIV/AIDS. This work was also supported by the VA HIV/AIDS Center of the South Texas Veterans Health Care System, NIH (R37046326), and the Doris Duke Distinguished Clinical Scientist Award to S.K.A. S.K.A. is also supported by a VA MERIT award and the Burroughs Wellcome Clinical Scientist Award in Translational Research.

References

1. 2008. Moving AHEAD with an international human epigenome project. Nature **454:**711-5.
2. **Abdolmaleky, H. M., K. H. Cheng, S. V. Faraone, M. Wilcox, S. J. Glatt, F. Gao, C. L. Smith, R. Shafa, B. Aeali, J. Carnevale, H. Pan, P. Papageorgis, J. F. Ponte, V. Sivaraman, M. T. Tsuang, and S. Thiagalingam.** 2006. Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. Hum Mol Genet **15:**3132-45.
3. **Brady, T., L. M. Agosto, N. Malani, C. C. Berry, U. O'Doherty, and F. Bushman.** 2009. HIV integration site distributions in resting and activated CD4+ T cells infected in culture. Aids **23:**1461-71.
4. **Brenchley, J. M., T. W. Schacker, L. E. Ruff, D. A. Price, J. H. Taylor, G. J. Beilman, P. L. Nguyen, A. Khoruts, M. Larson, A. T. Haase, and D. C. Douek.** 2004. CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J Exp Med **200:**749-59.
5. **Callinan, P. A., and A. P. Feinberg.** 2006. The emerging science of epigenomics. Hum Mol Genet **15 Spec No 1:**R95-101.
6. **Cardno, A. G., F. V. Rijsdijk, P. C. Sham, R. M. Murray, and P. McGuffin.** 2002. A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry **159:**539-45.
7. **Dolan, M. J., H. Kulkarni, J. F. Camargo, W. He, A. Smith, J. M. Anaya, T. Miura, F. M. Hecht, M. Mamtani, F. Pereyra, V. Marconi, A. Mangano, L. Sen, R. Bologna, R. A. Clark, S. A. Anderson, J. Delmar, R. J. O'Connell, A. Lloyd, J. Martin, S. S. Ahuja, B. K. Agan, B. D. Walker, S. G. Deeks, and S. K. Ahuja.** 2007. CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. Nat Immunol **8:**1324-36.
8. **Dupont, J. M., J. Tost, H. Jammes, and I. G. Gut.** 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal Biochem **333:**119-27.
9. **Egger, G., G. Liang, A. Aparicio, and P. A. Jones.** 2004. Epigenetics in human disease and prospects for epigenetic therapy. Nature **429:**457-63.
10. **Esteller, M.** 2005. DNA methylation and cancer therapy: new developments and expectations. Curr Opin Oncol **17:**55-60.
11. **Fang, J. Y., J. A. Mikovits, R. Bagni, C. L. Petrow-Sadowski, and F. W. Ruscetti.** 2001. Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases de novo methylation. J Virol **75:**9753-61.
12. **Foley, D. L., J. M. Craig, R. Morley, C. A. Olsson, T. Dwyer, K. Smith, and R. Saffery.** 2009. Prospects for epigenetic epidemiology. Am J Epidemiol **169:**389-400.
13. **Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y. Z. Wu, C. Plass, and M. Esteller.** 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A **102:**10604-9.
14. **Gartner, K.** 1990. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? Lab Anim **24:**71-7.
15. **Gomes, M. V., and R. A. Waterland.** 2008. Individual epigenetic variation: when, why, and so what? Nestle Nutr Workshop Ser Pediatr Program **62:**141-50; discussion 151-5.
16. **Gopalakrishnan, S., B. O. Van Emburgh, and K. D. Robertson.** 2008. DNA methylation in development and human disease. Mutat Res **647:**30-8.
17. **Gosden, R. G., and A. P. Feinberg.** 2007. Genetics and epigenetics--nature's pen-and-pencil set. N Engl J Med **356:**731-3.
18. **Grunau, C., S. J. Clark, and A. Rosenthal.** 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res **29:**E65-5.

19. **Gutekunst, K. A., F. Kashanchi, J. N. Brady, and D. P. Bednarik.** 1993. Transcription of the HIV-1 LTR is regulated by the density of DNA CpG methylation. J Acquir Immune Defic Syndr **6:**541-9.
20. **Hanash, S.** 2004. Integrated global profiling of cancer. Nat Rev Cancer **4:**638-44.
21. **Hatchwell, E., and J. M. Greally.** 2007. The potential role of epigenomic dysregulation in complex human disease. Trends Genet **23:**588-95.
22. **Hejnar, J., P. Hajkova, J. Plachy, D. Elleder, V. Stepanets, and J. Svoboda.** 2001. CpG island protects Rous sarcoma virus-derived vectors integrated into nonpermissive cells from DNA methylation and transcriptional suppression. Proc Natl Acad Sci U S A **98:**565-9.
23. **Henry, W. K., P. Tebas, and H. C. Lane.** 2006. Explaining, predicting, and treating HIV-associated CD4 cell loss: after 25 years still a puzzle. Jama **296:**1523-5.
24. **Jain, N., A. Rossi, and G. Garcia-Manero.** 2009. Epigenetic therapy of leukemia: An update. Int J Biochem Cell Biol **41:**72-80.
25. **Kaminsky, Z. A., T. Tang, S. C. Wang, C. Ptak, G. H. Oh, A. H. Wong, L. A. Feldcamp, C. Virtanen, J. Halfvarson, C. Tysk, A. F. McRae, P. M. Visscher, G. W. Montgomery, Gottesman, II, N. G. Martin, and A. Petronis.** 2009. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet **41:**240-5.
26. **Kauder, S. E., A. Bosque, A. Lindqvist, V. Planelles, and E. Verdin.** 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog **5:**e1000495.
27. **Kurkjian, C., S. Kummar, and A. J. Murgo.** 2008. DNA methylation: its role in cancer development and therapy. Curr Probl Cancer **32:**187-235.
28. **Lavie, L., M. Kitova, E. Maldener, E. Meese, and J. Mayer.** 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). J Virol **79:**876-83.
29. **Lim, H. G., K. Suzuki, D. A. Cooper, and A. D. Kelleher.** 2008. Promoter-targeted siRNAs induce gene silencing of simian immunodeficiency virus (SIV) infection in vitro. Mol Ther **16:**565-70.
30. **Liu, L., Y. Li, and T. O. Tollefsbol.** 2008. Gene-environment interactions and epigenetic basis of human diseases. Curr Issues Mol Biol **10:**25-36.
31. **Matouskova, M., J. Blazkova, P. Pajer, A. Pavlicek, and J. Hejnar.** 2006. CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues. Exp Cell Res **312:**1011-20.
32. **Mattapallil, J. J., D. C. Douek, B. Hill, Y. Nishimura, M. Martin, and M. Roederer.** 2005. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. Nature **434:**1093-7.
33. **Mattapallil, J. J., Z. Smit-McBride, M. McChesney, and S. Dandekar.** 1998. Intestinal intraepithelial lymphocytes are primed for gamma interferon and MIP-1beta expression and display antiviral cytotoxic activity despite severe CD4(+) T-cell depletion in primary simian immunodeficiency virus infection. J Virol **72:**6421-9.
34. **Mikovits, J. A., H. A. Young, P. Vertino, J. P. Issa, P. M. Pitha, S. Turcoski-Corrales, D. D. Taub, C. L. Petrow, S. B. Baylin, and F. W. Ruscetti.** 1998. Infection with human immunodeficiency virus type 1 upregulates DNA methyltransferase, resulting in de novo methylation of the gamma interferon (IFN-gamma) promoter and subsequent downregulation of IFN-gamma production. Mol Cell Biol **18:**5166-77.
35. **Mori, Y., J. Yin, F. Sato, A. Sterian, L. A. Simms, F. M. Selaru, K. Schulmann, Y. Xu, A. Olaru, S. Wang, E. Deacu, J. M. Abraham, J. Young, B. A. Leggett, and S. J. Meltzer.** 2004. Identification of genes uniquely involved in frequent microsatellite instability colon carcinogenesis by expression profiling combined with epigenetic scanning. Cancer Res **64:**2434-8.
36. **Neidhart, M., J. Rethage, S. Kuchen, P. Kunzler, R. M. Crowl, M. E. Billingham, R. E. Gay, and S. Gay.** 2000. Retrotransposable L1 elements expressed in rheumatoid arthritis synovial

tissue: association with genomic DNA hypomethylation and influence on gene expression. Arthritis Rheum **43:**2634-47.

37. **Nouzova, M., N. Holtan, M. M. Oshiro, R. B. Isett, J. L. Munoz-Rodriguez, A. F. List, M. L. Narro, S. J. Miller, N. C. Merchant, and B. W. Futscher.** 2004. Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. J Pharmacol Exp Ther **311:**968-81.
38. **Pons, D., F. R. de Vries, P. J. van den Elsen, B. T. Heijmans, P. H. Quax, and J. W. Jukema.** 2009. Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. Eur Heart J **30:**266-77.
39. **Ptak, C., and A. Petronis.** 2008. Epigenetics and complex disease: from etiology to new therapeutics. Annu Rev Pharmacol Toxicol **48:**257-76.
40. **Sandovici, I., S. Kassovska-Bratinova, J. C. Loredo-Osti, M. Leppert, A. Suarez, R. Stewart, F. D. Bautista, M. Schiraldi, and C. Sapienza.** 2005. Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. Hum Mol Genet **14:**2135-43.
41. **Silvestri, G., M. Paiardini, I. Pandrea, M. M. Lederman, and D. L. Sodora.** 2007. Understanding the benign nature of SIV infection in natural hosts. J Clin Invest **117:**3148-54.
42. **Smith, M.** 2009. The year in human and medical genetics. Highlights of 2007-2008. Ann N Y Acad Sci **1151:**1-21.
43. **Strickland, F. M., and B. C. Richardson.** 2008. Epigenetics in human autoimmunity. Epigenetics in autoimmunity - DNA methylation in systemic lupus erythematosus and beyond. Autoimmunity **41:**278-86.
44. **Szyf, M.** 2009. Epigenetics, DNA methylation, and chromatin modifying drugs. Annu Rev Pharmacol Toxicol **49:**243-63.
45. **Tufarelli, C., J. A. Stanley, D. Garrick, J. A. Sharpe, H. Ayyub, W. G. Wood, and D. R. Higgs.** 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet **34:**157-65.
46. **van der Maarel, S. M.** 2008. Epigenetic mechanisms in health and disease. Ann Rheum Dis **67 Suppl 3:**iii97-100.
47. **van Vliet, J., N. A. Oates, and E. Whitelaw.** 2007. Epigenetic mechanisms in the context of complex diseases. Cell Mol Life Sci **64:**1531-8.
48. **Veazey, R. S., M. DeMaria, L. V. Chalifoux, D. E. Shvetz, D. R. Pauley, H. L. Knight, M. Rosenzweig, R. P. Johnson, R. C. Desrosiers, and A. A. Lackner.** 1998. Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science **280:**427-31.
49. **Weinmann, A. S., P. S. Yan, M. J. Oberley, T. H. Huang, and P. J. Farnham.** 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev **16:**235-44.
50. **Wilson, A. G.** 2008. Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. J Periodontol **79:**1514-9.
51. **Wong, A. H., Gottesman, II, and A. Petronis.** 2005. Phenotypic differences in genetically identical organisms: the epigenetic perspective. Hum Mol Genet **14 Spec No 1:**R11-8.
52. **Zardo, G., M. I. Tiirikainen, C. Hong, A. Misra, B. G. Feuerstein, S. Volik, C. C. Collins, K. R. Lamborn, A. Bollen, D. Pinkel, D. G. Albertson, and J. F. Costello.** 2002. Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. Nat Genet **32:**453-8.

CHAPTER 7 - DISCUSSION

Host genetic factors have been studied to understand how these mechanisms achieve viral protection. The sex worker cohorts provides a great opportunity for testing these host genetic factors. Furthermore one of the areas that has been severely affected by the HIV pandemic is sub-Saharan African, yet most HIV studies are conducted in the western world. There has also been data showing that females have an elevated risk of acquiring HIV (94, 162) therefore our study is particularly important since we examined black female sex worker cohort for host genetic factors that may play a role in conferring resistance to HIV infection.

Not much is known about the HIV protective mechanisms in individuals of African descent. Therefore we decided to focus our attention on one of the genes that has shown to be the most effective and potent HIV protective mechanisms of all the host genetic factors studied. The CCR5 gene that has the CCR5 Δ32 polymorphism is predominantly found in individuals of European descent. However, since individuals of African descent have this polymorphism at very low levels, we focussed on different aspects of the CCR5 gene that have shown to play a role in HIV resistance, other genetic polymorphisms found in the gene region, the CCR5 ligands- *CCL3L* and *CCL4L*, CCR5 expression, DNA methylation in the CCR5 promoter and intron regions. We also examined another receptor for chemokines that has been shown to play a role in HIV protection in individuals of African descent, i.e. DARC (114), although this association has been disputed in some studies (129, 140, 308, 316).

There are various gaps in knowledge regarding the above mentioned mechanisms; this study is aimed to address some of these gaps in knowledge. After examining the results from this study indicates that we have determined various factors that play a crucial role in HIV

resistance and hence demonstrate that there are some vital protection mechanisms against disease progression. This is the first study, to our knowledge, that extensively examined these genetic associations in susceptibility to HIV-1 infection in an adult African cohort.

DARC

A study done by He *et al*, showed that a polymorphism found in the promoter region of the *DARC* gene conferred resistance to HIV infection and disease progression (114). However, four independent groups reported that they could not replicate He et al findings (129, 140, 308, 316). He et al in their riposte explained that their reported associations was not replicated due to population stratification, and suggested that the disparate associations of *DARC* genotype with HIV-AIDS susceptibility may be explained partly by differences in cohort characteristics and endpoints evaluated (113). A further study done by Kulkarni et al found that the influence of the Duffy-null state on clinical outcome was evident mainly in Duffy-null subjects who at HIV diagnosis during the early stages of their infection also had low WBC counts (158). Another study done by Detels et al in 1994, had demonstrated that in men from the Multicenter AIDS Cohort Study (MACS) resistance to HIV infection was due to the influence of circulating levels of leukocyte subsets with high neutrophil and CD8+ T cell counts (64). Reich et al showed that despite the strong association with African ancestry, *DARC -46C/C* may be a causal variant for the leukopenia/neutropenia observed in an African American cohort (250).

In this study, using the above information as a baseline of knowledge we explored the CAPRISA cohort since this cohort is particularly suited for these analyses. Our study yielded three major findings, high platelet and low WBC counts associated with increased HIV susceptibility, the *DARC*-negative trait is a strong predictor of neutrophil counts and finally *DARC*-negative subjects with <2,500 neutrophils/mm^3 had a ~2.5-fold increased risk and rate of acquiring HIV relative to all other study participants.

Results are reproducible with previous studies, that a high neutrophil count is associated with HIV resistance (64), in our study we show that a high platelet count and low WBC count is associated with an increased risk of HIV susceptibility. Furthermore, the low WBC is predominantly due to the low neutrophils, and demonstrated here that individuals with <2,500 neutrophils/mm^3 have an increased likelihood of acquiring HIV. When we compared individuals that had a baseline count of <2,500 neutrophils/mm^3, these individuals (27% of cohort at enrolment) had a nearly three-fold increased HIV risk when compared to individuals having higher initial neutrophil counts.

The GWAS result using the neutrophil count as a quantitative outcome indicated that the DARC-negative trait (*DARC* T-46C) associated most strongly with neutrophil counts with a p-value of $1.4x10^{-8}$, and after removing outliers p-value is $2.69x10^{-8}$, when we adjust the top ten PCs the p-value is $p=7.9x10^{-11}$. This result was consistent with previous findings (250).

When we examine the third and novel finding, it shows the DARC-negative individuals who have a low neutrophil count of <2,500 neutrophils/mm^3 are at a much greater risk (about 2.5

fold) of acquiring HIV. These results suggest that there is a close triangulation among a pre-infection low neutrophil count, Duffy-null state and HIV risk. However, because of the close association of the Duffy-null state with low neutrophil counts, it is not possible to ascribe an independent effect of low neutrophil counts or the DARC-null state on HIV risk, and hence these two host factors may as act either independently or as a conjoint genetic-cellular determinant.

The result obtained showing platelet counts associating with HIV risk, is consistent with multiple lines of evidence suggesting that platelets provide a key interface between coagulation and inflammation/immunity (62, 289), which in turn may play an important role in HIV risk; notably these pathways have also been implicated in HIV disease outcome (159). Also since the neutrophils are the primary defenders against infections and also play a vital role in the adaptive immunity against HIV (215). Also, since the neutrophils are a major source of β-defensins 1–4, this could possibly be another control mechanism used for HIV replication.

. Using HRW/CSW cohort from Africa we demonstrate that low neutrophil and high platelet counts associated with an increased risk of acquiring HIV, and the Duffy-null state associates strongly with a low neutrophil count, and finally showing that Duffy-null subjects those who also have a low neutrophil count have an increased risk and rate of acquiring HIV infection. Since a low neutrophil count and Duffy-null state is prevalent in individuals of African descent, the combination of these two detrimental factors may be responsible for some of the increase in HIV. Further work will need to be conducted in order to determine the impact of this unfavorable genotype-hematological determinant in people of African ancestry.

To the best of our knowledge this study represent the first detailed characterization of the combinatorial gene content generated within a CNVR in world-wide populations. The results indicate that complex CNVs within a chemokine gene-rich region on chromosome 17q12 results in a striking degree of inter-individual and inter-population differences in chemokine gene content. Furthermore, evaluation of a well-characterized cohort of South African women, that included commercial sex workers, demonstrated that it is the combinatorial content of *CCL3L*- and *CCL4L*-related genes rather than the gene dose of any one of these genes that affects both HIV susceptibility and disease progression. Specifically, our results point to an important role for the balance between the copy number of protective (*CCL3La* and *CCL3Lb*) versus detrimental (e.g., *CCL4Lb*) chemokine genes in influencing HIV susceptibility, viral replication and CD4+ T cell loss. Thus, our results highlight that the CNV of chemokine genes are an excellent model system to (i) evaluate the extent of genomic architectural complexity in the form of smaller CNV within a segmental duplication or CNVR, (ii) gain further insights into the evolutionary history of CNV of immune response genes, (iii) investigate the impact of this genomic complexity on susceptibility to human diseases especially infectious, autoimmune or inflammatory diseases, and (iv) advance the concept that failure to account for the combinatorial genomic content of distinct CNVs, including their population-specific distributions, may both confound association studies and pose challenge for the full characterization of human genome sequence.

The extensive combinatorial content of chemokine genes is unprecedented and is consequence of complex CNV of *CCL3L* and *CCL4L* genes that can be broadly categorized into those that encode classical secreted chemokines, i.e., *CCL3La* (*CCL3L* and *CCL3L2*) and *CCL4La* (*CCL4L*2), versus non-classical chemokines, i.e., *CCL3Lb* (*CCL3L2*) and *CCL4Lb* (*CCL4L1*); *CCL3L* and *CCL3L3* are classified as *CCL3La* because they both are predicted to encode identical proteins.

From an evolutionary perspective, the duplication of potent immune response chemokine genes may be reflective of eons of co-evolution of primate and viral genomes (100). This is underscored by what appears to be two evolutionary trade-offs: first, increasing dosages of genes that transcribe classical secreted chemokines (*CCL3La* and *CCL3Lb*) are offset by increasing dosages of genes that transcribe mRNA isoforms that are not classical chemokines; and second, the observed inverse correlation between the copy number of *CCL4L*a (secreted chemokine) and *CCL4L*b (non-secreted chemokine) suggests that in some populations there might be an evolutionary pressure to maintain a balance between classical and non-classical *CCL4L* genes. Because high gene doses of functional chemokines would have imparted a persistent pro-inflammatory state, it is conceivable that the presence of these non-classical chemokine genes may have evolved to counter-balance the pro-inflammatory effects associated with possession of a high gene content of secreted classical chemokines. This thesis is supported by the observation that activated CD8+ T cells from subjects who are null for the gene encoding *CCL4L*a (the classical secreted *CCL4L* variant), but possess *CCL4Lb* gene (encoding a non-classical chemokine) express lower amounts CCL4, a CCR5-specific chemokine (49). By analogy, in preliminary studies, we have found that possession of *CCL3Lb* is associated with reduced expression of the classical chemokine *CCL3L*

following stimulation of peripheral blood mononuclear cells with the microbial product lipopolysaccharide.

The cohort of HIV-uninfected and –infected South African women afforded the unique opportunity to determine whether the combinatorial content of chemokine genes versus the dose of a single chemokine gene (e.g. *CCL3L*) influenced HIV susceptibility. We identify concordant associations for duplicated chemokine genes across three phenotypic endpoints: HIV susceptibility, viral load setpoint, and rate of $CD4^+$ T cell decline. Specifically, we find that these phenotypic endpoints are influenced not only by the quantitative content of each chemokine gene *per se*, but by the combinatorial content or balance of these genes.

Our findings suggest that restricting CNV-phenotype analysis solely to the *CCL3L*-segmental duplication alone, although providing an aggregate estimate of all *CCL3L*-related genes, may in some instances provide an incomplete assessment of the phenotypic impact of chemokine CNVs for three main reasons. First, our findings demonstrate that there is a complex population-specific correlation pattern between the CNV of different chemokine genes and hence as reflected in this study, failure to account for the effects of both *CCL3L* and *CCL4L* genes obscures the associations of this locus with HIV-AIDS susceptibility. Second, a role for the CNV of *CCL4L* independent of *CCL3L* in HIV-AIDS has been identified in cohorts of European ancestry (49, 275): in a cohort of adults from Spain demonstrated that subjects lacking *CCL4La* but possessing *CCL4Lb* had an increased risk of acquiring HIV infection (49). A separate study, demonstrated that in Ukranian children exposed perinatally to HIV the impact on HIV-AIDS susceptibility of the chemokine gene-rich locus on 17q12 was dependent on the balance between the doses of genes conferring protective (*CCL3L*a and

*CCL4L*a) versus detrimental (*CCL4L*b) effects; the frequency of subjects with *CCL3Lb* was too low to determine an independent effect of this gene (275). Concordantly, we find that a similar balance also impacts on HIV susceptibility and laboratory parameters predictive of AIDS risk in the CAPRISA cohort. Third, the region on 17q12 that has undergone segmental duplication is large (~100 kb) and is thought to be replete with many breakpoints that can lead to duplicated segments that vary in chemokine gene content amongst individuals (40, 202, 249, 293).

On the basis of the known HIV-suppressive effects of *CCL3L*/3 (*CCL3La*) and our preliminary results suggesting that *CCL3Lb* gene product may impart anti-inflammatory properties (data not shown), we propose a model wherein *CCL3L*-related genes may influence HIV susceptibility through a dual mechanism: firstly, via the anti-HIV effects associated with *CCL3La* gene products (*CCL3L* and *CCL3L*3) and secondly, by dampening inflammation, a key component of HIV pathogenicity, via the *CCL3L*b gene product.

In conjunction with the results of previous studies (reviewed in (112)), including a recent report suggesting that HIV preferentially selects for subjects with a low gene dose of the *CCL3L*-containing segmental duplication, these findings support a strong link between chemokine gene CNV and HIV-AIDS susceptibility.

The associations of the *CCL3L*-containing segmental duplication with HIV-AIDS susceptibility have been contested by two groups (21, 298), and Field et al asserted that the methods employed to quantify *CCL3L*-related genes were inaccurate (79). He et al in their

riposte to these contestations outlined methodological, genetic and epidemiologic reasons why inter-cohort differences in the associations of *CCL3L* CNV may have been observed (112). The results obtained in this study support the original findings that individuals who possess low CCL3L gene containing segmental duplications have an increased likelihood of acquiring HIV.

When we calculate the effect of the complex patterns of chemokine gene content impact upon evaluation of vaccine efficacy, we notice that individuals that represent good candidates for vaccine trials are high-risk individuals such as those included in this study. We found that the women who seroconverted during prospective follow-up were more likely to possess a low dose of *CCL3L* genes at the time of enrollment into the study. Analyses of intravenous drug users, another high-risk group, suggested that HIV may preferentially select for subjects with a low dose of the *CCL3L*-containing segmental duplication (133). Hence, hypothetically enhanced susceptibility associated with a low copy number may overshadow or nullify the benefits of a partially protective vaccine. Furthermore, because communicability is related to the viral load of the transmitting partner (7, 47, 103, 243, 310), chemokine gene content, by influencing the degree of viremia and immune deficiency as reflected by CD4 cell declines, may also affect sexual transmission rate by vaccinated subjects who become infected with HIV subsequent to vaccination. Thus, as chemokine gene balance impacts on four parameters relevant to the evaluation of vaccine efficacy (acquisition, viral load, $CD4^+$ T cell decline, and transmission rate) (91, 294), accounting for combinatorial chemokine gene content should improve the ability to assess true vaccine efficacy. This premise is also supported by our modeling studies, wherein we find that even low levels of unbalanced trial arm allocation of participants with protective or detrimental *CCL3L-CCR5* genotypes may confound the outcome of a vaccine trial (157). Further, the immunological phenotypes linked to $CCL3L^{high}$-

CCL4L^{low} and *CCL3L*^{low}-*CCL4L*^{high}, the most protective and detrimental genotypes against HIV acquisition detected in this study, may provide important information regarding the immune correlates of protection.

The discovery of the complex combinatorial inheritance pattern of chemokine genes has broad relevance to gaining a better appreciation of the phenotypic impact of CNVs on a wide variety of diseases (53, 183, 194, 249, 268, 323). We suggest that given the population-specific nature of the chemokine CNVs, and possibly other CNVs, this appreciation will require careful attention to the epidemiologic features of the study as well as fine-scale analyses of the phenotypic impacts of distinct CNVs, rather than assessment of any single component of a genomic region that has undergone segmental duplications. A similar premise was also advanced by the authors of a study that used high-resolution array-based comparative genomic hybridization to interrogate known CNV regions (231). Hence, despite the considerable excitement over the potential functional significance of CNVs, genotype-phenotype studies directed at uncovering the disease phenotypes of CNVs may need to account for this largely uncharacterized complexity, which may be a difficult undertaking using solely whole genome-sequencing approaches.

DNA methylation and CCR5 expression

The levels of CCR5 expression have been shown to be extremely important for HIV infection, individuals with lower levels of CCR5 expression have decreased likelihood of acquiring the virus due to the reduced number of coreceptors sites available. In this study we measured the CCR5 expression and in order to help understand which factors play a role in the change in the CCR5 expression we also measured the DNA methylation and determined how these two factors correlate with each other.

There were a number of different and important findings in this study and can be summarised as follows; firstly, we show that there is a strong significant negative correlation associated with the DNA methylation levels in the CCR5 regulatory regions with the CCR5 expression on the cell surface. Secondly, our results also demonstrate that within the CCR5 regulatory regions there are significant differences in the DNA methylations levels between the HIV uninfected and –infected samples. Thirdly, the results also indicate that after HIV-1 infection the DNA methylation levels tend to increase over the course of disease progression. Individuals with a high viral load and low CD4 count decrease the DNA methylation levels, whereas individuals with a low viral load and high CD4 count are able to increase the DNA methylation levels. Finally we noticed differences in methylation levels between mucosal and peripheral blood samples displays the importance of examining samples at the site of infection.

These results demonstrate the importance of host epigenetics on infectious diseases, by showing changes in DNA methylation levels results in changes in the gene expression and

hence results susceptibily or protection to HIV-1. Further investigation is required to understand the mechanisms.

CHAPTER 8 - CONCLUSION

The various results obtained in this study support the proposal that there are genetic and epigenetic determinants in chemokines and chemokine receptor genes that influence HIV-1 susceptibility and disease progression. The cohort examined in this study included individuals of African ancestry, with HIV-1 high risk female sex workers, we show there are various independent factors that enhance or protect these individuals against HIV-1 infection and disease progression. Firstly, pre-seroconversion neutrophil and platelet counts influence risk of HIV infection. The trait of Duffy-null-associated low neutrophil counts influences HIV susceptibility. Secondly, the combinatorial content of the genes in the 17q12 region, namely, *CCL3L*- and *CCL4L*-related genes rather than the gene dose of any one of these genes influences both HIV susceptibility and disease progression. Specifically, our results point to an important role for the balance between the copy numbers of protective (*CCL3La* and *CCL3L*b) versus detrimental (*CCL4Lb*) chemokine genes in influencing HIV susceptibility, viral replication and CD4+ T cell loss. Our findings underscore that subjects with a low dose of *CCL3L* genes are preferentially infected with HIV. Finally, we demonstrate a strong negative correlation between the methylation levels from the CCR5 promoter and intron regions with the CCR5 expression levels and further show that the DNA methylation status is lower in HIV+ subjects compared to HIV- subjects.

There is extensive heterogeneity in the risk of acquiring HIV-1 and in the rate of disease progression among HIV infected persons. These studies reinforce that both genetic and epigenetic factors are likely to influence HIV-AIDS pathogenesis. Hence, when evaluating a vaccine in subjects enriched for protective genetic-epigenetic factors, misallocation of favorable versus unfavorable genetic-epigenetic factors in trial arms might result in spurious estimates of a vaccine's efficacy because it might be difficult to differentiate between protective effects of the HIV vaccine versus those of protective genetic-epigenetic factors.

CHAPTER 9 - REFERENCES

1. 2008. Moving AHEAD with an international human epigenome project. Nature **454:**711-5.
2. **Abdolmaleky, H. M., K. H. Cheng, S. V. Faraone, M. Wilcox, S. J. Glatt, F. Gao, C. L. Smith, R. Shafa, B. Aeali, J. Carnevale, H. Pan, P. Papageorgis, J. F. Ponte, V. Sivaraman, M. T. Tsuang, and S. Thiagalingam.** 2006. Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. Hum Mol Genet **15:**3132-45.
3. **Ahuja, S. K., H. Kulkarni, G. Catano, B. K. Agan, J. F. Camargo, W. He, R. J. O'Connell, V. C. Marconi, J. Delmar, J. Eron, R. A. Clark, S. Frost, J. Martin, S. S. Ahuja, S. G. Deeks, S. Little, D. Richman, F. M. Hecht, and M. J. Dolan.** 2008. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. Nat Med **14:**413-20.
4. **Aitman, T. J., R. Dong, T. J. Vyse, P. J. Norsworthy, M. D. Johnson, J. Smith, J. Mangion, C. Roberton-Lowe, A. J. Marshall, E. Petretto, M. D. Hodges, G. Bhangal, S. G. Patel, K. Sheehan-Rooney, M. Duda, P. R. Cook, D. J. Evans, J. Domin, J. Flint, J. J. Boyle, C. D. Pusey, and H. T. Cook.** 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature **439:**851-5.
5. **Alkhatib, G., C. Combadiere, C. C. Broder, Y. Feng, P. E. Kennedy, P. M. Murphy, and E. A. Berger.** 1996. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science **272:**1955-8.
6. **An, P., G. Bleiber, P. Duggal, G. Nelson, M. May, B. Mangeat, I. Alobwede, D. Trono, D. Vlahov, S. Donfield, J. J. Goedert, J. Phair, S. Buchbinder, S. J. O'Brien, A. Telenti, and C. A. Winkler.** 2004. APOBEC3G genetic variants and their influence on the progression to AIDS. J Virol **78:**11070-6.
7. **Anderson, R., and M. Hanson.** 2005. Potential public health impact of imperfect HIV type 1 vaccines. J Infect Dis **191 Suppl 1:**S85-96.
8. **Ansari-Lari, M. A., X. M. Liu, M. L. Metzker, A. R. Rut, and R. A. Gibbs.** 1997. The extent of genetic variation in the CCR5 gene. Nat Genet **16:**221-2.
9. **Antequera, F.** 2003. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci **60:**1647-58.
10. **Aoki, K., N. Sato, A. Yamaguchi, O. Kaminuma, T. Hosozawa, and S. Miyatake.** 2009. Regulation of DNA demethylation during maturation of CD4+ naive T cells by the conserved noncoding sequence 1. J Immunol **182:**7698-707.
11. **Arenzana-Seisdedos, F., and M. Parmentier.** 2006. Genetics of resistance to HIV infection: Role of co-receptors and co-receptor ligands. Semin Immunol **18:**387-403.
12. **Arhel, N., and F. Kirchhoff.** 2010. Host proteins involved in HIV infection: new therapeutic targets. Biochim Biophys Acta **1802:**313-21.
13. **Ashorn, P. A., E. A. Berger, and B. Moss.** 1990. Human immunodeficiency virus envelope glycoprotein/CD4-mediated fusion of nonprimate cells with human cells. J Virol **64:**2149-56.
14. **Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler.** 2002. Recent segmental duplications in the human genome. Science **297:**1003-7.
15. **Banki, Z., D. Wilflingseder, C. G. Ammann, M. Pruenster, B. Mullauer, K. Hollander, M. Meyer, G. M. Sprinzl, J. van Lunzen, H. J. Stellbrink, M. P. Dierich, and H. Stoiber.** 2006. Factor I-mediated processing of complement fragments on HIV immune complexes targets HIV to CR2-expressing B cells and facilitates B cell-mediated transmission of opsonized HIV to T cells. J Immunol **177:**3469-76.
16. **Barreto, G., A. Schafer, J. Marhold, D. Stach, S. K. Swaminathan, V. Handa, G. Doderlein, N. Maltry, W. Wu, F. Lyko, and C. Niehrs.** 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature **445:**671-5.
17. **Beckmann, J. S., X. Estivill, and S. E. Antonarakis.** 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet **8:**639-46.

18. **Begaud, E., L. Chartier, V. Marechal, J. Ipero, J. Leal, P. Versmisse, G. Breton, A. Fontanet, C. Capoulade-Metay, H. Fleury, F. Barre-Sinoussi, D. Scott-Algara, and G. Pancino.** 2006. Reduced CD4 T cell activation and in vitro susceptibility to HIV-1 infection in exposed uninfected Central Africans. Retrovirology **3:**35.
19. **Beretta, A., S. H. Weiss, G. Rappocciolo, R. Mayur, C. De Santis, J. Quirinale, A. Cosma, P. Robbioni, G. M. Shearer, J. A. Berzofsky, and et al.** 1996. Human immunodeficiency virus type 1 (HIV-1)-seronegative injection drug users at risk for HIV exposure have antibodies to HLA class I antigens and T cells specific for HIV envelope. J Infect Dis **173:**472-6.
20. **Bestor, T. H.** 1988. Cloning of a mammalian DNA methyltransferase. Gene **74:**9-12.
21. **Bhattacharya, T., J. Stanton, E. Y. Kim, K. J. Kunstman, J. P. Phair, L. P. Jacobson, and S. M. Wolinsky.** 2009. CCL3L1 and HIV/AIDS susceptibility. Nat Med **15:**1112-5.
22. **Bird, A.** 2002. DNA methylation patterns and epigenetic memory. Genes Dev **16:**6-21.
23. **Bird, A.** 2007. Perceptions of epigenetics. Nature **447:**396-8.
24. **Blanpain, C., B. Lee, M. Tackoen, B. Puffer, A. Boom, F. Libert, M. Sharron, V. Wittamer, G. Vassart, R. W. Doms, and M. Parmentier.** 2000. Multiple nonfunctional alleles of CCR5 are frequent in various human populations. Blood **96:**1638-45.
25. **Bleul, C. C., L. Wu, J. A. Hoxie, T. A. Springer, and C. R. Mackay.** 1997. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. Proc Natl Acad Sci U S A **94:**1925-30.
26. **Blewitt, M. E., S. Chong, and E. Whitelaw.** 2004. How the mouse got its spots. Trends Genet **20:**550-4.
27. **Brady, T., L. M. Agosto, N. Malani, C. C. Berry, U. O'Doherty, and F. Bushman.** 2009. HIV integration site distributions in resting and activated CD4+ T cells infected in culture. Aids **23:**1461-71.
28. **Brass, A. L., D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, J. Lieberman, and S. J. Elledge.** 2008. Identification of host proteins required for HIV infection through a functional genomic screen. Science **319:**921-6.
29. **Broder, C. C., and R. G. Collman.** 1997. Chemokine receptors and HIV. J Leukoc Biol **62:**20-9.
30. **Bruder, C. E., A. Piotrowski, A. A. Gijsbers, R. Andersson, S. Erickson, T. D. de Stahl, U. Menzel, J. Sandgren, D. von Tell, A. Poplawski, M. Crowley, C. Crasto, E. C. Partridge, H. Tiwari, D. B. Allison, J. Komorowski, G. J. van Ommen, D. I. Boomsma, N. L. Pedersen, J. T. den Dunnen, K. Wirdefeldt, and J. P. Dumanski.** 2008. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet **82:**763-71.
31. **Bruniquel, D., and R. H. Schwartz.** 2003. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol **4:**235-40.
32. **Burns, J. C., C. Shimizu, E. Gonzalez, H. Kulkarni, S. Patel, H. Shike, R. S. Sundel, J. W. Newburger, and S. K. Ahuja.** 2005. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. J Infect Dis **192:**344-9.
33. **Bushman, F. D., N. Malani, J. Fernandes, I. D'Orso, G. Cagney, T. L. Diamond, H. Zhou, D. J. Hazuda, A. S. Espeseth, R. Konig, S. Bandyopadhyay, T. Ideker, S. P. Goff, N. J. Krogan, A. D. Frankel, J. A. Young, and S. K. Chanda.** 2009. Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS Pathog **5:**e1000437.
34. **Callinan, P. A., and A. P. Feinberg.** 2006. The emerging science of epigenomics. Hum Mol Genet **15 Spec No 1:**R95-101.
35. **Camargo, J. F., M. P. Quinones, S. Mummidi, S. Srinivas, A. A. Gaitan, K. Begum, F. Jimenez, S. VanCompernolle, D. Unutmaz, S. S. Ahuja, and S. K. Ahuja.** 2009. CCR5 expression levels influence NFAT translocation, IL-2 production, and subsequent signaling events during T lymphocyte activation. J Immunol **182:**171-82.

36. **Campbell, J. D., M. J. Stinson, F. E. Simons, and K. T. HayGlass.** 2002. Systemic chemokine and chemokine receptor responses are divergent in allergic versus non-allergic humans. Int Immunol **14:**1255-62.
37. **Campbell, J. D., M. J. Stinson, F. E. Simons, E. S. Rector, and K. T. HayGlass.** 2001. In vivo stability of human chemokine and chemokine receptor expression. Hum Immunol **62:**668-78.
38. **Capoulade-Metay, C., L. Ma, L. X. Truong, Y. Dudoit, P. Versmisse, N. V. Nguyen, M. Nguyen, D. Scott-Algara, F. Barre-Sinoussi, P. Debre, G. Bismuth, G. Pancino, and I. Theodorou.** 2004. New CCR5 variants associated with reduced HIV coreceptor function in southeast Asia. AIDS **18:**2243-52.
39. **Cardno, A. G., F. V. Rijsdijk, P. C. Sham, R. M. Murray, and P. McGuffin.** 2002. A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry **159:**539-45.
40. **Cardone, M. F., Z. Jiang, P. D'Addabbo, N. Archidiacono, M. Rocchi, E. E. Eichler, and M. Ventura.** 2008. Hominoid chromosomal rearrangements on 17q map to complex regions of segmental duplication. Genome Biol **9:**R28.
41. **Carrington, M., M. Dean, M. P. Martin, and S. J. O'Brien.** 1999. Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Hum Mol Genet **8:**1939-45.
42. **Carrington, M., T. Kissner, B. Gerrard, S. Ivanov, S. J. O'Brien, and M. Dean.** 1997. Novel alleles of the chemokine-receptor gene CCR5. Am J Hum Genet **61:**1261-7.
43. **Choe, H., M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P. D. Ponath, L. Wu, C. R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard, and J. Sodroski.** 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell **85:**1135-48.
44. **Chong, S., and E. Whitelaw.** 2004. Epigenetic germline inheritance. Curr Opin Genet Dev **14:**692-6.
45. **Clegg, A. O., L. J. Ashton, R. A. Biti, P. Badhwar, P. Williamson, J. M. Kaldor, and G. J. Stewart.** 2000. CCR5 promoter polymorphisms, CCR5 59029A and CCR5 59353C, are under represented in HIV-1-infected long-term non-progressors. The Australian Long-Term Non-Progressor Study Group. AIDS **14:**103-8.
46. **Clerici, M., J. M. Levin, H. A. Kessler, A. Harris, J. A. Berzofsky, A. L. Landay, and G. M. Shearer.** 1994. HIV-specific T-helper activity in seronegative health care workers exposed to contaminated blood. JAMA **271:**42-6.
47. **Cohen, M. S., N. Hellmann, J. A. Levy, K. DeCock, and J. Lange.** 2008. The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. J Clin Invest **118:**1244-54.
48. **Colin, L., and C. Van Lint.** 2009. Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. Retrovirology **6:**111.
49. **Colobran, R., P. Adreani, Y. Ashhab, A. Llano, J. A. Este, O. Dominguez, R. Pujol-Borrell, and M. Juan.** 2005. Multiple products derived from two CCL4 loci: high incidence of a new polymorphism in HIV+ patients. J Immunol **174:**5655-64.
50. **Colobran, R., N. Casamitjana, A. Roman, R. Faner, E. Pedrosa, J. I. Arostegui, R. Pujol-Borrell, M. Juan, and E. Palou.** 2009. Copy number variation in the CCL4L gene is associated with susceptibility to acute rejection in lung transplantation. Genes Immun **10:**254-9.
51. **Colobran, R., D. Comas, R. Faner, E. Pedrosa, R. Anglada, R. Pujol-Borrell, J. Bertranpetit, and M. Juan.** 2008. Population structure in copy number variation and SNPs in the CCL4L chemokine gene. Genes Immun **9:**279-88.
52. **Colobran, R., E. Pedrosa, L. Carretero-Iglesia, and M. Juan.** 2010. Copy number variation in chemokine superfamily: the complex scene of CCL3L-CCL4L genes in health and disease. Clin Exp Immunol.
53. **Conrad, D. F., and M. E. Hurles.** 2007. The population genetics of structural variation. Nat Genet **39:**S30-6.
54. **Corwin, A. L., J. G. Olson, M. A. Omar, A. Razaki, and D. M. Watts.** 1991. HIV-1 in Somalia: prevalence and knowledge among prostitutes. AIDS **5:**902-4.
55. **Cutbush, M., and P. L. Mollison.** 1950. The Duffy blood group system. Heredity **4:**383-9.

56. **Dalgleish, A. G., P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss.** 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature **312:**763-7.
57. **Davenport, R. D.** 2009. An introduction to chemokines and their roles in transfusion medicine. Vox Sang **96:**183-98.
58. **de Roda Husman, A. M., H. Blaak, M. Brouwer, and H. Schuitemaker.** 1999. CC chemokine receptor 5 cell-surface expression in relation to CC chemokine receptor 5 genotype and the clinical course of HIV-1 infection. J Immunol **163:**4597-603.
59. **Dean, M., M. Carrington, C. Winkler, G. A. Huttley, M. W. Smith, R. Allikmets, J. J. Goedert, S. P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels, and S. J. O'Brien.** 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:**1856-62.
60. **Deeks, S. G., and B. D. Walker.** 2007. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. Immunity **27:**406-16.
61. **Degenhardt, J. D., P. de Candia, A. Chabot, S. Schwartz, L. Henderson, B. Ling, M. Hunter, Z. Jiang, R. E. Palermo, M. Katze, E. E. Eichler, M. Ventura, J. Rogers, P. Marx, Y. Gilad, and C. D. Bustamante.** 2009. Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). PLoS Genet **5:**e1000346.
62. **Delvaeye, M., and E. M. Conway.** 2009. Coagulation and innate immune responses: can we view them separately? Blood **114:**2367-74.
63. **Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, and N. R. Landau.** 1996. Identification of a major co-receptor for primary isolates of HIV-1. Nature **381:**661-6.
64. **Detels, R., Z. Liu, K. Hennessey, J. Kan, B. R. Visscher, J. M. Taylor, D. R. Hoover, C. R. Rinaldo, Jr., J. P. Phair, A. J. Saah, and *et al*.** 1994. Resistance to HIV-1 infection. Multicenter AIDS Cohort Study. J Acquir Immune Defic Syndr **7:**1263-9.
65. **Dolan, M. J., M. Clerici, S. P. Blatt, C. W. Hendrix, G. P. Melcher, R. N. Boswell, T. M. Freeman, W. Ward, R. Hensley, and G. M. Shearer.** 1995. In vitro T cell function, delayed-type hypersensitivity skin testing, and CD4+ T cell subset phenotyping independently predict survival time in patients infected with human immunodeficiency virus. J Infect Dis **172:**79-87.
66. **Dolan, M. J., H. Kulkarni, J. F. Camargo, W. He, A. Smith, J. M. Anaya, T. Miura, F. M. Hecht, M. Mamtani, F. Pereyra, V. Marconi, A. Mangano, L. Sen, R. Bologna, R. A. Clark, S. A. Anderson, J. Delmar, R. J. O'Connell, A. Lloyd, J. Martin, S. S. Ahuja, B. K. Agan, B. D. Walker, S. G. Deeks, and S. K. Ahuja.** 2007. CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. Nat Immunol **8:**1324-36.
67. **Doranz, B. J., J. Rucker, Y. Yi, R. J. Smyth, M. Samson, S. C. Peiper, M. Parmentier, R. G. Collman, and R. W. Doms.** 1996. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. Cell **85:**1149-58.
68. **Dragic, T., V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore, and W. A. Paxton.** 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature **381:**667-73.
69. **Dupont, J. M., J. Tost, H. Jammes, and I. G. Gut.** 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal Biochem **333:**119-27.
70. **Egger, G., G. Liang, A. Aparicio, and P. A. Jones.** 2004. Epigenetics in human disease and prospects for epigenetic therapy. Nature **429:**457-63.
71. **Esteller, M.** 2005. DNA methylation and cancer therapy: new developments and expectations. Curr Opin Oncol **17:**55-60.

72. **Esteller, M.** 2008. Epigenetics in cancer. N Engl J Med **358:**1148-59.
73. **Fakhrai-Rad, H., N. Pourmand, and M. Ronaghi.** 2002. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Hum Mutat **19:**479-85.
74. **Fang, J. Y., J. A. Mikovits, R. Bagni, C. L. Petrow-Sadowski, and F. W. Ruscetti.** 2001. Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases de novo methylation. J Virol **75:**9753-61.
75. **Fellay, J., K. V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J. M. Miro, N. Obel, J. P. Smith, J. Wyniger, P. Descombes, S. E. Antonarakis, N. L. Letvin, A. J. McMichael, B. F. Haynes, A. Telenti, and D. B. Goldstein.** 2007. A whole-genome association study of major determinants for host control of HIV-1. Science **317:**944-7.
76. **Fellermann, K., D. E. Stange, E. Schaeffeler, H. Schmalzl, J. Wehkamp, C. L. Bevins, W. Reinisch, A. Teml, M. Schwab, P. Lichter, B. Radlwimmer, and E. F. Stange.** 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet **79:**439-48.
77. **Felsenfeld, G., and M. Groudine.** 2003. Controlling the double helix. Nature **421:**448-53.
78. **Feng, Y., C. C. Broder, P. E. Kennedy, and E. A. Berger.** 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science **272:**872-7.
79. **Field, S. F., J. M. Howson, L. M. Maier, S. Walker, N. M. Walker, D. J. Smyth, J. A. Armour, D. G. Clayton, and J. A. Todd.** 2009. Experimental aspects of copy number variant assays at CCL3L1. Nat Med **15:**1115-7.
80. **Filion, G. J., S. Zhenilo, S. Salozhin, D. Yamada, E. Prokhortchouk, and P. A. Defossez.** 2006. A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol **26:**169-81.
81. **Foley, D. L., J. M. Craig, R. Morley, C. A. Olsson, T. Dwyer, K. Smith, and R. Saffery.** 2009. Prospects for epigenetic epidemiology. Am J Epidemiol **169:**389-400.
82. **Fowke, K. R., N. J. Nagelkerke, J. Kimani, J. N. Simonsen, A. O. Anzala, J. J. Bwayo, K. S. MacDonald, E. N. Ngugi, and F. A. Plummer.** 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet **348:**1347-51.
83. **Fox, R. J., P. Kivisakk, J. C. Lee, B. Tucky, C. Lucchinetti, R. A. Rudick, and R. M. Ransohoff.** 2006. Chemokine receptors as biomarkers in multiple sclerosis. Dis Markers **22:**227-33.
84. **Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y. Z. Wu, C. Plass, and M. Esteller.** 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A **102:**10604-9.
85. **Frank, B., J. L. Bermejo, K. Hemminki, C. Sutter, B. Wappenschmidt, A. Meindl, M. Kiechle-Bahat, P. Bugert, R. K. Schmutzler, C. R. Bartram, and B. Burwinkel.** 2007. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. Carcinogenesis **28:**1442-5.
86. **Freedman, D. S., L. Gates, W. D. Flanders, O. W. Van Assendelft, J. J. Barboriak, M. R. Joesoef, and T. Byers.** 1997. Black/white differences in leukocyte subpopulations in men. Int J Epidemiol **26:**757-64.
87. **Freiman, R. N., and R. Tjian.** 2003. Regulating the regulators: lysine modifications make their mark. Cell **112:**11-7.
88. **Fu, W., B. E. Sanders-Beer, K. S. Katz, D. R. Maglott, K. D. Pruitt, and R. G. Ptak.** 2009. Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic Acids Res **37:**D417-22.

89. **Gartner, K.** 1990. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? Lab Anim **24:**71-7.
90. **Gervaix, A., J. Nicolas, P. Portales, K. Posfay-Barbe, C. A. Wyler, M. Segondy, O. Avinens, C. A. Siegrist, J. Clot, J. F. Eliaou, J. Astruc, and P. Corbeau.** 2002. Response to treatment and disease progression linked to CD4+ T cell surface CC chemokine receptor 5 density in human immunodeficiency virus type 1 vertical infection. J Infect Dis **185:**1055-61.
91. **Gilbert, P. B., V. G. DeGruttola, M. G. Hudgens, S. G. Self, S. M. Hammer, and L. Corey.** 2003. What constitutes efficacy for a human immunodeficiency virus vaccine that ameliorates viremia: issues involving surrogate end points in phase 3 trials. J Infect Dis **188:**179-93.
92. **Gius, D., H. Cui, C. M. Bradbury, J. Cook, D. K. Smart, S. Zhao, L. Young, S. A. Brandenburg, Y. Hu, K. S. Bisht, A. S. Ho, D. Mattson, L. Sun, P. J. Munson, E. Y. Chuang, J. B. Mitchell, and A. P. Feinberg.** 2004. Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. Cancer Cell **6:**361-71.
93. **Glass, W. G., D. H. McDermott, J. K. Lim, S. Lekhong, S. F. Yu, W. A. Frank, J. Pape, R. C. Cheshier, and P. M. Murphy.** 2006. CCR5 deficiency increases risk of symptomatic West Nile virus infection. J Exp Med **203:**35-40.
94. **Glynn, J. R., M. Carael, B. Auvert, M. Kahindo, J. Chege, R. Musonda, F. Kaona, and A. Buve.** 2001. Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia. AIDS **15 Suppl 4:**S51-60.
95. **Gomes, M. V., and R. A. Waterland.** 2008. Individual epigenetic variation: when, why, and so what? Nestle Nutr Workshop Ser Pediatr Program **62:**141-50; discussion 151-5.
96. **Gonzalez, E., M. Bamshad, N. Sato, S. Mummidi, R. Dhanda, G. Catano, S. Cabrera, M. McBride, X. H. Cao, G. Merrill, P. O'Connell, D. W. Bowden, B. I. Freedman, S. A. Anderson, E. A. Walter, J. S. Evans, K. T. Stephan, R. A. Clark, S. Tyagi, S. S. Ahuja, M. J. Dolan, and S. K. Ahuja.** 1999. Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. Proc Natl Acad Sci U S A **96:**12004-9.
97. **Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A. Clark, S. A. Anderson, J. O'Connell R, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja.** 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science **307:**1434-40.
98. **Gopalakrishnan, S., B. O. Van Emburgh, and K. D. Robertson.** 2008. DNA methylation in development and human disease. Mutat Res **647:**30-8.
99. **Gornalusse, G., A. A. Gaitan, S. Mummidi, A. M. Murillo, V. Ramsuran, W. He, S. Valera, C. Ratliff, E. A. Walter, Z. A. Chykarenko, R. Hutt, D. Daskalakis, L. Shostakovich-Koretskaya, S. A. karim, J. Martin, S. G. Deeks, F. M. Hecht, E. Sinclair, A. R. ClarK, T. Ndung'u, F. T. Valentine, P. W. Hunt, and S. K. Ahuja.** 2011. DNA methylation regulates T-cell expression levels of the major HIV co-receptor CCR5 and associates with disease phenotypes. Nat Immunol **Submitted for publication**.
100. **Gornalusse, G., S. Mummidi, W. He, G. Silvestri, M. Bamshad, and S. K. Ahuja.** 2009. CCL3L Copy number variation and the co-evolution of primate and viral genomes. PLoS Genet **5:**e1000359.
101. **Gosden, R. G., and A. P. Feinberg.** 2007. Genetics and epigenetics--nature's pen-and-pencil set. N Engl J Med **356:**731-3.
102. **Granelli-Piperno, A., B. Moser, M. Pope, D. Chen, Y. Wei, F. Isdell, U. O'Doherty, W. Paxton, R. Koup, S. Mojsov, N. Bhardwaj, I. Clark-Lewis, M. Baggiolini, and R. M. Steinman.** 1996. Efficient interaction of HIV-1 with purified dendritic cells via multiple chemokine coreceptors. J Exp Med **184:**2433-8.

103. **Gray, R. H., M. J. Wawer, R. Brookmeyer, N. K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. vanCott, and T. C. Quinn.** 2001. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. Lancet **357:**1149-53.
104. **Grunau, C., S. J. Clark, and A. Rosenthal.** 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res **29:**E65-5.
105. **Grunhage, F., J. Nattermann, O. A. Gressner, H. E. Wasmuth, C. Hellerbrand, T. Sauerbruch, U. Spengler, and F. Lammert.** 2010. Lower copy numbers of the chemokine CCL3L1 gene in patients with chronic hepatitis C. J Hepatol **52:**153-159.
106. **Guerin, S., L. Meyer, I. Theodorou, F. Boufassa, M. Magierowska, C. Goujard, C. Rouzioux, P. Debre, and J. F. Delfraissy.** 2000. CCR5 delta32 deletion and response to highly active antiretroviral therapy in HIV-1-infected patients. Aids **14:**2788-90.
107. **Gutekunst, K. A., F. Kashanchi, J. N. Brady, and D. P. Bednarik.** 1993. Transcription of the HIV-1 LTR is regulated by the density of DNA CpG methylation. J Acquir Immune Defic Syndr **6:**541-9.
108. **Haddy, T. B., S. R. Rana, and O. Castro.** 1999. Benign ethnic neutropenia: what is a normal absolute neutrophil count? J Lab Clin Med **133:**15-22.
109. **Hadley, T. J., and S. C. Peiper.** 1997. From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. Blood **89:**3077-91.
110. **Hanash, S.** 2004. Integrated global profiling of cancer. Nat Rev Cancer **4:**638-44.
111. **Hatchwell, E., and J. M. Greally.** 2007. The potential role of epigenomic dysregulation in complex human disease. Trends Genet **23:**588-95.
112. **He, W., H. Kulkarni, J. Castiblanco, C. Shimizu, U. Aluyen, R. Maldonado, A. Carrillo, M. Griffin, A. Lipsitt, L. Beachy, L. Shostakovich-Koretskaya, A. Mangano, L. Sen, R. J. Nibbs, C. T. Tiemessen, H. Bolivar, M. J. Bamshad, R. A. Clark, J. C. Burns, M. J. Dolan, and S. K. Ahuja.** 2009. Reply to: "Experimental aspects of copy number variant assays at CCL3L1". Nat Med **15:**1117-20.
113. **He, W., V. Marconi, J. Castiblanco, H. Kulkarni, R. A. Clark, M. J. Dolan, R. A. Weiss, and S. K. Ahuja.** 2009. Response: Association of Duffy Antigen Genotypes with HIV-AIDS Susceptibility. Cell Host Microbe **5:**418-419.
114. **He, W., S. Neil, H. Kulkarni, E. Wright, B. K. Agan, V. C. Marconi, M. J. Dolan, R. A. Weiss, and S. K. Ahuja.** 2008. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. Cell Host Microbe **4:**52-62.
115. **Heijmans, B. T., D. Kremer, E. W. Tobi, D. I. Boomsma, and P. E. Slagboom.** 2007. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum Mol Genet **16:**547-54.
116. **Hejnar, J., P. Hajkova, J. Plachy, D. Elleder, V. Stepanets, and J. Svoboda.** 2001. CpG island protects Rous sarcoma virus-derived vectors integrated into nonpermissive cells from DNA methylation and transcriptional suppression. Proc Natl Acad Sci U S A **98:**565-9.
117. **Hendrickson, S. L., L. P. Jacobson, G. W. Nelson, J. P. Phair, J. Lautenberger, R. C. Johnson, L. Kingsley, J. B. Margolick, R. Detels, J. J. Goedert, and S. J. O'Brien.** 2008. Host genetic influences on highly active antiretroviral therapy efficacy and AIDS-free survival. J Acquir Immune Defic Syndr **48:**263-71.
118. **Henry, W. K., P. Tebas, and H. C. Lane.** 2006. Explaining, predicting, and treating HIV-associated CD4 cell loss: after 25 years still a puzzle. Jama **296:**1523-5.
119. **Heredia, A., B. Gilliam, A. DeVico, N. Le, D. Bamba, R. Flinko, G. Lewis, R. C. Gallo, and R. R. Redfield.** 2007. CCR5 density levels on primary CD4 T cells impact the replication and Enfuvirtide susceptibility of R5 HIV-1. AIDS **21:**1317-22.

120. **Heredia, A., O. Latinovic, R. C. Gallo, G. Melikyan, M. Reitz, N. Le, and R. R. Redfield.** 2008. Reduction of CCR5 with low-dose rapamycin enhances the antiviral activity of vicriviroc against both sensitive and drug-resistant HIV-1. Proc Natl Acad Sci U S A **105:**20476-81.
121. **Hermann, A., S. Schmitt, and A. Jeltsch.** 2003. The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. J Biol Chem **278:**31717-21.
122. **Hess, C., T. Klimkait, L. Schlapbach, V. Del Zenero, S. Sadallah, E. Horakova, G. Balestra, V. Werder, C. Schaefer, M. Battegay, and J. A. Schifferli.** 2002. Association of a pool of HIV-1 with erythrocytes in vivo: a cohort study. Lancet **359:**2230-4.
123. **Hladik, F., H. Liu, E. Speelmon, D. Livingston-Rosanoff, S. Wilson, P. Sakchalathorn, Y. Hwangbo, B. Greene, T. Zhu, and M. J. McElrath.** 2005. Combined effect of CCR5-Delta32 heterozygosity and the CCR5 promoter polymorphism -2459 A/G on CCR5 expression and resistance to human immunodeficiency virus type 1 transmission. J Virol **79:**11677-84.
124. **Holliday, R.** 1990. Mechanisms for the control of gene activity during development. Biol Rev Camb Philos Soc **65:**431-71.
125. **Holliday, R., and J. E. Pugh.** 1975. DNA modification mechanisms and gene activity during development. Science **187:**226-32.
126. **Hollox, E. J., J. A. Armour, and J. C. Barber.** 2003. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. Am J Hum Genet **73:**591-600.
127. **Hollox, E. J., U. Huffmeier, P. L. Zeeuwen, R. Palla, J. Lascorz, D. Rodijk-Olthuis, P. C. van de Kerkhof, H. Traupe, G. de Jongh, M. den Heijer, A. Reis, J. A. Armour, and J. Schalkwijk.** 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet **40:**23-5.
128. **Horakova, E., O. Gasser, S. Sadallah, J. M. Inal, G. Bourgeois, I. Ziekau, T. Klimkait, and J. A. Schifferli.** 2004. Complement mediates the binding of HIV to erythrocytes. J Immunol **173:**4236-41.
129. **Horne, K. C., X. Li, L. P. Jacobson, F. Palella, B. D. Jamieson, J. B. Margolick, J. Martinson, V. Turkozu, K. Visvanathan, and I. J. Woolley.** 2009. Duffy antigen polymorphisms do not alter progression of HIV in African Americans in the MACS cohort. Cell Host Microbe **5:**415-7; author reply 418-9.
130. **Horuk, R.** 1999. Chemokine receptors and HIV-1: the fusion of two major research fields. Immunol Today **20:**89-94.
131. **Horuk, R.** 1994. The interleukin-8-receptor family: from chemokines to malaria. Immunol Today **15:**169-74.
132. **Hsieh, M. M., J. E. Everhart, D. D. Byrd-Holt, J. F. Tisdale, and G. P. Rodgers.** 2007. Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. Ann Intern Med **146:**486-92.
133. **Huik, K., M. Sadam, T. Karki, R. Avi, T. Krispin, P. Paap, K. Ruutel, A. Uuskula, A. Talu, K. Abel-Ollo, and I. Lutsar.** 2010. CCL3L1 copy number is a strong genetic determinant of HIV seropositivity in Caucasian intravenous drug users. J Infect Dis **201:**730-9.
134. **Hummel, S., D. Schmidt, B. Kremeyer, B. Herrmann, and M. Oppermann.** 2005. Detection of the CCR5-Delta32 HIV resistance gene in Bronze Age skeletons. Genes Immun **6:**371-4.
135. **Irving, S. G., P. F. Zipfel, J. Balke, O. W. McBride, C. C. Morton, P. R. Burd, U. Siebenlist, and K. Kelly.** 1990. Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. Nucleic Acids Res **18:**3261-70.
136. **Jaenisch, R., and A. Bird.** 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet **33 Suppl:**245-54.
137. **Jahner, D., H. Stuhlmann, C. L. Stewart, K. Harbers, J. Lohler, I. Simon, and R. Jaenisch.** 1982. De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature **298:**623-8.
138. **Jain, N., A. Rossi, and G. Garcia-Manero.** 2009. Epigenetic therapy of leukemia: An update. Int J Biochem Cell Biol **41:**72-80.

139. **Jones, P. A., S. M. Taylor, T. Mohandas, and L. J. Shapiro.** 1982. Cell cycle-specific reactivation of an inactive X-chromosome locus by 5-azadeoxycytidine. Proc Natl Acad Sci U S A **79:**1215-9.
140. **Julg, B., S. Reddy, M. van der Stok, S. Kulkarni, Y. Qi, S. Bass, B. Gold, M. A. Nalls, G. W. Nelson, B. D. Walker, M. Carrington, and T. Ndung'u.** 2009. Lack of Duffy antigen receptor for chemokines: no influence on HIV disease progression in an African treatment-naive population. Cell Host Microbe **5:**413-5; author reply 418-9.
141. **Kaminsky, Z. A., T. Tang, S. C. Wang, C. Ptak, G. H. Oh, A. H. Wong, L. A. Feldcamp, C. Virtanen, J. Halfvarson, C. Tysk, A. F. McRae, P. M. Visscher, G. W. Montgomery, Gottesman, II, N. G. Martin, and A. Petronis.** 2009. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet **41:**240-5.
142. **Karim, Q. A., S. S. Karim, K. Soldan, and M. Zondi.** 1995. Reducing the risk of HIV infection among South African sex workers: socioeconomic and gender barriers. Am J Public Health **85:**1521-5.
143. **Karpus, W. J., N. W. Lukacs, K. J. Kennedy, W. S. Smith, S. D. Hurst, and T. A. Barrett.** 1997. Differential CC chemokine-induced enhancement of T helper cell cytokine production. J Immunol **158:**4129-36.
144. **Kasten, S., A. Goldwich, M. Schmitt, A. Rascu, M. Grunke, C. Dechant, J. R. Kalden, and T. Harrer.** 2000. Positive influence of the Delta32CCR5 allele on response to highly active antiretroviral therapy (HAART) in HIV-1 infected patients. Eur J Med Res **5:**323-8.
145. **Kauder, S. E., A. Bosque, A. Lindqvist, V. Planelles, and E. Verdin.** 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog **5:**e1000495.
146. **Kaur, G., and N. Mehra.** 2009. Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection. Tissue Antigens **73:**289-301.
147. **Kaur, G., P. Singh, C. C. Rapthap, N. Kumar, M. Vajpayee, S. K. Sharma, A. Wanchu, and N. K. Mehra.** 2007. Polymorphism in the CCR5 gene promoter and HIV-1 infection in North Indians. Hum Immunol **68:**454-61.
148. **Kim, A., M. Pettoello-Mantovani, and H. Goldstein.** 1998. Decreased susceptibility of peripheral blood mononuclear cells from individuals heterozygous for a mutant CCR5 allele to HIV infection. J Acquir Immune Defic Syndr Hum Retrovirol **19:**145-9.
149. **Kivisakk, P., C. Trebst, J. C. Lee, B. H. Tucky, R. A. Rudick, J. J. Campbell, and R. M. Ransohoff.** 2003. Expression of CCR2, CCR5, and CXCR3 by CD4+ T cells is stable during a 2-year longitudinal study but varies widely between individuals. J Neurovirol **9:**291-9.
150. **Klose, R. J., and A. P. Bird.** 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci **31:**89-97.
151. **Knight, S. C., S. E. Macatonia, and S. Patterson.** 1990. HIV I infection of dendritic cells. Int Rev Immunol **6:**163-75.
152. **Kondrashov, A. S.** 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat **21:**12-27.
153. **Konig, R., Y. Zhou, D. Elleder, T. L. Diamond, G. M. Bonamy, J. T. Irelan, C. Y. Chiang, B. P. Tu, P. D. De Jesus, C. E. Lilley, S. Seidel, A. M. Opaluch, J. S. Caldwell, M. D. Weitzman, K. L. Kuhen, S. Bandyopadhyay, T. Ideker, A. P. Orth, L. J. Miraglia, F. D. Bushman, J. A. Young, and S. K. Chanda.** 2008. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell **135:**49-60.
154. **Kosoy, R., R. Nassir, C. Tian, P. A. White, L. M. Butler, G. Silva, R. Kittles, M. E. Alarcon-Riquelme, P. K. Gregersen, J. W. Belmont, F. M. De La Vega, and M. F. Seldin.** 2009. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat **30:**69-78.
155. **Kuhmann, S. E., E. J. Platt, S. L. Kozak, and D. Kabat.** 2000. Cooperation of multiple CCR5 coreceptors is required for infections by human immunodeficiency virus type 1. J Virol **74:**7005-15.

156. **Kuhn, L., D. B. Schramm, S. Donninger, S. Meddows-Taylor, A. H. Coovadia, G. G. Sherman, G. E. Gray, and C. T. Tiemessen.** 2007. African infants' CCL3 gene copies influence perinatal HIV transmission in the absence of maternal nevirapine. AIDS **21:**1753-61.
157. **Kulkarni, H., V. C. Marconi, B. K. Agan, C. McArthur, G. Crawford, R. A. Clark, M. J. Dolan, and S. K. Ahuja.** 2008. Role of CCL3L1-CCR5 genotypes in the epidemic spread of HIV-1 and evaluation of vaccine efficacy. PLoS One **3:**e3671.
158. **Kulkarni, H., V. C. Marconi, W. He, M. L. Landrum, J. F. Okulicz, J. Delmar, D. Kazandjian, J. Castiblanco, S. S. Ahuja, E. J. Wright, R. A. Weiss, R. A. Clark, M. J. Dolan, and S. K. Ahuja.** 2009. The Duffy-null state is associated with a survival advantage in leukopenic HIV-infected persons of African ancestry. Blood **114:**2570-1.
159. **Kuller, L. H., R. Tracy, W. Belloso, S. De Wit, F. Drummond, H. C. Lane, B. Ledergerber, J. Lundgren, J. Neuhaus, D. Nixon, N. I. Paton, and J. D. Neaton.** 2008. Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med **5:**e203.
160. **Kurkjian, C., S. Kummar, and A. J. Murgo.** 2008. DNA methylation: its role in cancer development and therapy. Curr Probl Cancer **32:**187-235.
161. **Lachgar, A., G. Jaureguiberry, H. Le Buenac, B. Bizzini, J. F. Zagury, J. Rappaport, and D. Zagury.** 1998. Binding of HIV-1 to RBCs involves the Duffy antigen receptors for chemokines (DARC). Biomed Pharmacother **52:**436-9.
162. **Laga, M., B. Schwartlander, E. Pisani, P. S. Sow, and M. Carael.** 2001. To stem HIV in Africa, prevent transmission to young women. AIDS **15:**931-4.
163. **Laing, K. J., and C. J. Secombes.** 2004. Chemokines. Dev Comp Immunol **28:**443-60.
164. **Laird, P. W.** 2005. Cancer epigenetics. Hum Mol Genet **14 Spec No 1:**R65-76.
165. **Lama, J., and V. Planelles.** 2007. Host factors influencing susceptibility to HIV infection and AIDS progression. Retrovirology **4:**52.
166. **Landau, N. R., M. Warton, and D. R. Littman.** 1988. The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4. Nature **334:**159-62.
167. **Langlade-Demoyen, P., N. Ngo-Giang-Huong, F. Ferchal, and E. Oksenhendler.** 1994. Human immunodeficiency virus (HIV) nef-specific cytotoxic T lymphocytes in noninfected heterosexual contact of HIV-infected patients. J Clin Invest **93:**1293-7.
168. **Lavie, L., M. Kitova, E. Maldener, E. Meese, and J. Mayer.** 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). J Virol **79:**876-83.
169. **Lederman, M. M., A. Penn-Nicholson, M. Cho, and D. Mosier.** 2006. Biology of CCR5 and its role in HIV infection and treatment. JAMA **296:**815-26.
170. **Lee, B., B. J. Doranz, S. Rana, Y. Yi, M. Mellado, J. M. Frade, A. C. Martinez, S. J. O'Brien, M. Dean, R. G. Collman, and R. W. Doms.** 1998. Influence of the CCR2-V64I polymorphism on human immunodeficiency virus type 1 coreceptor activity and on chemokine receptor function of CCR2b, CCR3, CCR5, and CXCR4. J Virol **72:**7450-8.
171. **Lelievre, J. D., F. Petit, L. Perrin, F. Mammano, D. Arnoult, J. C. Ameisen, J. Corbeil, A. Gervaix, and J. Estaquier.** 2004. The density of coreceptors at the surface of CD4+ T cells contributes to the extent of human immunodeficiency virus type 1 viral replication-mediated T cell death. AIDS Res Hum Retroviruses **20:**1230-43.
172. **Levy, J. A.** 2002. HIV-1: hitching a ride on erythrocytes. Lancet **359:**2212-3.
173. **Liang, G., F. A. Gonzales, P. A. Jones, T. F. Orntoft, and T. Thykjaer.** 2002. Analysis of gene induction in human fibroblasts and bladder cancer cells exposed to the methylation inhibitor 5-aza-2'-deoxycytidine. Cancer Res **62:**961-6.
174. **Lim, H. G., K. Suzuki, D. A. Cooper, and A. D. Kelleher.** 2008. Promoter-targeted siRNAs induce gene silencing of simian immunodeficiency virus (SIV) infection in vitro. Mol Ther **16:**565-70.

175. **Lin, Y. L., C. Mettling, P. Portales, B. Reant, V. Robert-Hebmann, J. Reynes, J. Clot, and P. Corbeau.** 2006. The efficiency of R5 HIV-1 infection is determined by CD4 T-cell surface CCR5 density through G alpha i-protein signalling. Aids **20:**1369-77.
176. **Lin, Y. L., C. Mettling, P. Portales, J. Reynes, J. Clot, and P. Corbeau.** 2002. Cell surface CCR5 density determines the postentry efficiency of R5 HIV-1 infection. Proc Natl Acad Sci U S A **99:**15590-5.
177. **Ling, B., R. S. Veazey, A. Luckay, C. Penedo, K. Xu, J. D. Lifson, and P. A. Marx.** 2002. SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. AIDS **16:**1489-96.
178. **Liu, L., Y. Li, and T. O. Tollefsbol.** 2008. Gene-environment interactions and epigenetic basis of human diseases. Curr Issues Mol Biol **10:**25-36.
179. **Liu, R., W. A. Paxton, S. Choe, D. Ceradini, S. R. Martin, R. Horuk, M. E. MacDonald, H. Stuhlmann, R. A. Koup, and N. R. Landau.** 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell **86:**367-77.
180. **Loetscher, P., M. Uguccioni, L. Bordoli, M. Baggiolini, B. Moser, C. Chizzolini, and J. M. Dayer.** 1998. CCR5 is characteristic of Th1 lymphocytes. Nature **391:**344-5.
181. **Lopez, C., M. Comabella, M. Tintore, J. Sastre-Garriga, and X. Montalban.** 2006. Variations in chemokine receptor and cytokine expression during pregnancy in multiple sclerosis patients. Mult Scler **12:**421-7.
182. **Louisirirotchanakul, S., H. Liu, A. Roongpisuthipong, E. E. Nakayama, Y. Takebe, T. Shioda, and C. Wasi.** 2002. Genetic analysis of HIV-1 discordant couples in Thailand: association of CCR2 64I homozygosity with HIV-1-negative status. J Acquir Immune Defic Syndr **29:**314-5.
183. **Lupski, J. R.** 2007. Genomic rearrangements and sporadic disease. Nat Genet **39:**S43-7.
184. **Luster, A. D., and M. E. Rothenberg.** 1997. Role of the monocyte chemoattractant protein and eotaxin subfamily of chemokines in allergic inflammation. J Leukoc Biol **62:**620-33.
185. **Luther, S. A., and J. G. Cyster.** 2001. Chemokines as regulators of T cell differentiation. Nat Immunol **2:**102-7.
186. **Mackay, C. R.** 2005. CCL3L1 dose and HIV-1 susceptibility. Trends Mol Med **11:**203-6.
187. **Maddon, P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and R. Axel.** 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell **47:**333-48.
188. **Mamtani, M., J. M. Anaya, W. He, and S. K. Ahuja.** Association of copy number variation in the FCGR3B gene with risk of autoimmune diseases. Genes Immun **11:**155-60.
189. **Marmor, M., K. Hertzmark, S. M. Thomas, P. N. Halkitis, and M. Vogler.** 2006. Resistance to HIV infection. J Urban Health **83:**5-17.
190. **Mars, W. M., P. Patmasiriwat, T. Maity, V. Huff, M. M. Weil, and G. F. Saunders.** 1995. Inheritance of unequal numbers of the genes encoding the human neutrophil defensins HP-1 and HP-3. J Biol Chem **270:**30371-6.
191. **Martin, M. P., M. Dean, M. W. Smith, C. Winkler, B. Gerrard, N. L. Michael, B. Lee, R. W. Doms, J. Margolick, S. Buchbinder, J. J. Goedert, T. R. O'Brien, M. W. Hilgartner, D. Vlahov, S. J. O'Brien, and M. Carrington.** 1998. Genetic acceleration of AIDS progression by a promoter variant of CCR5. Science **282:**1907-11.
192. **Martin, M. P., M. M. Lederman, H. B. Hutcheson, J. J. Goedert, G. W. Nelson, Y. van Kooyk, R. Detels, S. Buchbinder, K. Hoots, D. Vlahov, S. J. O'Brien, and M. Carrington.** 2004. Association of DC-SIGN promoter polymorphism with increased risk for parenteral, but not mucosal, acquisition of human immunodeficiency virus type 1 infection. J Virol **78:**14053-6.
193. **Matouskova, M., J. Blazkova, P. Pajer, A. Pavlicek, and J. Hejnar.** 2006. CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues. Exp Cell Res **312:**1011-20.

194. **McCarroll, S. A., and D. M. Altshuler.** 2007. Copy-number variation and association studies of human disease. Nat Genet **39:**S37-42.
195. **McGowan, I., J. Elliott, G. Cortina, K. Tanner, C. Siboliban, A. Adler, D. Cho, W. J. Boscardin, L. Soto-Torres, and P. A. Anton.** 2007. Characterization of baseline intestinal mucosal indices of injury and inflammation in men for use in rectal microbicide trials (HIV Prevention Trials Network-056). J Acquir Immune Defic Syndr **46:**417-25.
196. **McKinney, C., M. E. Merriman, P. T. Chapman, P. J. Gow, A. A. Harrison, J. Highton, P. B. Jones, L. McLean, J. L. O'Donnell, V. Pokorny, M. Spellerberg, L. K. Stamp, J. Willis, S. Steer, and T. R. Merriman.** 2008. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Annual Rheum Dis. **67:**409-413.
197. **McNicholl, J. M., D. K. Smith, S. H. Qari, and T. Hodge.** 1997. Host genes and HIV: the role of the chemokine receptor gene CCR5 and its allele. Emerg Infect Dis **3:**261-71.
198. **Meddows-Taylor, S., S. L. Donninger, M. Paximadis, D. B. Schramm, F. S. Anthony, G. E. Gray, L. Kuhn, and C. T. Tiemessen.** 2006. Reduced ability of newborns to produce CCL3 is associated with increased susceptibility to perinatal human immunodeficiency virus 1 transmission. J Gen Virol **87:**2055-65.
199. **Melzer, D., J. R. Perry, D. Hernandez, A. M. Corsi, K. Stevens, I. Rafferty, F. Lauretani, A. Murray, J. R. Gibbs, G. Paolisso, S. Rafiq, J. Simon-Sanchez, H. Lango, S. Scholz, M. N. Weedon, S. Arepalli, N. Rice, N. Washecka, A. Hurst, A. Britton, W. Henley, J. van de Leemput, R. Li, A. B. Newman, G. Tranah, T. Harris, V. Panicker, C. Dayan, A. Bennett, M. I. McCarthy, A. Ruokonen, M. R. Jarvelin, J. Guralnik, S. Bandinelli, T. M. Frayling, A. Singleton, and L. Ferrucci.** 2008. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet **4:**e1000072.
200. **Menten, P., A. Wuyts, and J. Van Damme.** 2002. Macrophage inflammatory protein-1. Cytokine Growth Factor Rev **13:**455-81.
201. **Mikovits, J. A., H. A. Young, P. Vertino, J. P. Issa, P. M. Pitha, S. Turcoski-Corrales, D. D. Taub, C. L. Petrow, S. B. Baylin, and F. W. Ruscetti.** 1998. Infection with human immunodeficiency virus type 1 upregulates DNA methyltransferase, resulting in de novo methylation of the gamma interferon (IFN-gamma) promoter and subsequent downregulation of IFN-gamma production. Mol Cell Biol **18:**5166-77.
202. **Modi, W. S.** 2004. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. Genomics **83:**735-8.
203. **Modi, W. S., J. Bergeron, and M. Sanford.** 2001. The human MIP-1beta chemokine is encoded by two paralogous genes, ACT-2 and LAG-1. Immunogenetics **53:**543-9.
204. **Molon, B., G. Gri, M. Bettella, C. Gomez-Mouton, A. Lanzavecchia, A. C. Martinez, S. Manes, and A. Viola.** 2005. T cell costimulation by chemokine receptors. Nat Immunol **6:**465-71.
205. **Morak, M., H. K. Schackert, N. Rahner, B. Betz, M. Ebert, C. Walldorf, B. Royer-Pokora, K. Schulmann, M. von Knebel-Doeberitz, W. Dietmaier, G. Keller, B. Kerker, G. Leitner, and E. Holinski-Feder.** 2008. Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. Eur J Hum Genet **16:**804-11.
206. **Morgan, H. D., W. Dean, H. A. Coker, W. Reik, and S. K. Petersen-Mahrt.** 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem **279:**52353-60.
207. **Mori, Y., J. Yin, F. Sato, A. Sterian, L. A. Simms, F. M. Selaru, K. Schulmann, Y. Xu, A. Olaru, S. Wang, E. Deacu, J. M. Abraham, J. Young, B. A. Leggett, and S. J. Meltzer.** 2004. Identification of genes uniquely involved in frequent microsatellite instability colon carcinogenesis by expression profiling combined with epigenetic scanning. Cancer Res **64:**2434-8.

208. **Mummidi, S., M. Bamshad, S. S. Ahuja, E. Gonzalez, P. M. Feuillet, K. Begum, M. C. Galvis, V. Kostecki, A. J. Valente, K. K. Murthy, L. Haro, M. J. Dolan, J. S. Allan, and S. K. Ahuja.** 2000. Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. J Biol Chem **275:**18946-61.

209. **Murooka, T. T., M. M. Wong, R. Rahbar, B. Majchrzak-Kita, A. E. Proudfoot, and E. N. Fish.** 2006. CCL5-CCR5-mediated apoptosis in T cells: Requirement for glycosaminoglycan binding and CCL5 aggregation. J Biol Chem **281:**25184-94.

210. **Murphy, P. M., M. Baggiolini, I. F. Charo, C. A. Hebert, R. Horuk, K. Matsushima, L. H. Miller, J. J. Oppenheim, and C. A. Power.** 2000. International union of pharmacology. XXII. Nomenclature for chemokine receptors. Pharmacol Rev **52:**145-76.

211. **Nachman, M. W., and S. L. Crowell.** 2000. Estimate of the mutation rate per nucleotide in humans. Genetics **156:**297-304.

212. **Nakajima, T., G. Kaur, N. Mehra, and A. Kimura.** 2008. HIV-1/AIDS susceptibility and copy number variation in CCL3L1, a gene encoding a natural ligand for HIV-1 co-receptor CCR5. Cytogenet Genome Res **123:**156-60.

213. **Nakajima, T., H. Ohtani, T. Naruse, H. Shibata, J. I. Mimaya, H. Terunuma, and A. Kimura.** 2007. Copy number variations of CCL3L1 and long-term prognosis of HIV-1 infection in asymptomatic HIV-infected Japanese with hemophilia. Immunogenetics **59:**793-8.

214. **Nalls, M. A., J. G. Wilson, N. J. Patterson, A. Tandon, J. M. Zmuda, S. Huntsman, M. Garcia, D. Hu, R. Li, B. A. Beamer, K. V. Patel, E. L. Akylbekova, J. C. Files, C. L. Hardy, S. G. Buxbaum, H. A. Taylor, D. Reich, T. B. Harris, and E. Ziv.** 2008. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. Am J Hum Genet **82:**81-7.

215. **Nathan, C.** 2006. Neutrophils and immunity: challenges and opportunities. Nat Rev Immunol **6:**173-82.

216. **Neidhart, M., J. Rethage, S. Kuchen, P. Kunzler, R. M. Crowl, M. E. Billingham, R. E. Gay, and S. Gay.** 2000. Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. Arthritis Rheum **43:**2634-47.

217. **Nguyen, L., M. Li, T. Chaowanachan, D. J. Hu, S. Vanichseni, P. A. Mock, F. van Griensven, M. Martin, U. Sangkum, K. Choopanya, J. W. Tappero, R. B. Lal, and C. Yang.** 2004. CCR5 promoter human haplogroups associated with HIV-1 disease progression in Thai injection drug users. AIDS **18:**1327-33.

218. **Nibbs, R. J., S. M. Wylie, I. B. Pragnell, and G. J. Graham.** 1997. Cloning and characterization of a novel murine beta chemokine receptor, D6. Comparison to three other related macrophage inflammatory protein-1alpha receptors, CCR-1, CCR-3, and CCR-5. J Biol Chem **272:**12495-504.

219. **Nicholson, J. K., S. W. Browning, R. L. Hengel, E. Lew, L. E. Gallagher, D. Rimland, and J. S. McDougal.** 2001. CCR5 and CXCR4 expression on memory and naive T cells in HIV-1 infection and response to highly active antiretroviral therapy. J Acquir Immune Defic Syndr **27:**105-15.

220. **Nouzova, M., N. Holtan, M. M. Oshiro, R. B. Isett, J. L. Munoz-Rodriguez, A. F. List, M. L. Narro, S. J. Miller, N. C. Merchant, and B. W. Futscher.** 2004. Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. J Pharmacol Exp Ther **311:**968-81.

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

222. **Nzila, N., M. Laga, M. A. Thiam, K. Mayimona, B. Edidi, E. Van Dyck, F. Behets, S. Hassig, A. Nelson, K. Mokwa, and et al.** 1991. HIV and other sexually transmitted diseases among female prostitutes in Kinshasa. AIDS **5:**715-21.

223. **O'Brien, S. J., and G. W. Nelson.** 2004. Human genes that limit AIDS. Nat Genet **36:**565-74.
224. **Okano, M., D. W. Bell, D. A. Haber, and E. Li.** 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:**247-57.
225. **Ometto, L., M. Zanchetta, A. Cabrelle, G. Esposito, M. Mainardi, L. Chieco-Bianchi, and A. De Rossi.** 1999. Restriction of HIV type 1 infection in macrophages heterozygous for a deletion in the CC-chemokine receptor 5 gene. AIDS Res Hum Retroviruses **15:**1441-52.
226. **Oppenheim, J. J., O. M. Z. Howard, and E. Goetzl.** 2000. Chemotactic factors, neuropeptides, and other ligands for seven transmembrane receptors In: Cytokine reference: a compendium of cytokines and other mediators of host defence, London.
227. **Ordway, J. M., and T. Curran.** 2002. Methylation matters: modeling a manageable genome. Cell Growth Differ **13:**149-62.
228. **Ostrowski, M. A., S. J. Justement, A. Catanzaro, C. A. Hallahan, L. A. Ehler, S. B. Mizell, P. N. Kumar, J. A. Mican, T. W. Chun, and A. S. Fauci.** 1998. Expression of chemokine receptors CXCR4 and CCR5 in HIV-1-infected and uninfected individuals. J Immunol **161:**3195-201.
229. **Page, K. A., N. R. Landau, and D. R. Littman.** 1990. Construction and use of a human immunodeficiency virus vector for analysis of virus infectivity. J Virol **64:**5270-6.
230. **Paxton, W. A., R. Liu, S. Kang, L. Wu, T. R. Gingeras, N. R. Landau, C. R. Mackay, and R. A. Koup.** 1998. Reduced HIV-1 infectability of CD4+ lymphocytes from exposed-uninfected individuals: association with low expression of CCR5 and high production of beta-chemokines. Virology **244:**66-73.
231. **Perry, G. H., A. Ben-Dor, A. Tsalenko, N. Sampas, L. Rodriguez-Revenga, C. W. Tran, A. Scheffer, I. Steinfeld, P. Tsang, N. A. Yamada, H. S. Park, J. I. Kim, J. S. Seo, Z. Yakhini, S. Laderman, L. Bruhn, and C. Lee.** 2008. The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet **82:**685-95.
232. **Petersen, D. C., M. J. Kotze, M. D. Zeier, A. Grimwood, D. Pretorius, E. Vardas, E. J. van Rensburg, and V. M. Hayes.** 2001. Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis assay. AIDS **15:**171-7.
233. **Petronis, A.** 2006. Epigenetics and twins: three variations on the theme. Trends Genet **22:**347-50.
234. **Pinney, J. W., J. E. Dickerson, W. Fu, B. E. Sanders-Beer, R. G. Ptak, and D. L. Robertson.** 2009. HIV-host interactions: a map of viral perturbation of the host system. AIDS **23:**549-54.
235. **Pinto, L. A., J. Sullivan, J. A. Berzofsky, M. Clerici, H. A. Kessler, A. L. Landay, and G. M. Shearer.** 1995. ENV-specific cytotoxic T lymphocyte responses in HIV seronegative health care workers occupationally exposed to HIV-contaminated body fluids. J Clin Invest **96:**867-76.
236. **Pinto, L. A., M. S. Williams, M. J. Dolan, P. A. Henkart, and G. M. Shearer.** 2000. Beta-chemokines inhibit activation-induced death of lymphocytes from HIV-infected individuals. Eur J Immunol **30:**2048-55.
237. **Platt, E. J., K. Wehrly, S. E. Kuhmann, B. Chesebro, and D. Kabat.** 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol **72:**2855-64.
238. **Pons, D., F. R. de Vries, P. J. van den Elsen, B. T. Heijmans, P. H. Quax, and J. W. Jukema.** 2009. Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. Eur Heart J **30:**266-77.
239. **Potocki, L., W. Bi, D. Treadwell-Deering, C. M. Carvalho, A. Eifert, E. M. Friedman, D. Glaze, K. Krull, J. A. Lee, R. A. Lewis, R. Mendoza-Londono, P. Robbins-Furman, C. Shaw, X. Shi, G. Weissenberger, M. Withers, S. A. Yatsenko, E. H. Zackai, P. Stankiewicz, and J. R. Lupski.** 2007. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. Am J Hum Genet **80:**633-49.

240. **Premack, B. A., and T. J. Schall.** 1996. Chemokine receptors: gateways to inflammation and infection. Nat Med **2:**1174-8.
241. **Ptak, C., and A. Petronis.** 2008. Epigenetics and complex disease: from etiology to new therapeutics. Annu Rev Pharmacol Toxicol **48:**257-76.
242. **Ptak, R. G., W. Fu, B. E. Sanders-Beer, J. E. Dickerson, J. W. Pinney, D. L. Robertson, M. N. Rozanov, K. S. Katz, D. R. Maglott, K. D. Pruitt, and C. W. Dieffenbach.** 2008. Cataloguing the HIV type 1 human protein interaction network. AIDS Res Hum Retroviruses **24:**1497-502.
243. **Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, and R. H. Gray.** 2000. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med **342:**921-9.
244. **Rakyan, V., and E. Whitelaw.** 2003. Transgenerational epigenetic inheritance. Curr Biol **13:**R6.
245. **Rakyan, V. K., M. E. Blewitt, R. Druker, J. I. Preis, and E. Whitelaw.** 2002. Metastable epialleles in mammals. Trends Genet **18:**348-51.
246. **Rakyan, V. K., J. Preis, H. D. Morgan, and E. Whitelaw.** 2001. The marks, mechanisms and memory of epigenetic states in mammals. Biochem J **356:**1-10.
247. **Ramjee, G., N. S. Morar, M. Alary, L. Mukenge-Tshibaka, B. Vuylsteke, V. Ettiegne-Traore, V. Chandeying, S. A. Karim, and L. Van Damme.** 2000. Challenges in the conduct of vaginal microbicide effectiveness trials in the developing world. AIDS **14:**2553-7.
248. **Reddy, K., C. A. Winkler, L. Werner, K. Mlisana, S. S. Abdool Karim, and T. Ndung'u.** APOBEC3G expression is dysregulated in primary HIV-1 infection and polymorphic variants influence CD4+ T-cell counts and plasma viral load. AIDS **24:**195-204.
249. **Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles.** 2006. Global variation in copy number in the human genome. Nature **444:**444-54.
250. **Reich, D., M. A. Nalls, W. H. Kao, E. L. Akylbekova, A. Tandon, N. Patterson, J. Mullikin, W. C. Hsueh, C. Y. Cheng, J. Coresh, E. Boerwinkle, M. Li, A. Waliszewska, J. Neubauer, R. Li, T. S. Leak, L. Ekunwe, J. C. Files, C. L. Hardy, J. M. Zmuda, H. A. Taylor, E. Ziv, T. B. Harris, and J. G. Wilson.** 2009. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS Genet **5:**e1000360.
251. **Reid, G., R. Gallais, and R. Metivier.** 2009. Marking time: the dynamic role of chromatin and covalent modification in transcription. Int J Biochem Cell Biol **41:**155-63.
252. **Reuse, S., M. Calao, K. Kabeya, A. Guiguen, J. S. Gatot, V. Quivy, C. Vanhulle, A. Lamine, D. Vaira, D. Demonte, V. Martinelli, E. Veithen, T. Cherrier, V. Avettand, S. Poutrel, J. Piette, Y. de Launoit, M. Moutschen, A. Burny, C. Rouzioux, S. De Wit, G. Herbein, O. Rohr, Y. Collette, O. Lambotte, N. Clumeck, and C. Van Lint.** 2009. Synergistic activation of HIV-1 expression by deacetylase inhibitors and prostratin: implications for treatment of latent infection. PLoS One **4:**e6093.
253. **Reynes, J., V. Baillat, P. Portales, J. Clot, and P. Corbeau.** 2003. Low CD4+ T-cell surface CCR5 density as a cause of resistance to in vivo HIV-1 infection. J Acquir Immune Defic Syndr **34:**114-6.
254. **Reynes, J., V. Baillat, P. Portales, J. Clot, and P. Corbeau.** 2004. Relationship between CCR5 density and viral load after discontinuation of antiretroviral therapy. JAMA **291:**46.
255. **Reynes, J., P. Portales, M. Segondy, V. Baillat, P. Andre, O. Avinens, M. C. Picot, J. Clot, J. F. Eliaou, and P. Corbeau.** 2001. CD4 T cell surface CCR5 density as a host factor in HIV-1 disease progression. Aids **15:**1627-34.

256. **Reynes, J., P. Portales, M. Segondy, V. Baillat, P. Andre, B. Reant, O. Avinens, G. Couderc, M. Benkirane, J. Clot, J. F. Eliaou, and P. Corbeau.** 2000. CD4+ T cell surface CCR5 density as a determining factor of virus load in persons infected with human immunodeficiency virus type 1. J Infect Dis **181:**927-32.
257. **Riggs, A. D.** 1975. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet **14:**9-25.
258. **Rollins, R. A., F. Haghighi, J. R. Edwards, R. Das, M. Q. Zhang, J. Ju, and T. H. Bestor.** 2006. Large-scale structure of genomic methylation patterns. Genome Res **16:**157-63.
259. **Rot, A.** 2005. Contribution of Duffy antigen to chemokine function. Cytokine Growth Factor Rev **16:**687-94.
260. **Rowland-Jones, S. L., D. F. Nixon, M. C. Aldhous, F. Gotch, K. Ariyoshi, N. Hallam, J. S. Kroll, K. Froebel, and A. McMichael.** 1993. HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant. Lancet **341:**860-1.
261. **Rubbert, A., C. Combadiere, M. Ostrowski, J. Arthos, M. Dybul, E. Machado, M. A. Cohn, J. A. Hoxie, P. M. Murphy, A. S. Fauci, and D. Weissman.** 1998. Dendritic cells express multiple chemokine receptors used as coreceptors for HIV entry. J Immunol **160:**3933-41.
262. **Rucker, J., A. L. Edinger, M. Sharron, M. Samson, B. Lee, J. F. Berson, Y. Yi, B. Margulies, R. G. Collman, B. J. Doranz, M. Parmentier, and R. W. Doms.** 1997. Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses. J Virol **71:**8999-9007.
263. **Ruiz, M. E., C. Cicala, J. Arthos, A. Kinter, A. T. Catanzaro, J. Adelsberger, K. L. Holmes, O. J. Cohen, and A. S. Fauci.** 1998. Peripheral blood-derived CD34+ progenitor cells: CXC chemokine receptor 4 and CC chemokine receptor 5 expression and infection by HIV. J Immunol **161:**4169-76.
264. **Russo, V. E. A., R. A. Martienssen, and A. D. Riggs.** 1996. Epigenetic Mechanisms of Gene Regulation. Cold Spring Harbor Laboratory Press,, Woodbury.
265. **Samson, M., F. Libert, B. J. Doranz, J. Rucker, C. Liesnard, C. M. Farber, S. Saragosti, C. Lapoumeroulie, J. Cognaux, C. Forceille, G. Muyldermans, C. Verhofstede, G. Burtonboy, M. Georges, T. Imai, S. Rana, Y. Yi, R. J. Smyth, R. G. Collman, R. W. Doms, G. Vassart, and M. Parmentier.** 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature **382:**722-5.
266. **Sandovici, I., S. Kassovska-Bratinova, J. C. Loredo-Osti, M. Leppert, A. Suarez, R. Stewart, F. D. Bautista, M. Schiraldi, and C. Sapienza.** 2005. Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. Hum Mol Genet **14:**2135-43.
267. **Schaefer, M., S. Hagemann, K. Hanna, and F. Lyko.** 2009. Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines. Cancer Res **69:**8127-32.
268. **Sebat, J.** 2007. Major changes in our DNA lead to major changes in our thinking. Nat Genet **39:**S3-5.
269. **Seligmann, M., A. J. Pinching, F. S. Rosen, J. L. Fahey, R. M. Khaitov, D. Klatzmann, S. Koenig, N. Luo, J. Ngu, G. Riethmuller, and et al.** 1987. Immunology of human immunodeficiency virus infection and the acquired immunodeficiency syndrome. An update. Ann Intern Med **107:**234-42.
270. **Sewram, S., R. Singh, E. Kormuth, L. Werner, K. Mlisana, S. S. Karim, and T. Ndung'u.** 2009. Human TRIM5alpha expression levels and reduced susceptibility to HIV-1 infection. J Infect Dis **199:**1657-63.
271. **Shalekoff, S., S. Meddows-Taylor, D. B. Schramm, S. L. Donninger, G. E. Gray, G. G. Sherman, A. H. Coovadia, L. Kuhn, and C. T. Tiemessen.** 2008. Host CCL3L1 gene copy number in relation to HIV-1-specific CD4+ and CD8+ T-cell responses and viral load in South African women. J Acquir Immune Defic Syndr **48:**245-54.

272. **Shao, W., J. Tang, W. Song, C. Wang, Y. Li, C. M. Wilson, and R. A. Kaslow.** 2007. CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. Genes Immun **8:**224-31.
273. **Shen, L., Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland, and J. P. Issa.** 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet **3:**2023-36.
274. **Shi, H., S. H. Wei, Y. W. Leu, F. Rahmatpanah, J. C. Liu, P. S. Yan, K. P. Nephew, and T. H. Huang.** 2003. Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. Cancer Res **63:**2164-71.
275. **Shostakovich-Koretskaya, L., G. Catano, Z. A. Chykarenko, W. He, G. Gornalusse, S. Mummidi, R. Sanchez, M. J. Dolan, S. S. Ahuja, R. A. Clark, H. Kulkarni, and S. K. Ahuja.** 2009. Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. AIDS **23:**679-88.
276. **Silvestri, G., M. Paiardini, I. Pandrea, M. M. Lederman, and D. L. Sodora.** 2007. Understanding the benign nature of SIV infection in natural hosts. J Clin Invest **117:**3148-54.
277. **Simonsen, J. N., F. A. Plummer, E. N. Ngugi, C. Black, J. K. Kreiss, M. N. Gakinya, P. Waiyaki, L. J. D'Costa, J. O. Ndinya-Achola, P. Piot, and et al.** 1990. HIV infection among lower socioeconomic strata prostitutes in Nairobi. AIDS **4:**139-44.
278. **Smith, K. Y., S. Kumar, J. J. Pulvirenti, M. Gianesin, H. A. Kessler, and A. Landay.** 2002. CCR5 and CXCR4 expression after highly active antiretroviral therapy (HAART). J Acquir Immune Defic Syndr **30:**458-60.
279. **Smith, M.** 2009. The year in human and medical genetics. Highlights of 2007-2008. Ann N Y Acad Sci **1151:**1-21.
280. **Smith, M. W., M. Carrington, C. Winkler, D. Lomb, M. Dean, G. Huttley, and S. J. O'Brien.** 1997. CCR2 chemokine receptor and AIDS progression. Nat Med **3:**1052-3.
281. **Stephens, J. C., D. E. Reich, D. B. Goldstein, H. D. Shin, M. W. Smith, M. Carrington, C. Winkler, G. A. Huttley, R. Allikmets, L. Schriml, B. Gerrard, M. Malasky, M. D. Ramos, S. Morlot, M. Tzetis, C. Oddoux, F. S. di Giovine, G. Nasioulas, D. Chandler, M. Aseev, M. Hanson, L. Kalaydjieva, D. Glavac, P. Gasparini, E. Kanavakis, M. Claustres, M. Kambouris, H. Ostrer, G. Duff, V. Baranov, H. Sibul, A. Metspalu, D. Goldman, N. Martin, D. Duffy, J. Schmidtke, X. Estivill, S. J. O'Brien, and M. Dean.** 1998. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet **62:**1507-15.
282. **Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S. W. Scherer, S. Tavare, P. Deloukas, M. E. Hurles, and E. T. Dermitzakis.** 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science **315:**848-53.
283. **Strickland, F. M., and B. C. Richardson.** 2008. Epigenetics in human autoimmunity. Epigenetics in autoimmunity - DNA methylation in systemic lupus erythematosus and beyond. Autoimmunity **41:**278-86.
284. **Suzuki, H., E. Gabrielson, W. Chen, R. Anbazhagan, M. van Engeland, M. P. Weijenberg, J. G. Herman, and S. B. Baylin.** 2002. A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. Nat Genet **31:**141-9.
285. **Suzuki, M. M., and A. Bird.** 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet **9:**465-76.
286. **Szyf, M.** 2009. Epigenetics, DNA methylation, and chromatin modifying drugs. Annu Rev Pharmacol Toxicol **49:**243-63.
287. **Takai, D., and P. A. Jones.** 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A **99:**3740-5.

288. **Thomas, S. M., D. B. Tse, D. S. Ketner, G. Rochford, D. A. Meyer, D. D. Zade, P. N. Halkitis, A. Nadas, W. Borkowsky, and M. Marmor.** 2006. CCR5 expression and duration of high risk sexual activity among HIV-seronegative men who have sex with men. Aids **20:**1879-83.
289. **Torre, D., and A. Pugliese.** 2008. Platelets and HIV-1 infection: old and new aspects. Curr HIV Res **6:**411-8.
290. **Tost, J., and I. G. Gut.** 2007. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. Methods Mol Biol **373:**89-102.
291. **Tost, J., and I. G. Gut.** 2007. DNA methylation analysis by pyrosequencing. Nat Protoc **2:**2265-75.
292. **Tournamille, C., Y. Colin, J. P. Cartron, and C. Le Van Kim.** 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet **10:**224-8.
293. **Townson, J. R., L. F. Barcellos, and R. J. Nibbs.** 2002. Gene copy number regulates the production of the human chemokine CCL3-L1. Eur J Immunol **32:**3016-26.
294. **Trials, A. V.** 2001, posting date. Considerations for Phase III Trial Design and Endpoints N http://www.niaid.nih.gov/vrc/pdf/p3trialsend.pdf. [Online.]
295. **Tufarelli, C., J. A. Stanley, D. Garrick, J. A. Sharpe, H. Ayyub, W. G. Wood, and D. R. Higgs.** 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet **34:**157-65.
296. **UNAIDS.** 2007. AIDS epidemic update.
297. **UNAIDS.** 2006. AIDS Epidemic Update. Geneva: WHO/UNAIDS.
298. **Urban, T. J., A. C. Weintrob, J. Fellay, S. Colombo, K. V. Shianna, C. Gumbs, M. Rotger, K. Pelak, K. K. Dang, R. Detels, J. J. Martinson, S. J. O'Brien, N. L. Letvin, A. J. McMichael, B. F. Haynes, M. Carrington, A. Telenti, N. L. Michael, and D. B. Goldstein.** 2009. CCL3L1 and HIV/AIDS susceptibility. Nat Med **15:**1110-2.
299. **Valdez, H., S. F. Purvis, M. M. Lederman, M. Fillingame, and P. A. Zimmerman.** 1999. Association of the CCR5delta32 mutation with improved response to antiretroviral therapy. Jama **282:**734.
300. **van der Maarel, S. M.** 2008. Epigenetic mechanisms in health and disease. Ann Rheum Dis **67 Suppl 3:**iii97-100.
301. **van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. Abdool Karim, A. Grobler, N. Barnabas, I. Iriogbe, and S. S. Abdool Karim.** 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One **3:**e1954.
302. **van Manen, D., M. A. Rits, C. Beugeling, K. van Dort, H. Schuitemaker, and N. A. Kootstra.** 2008. The effect of Trim5 polymorphisms on the clinical course of HIV-1 infection. PLoS Pathog **4:**e18.
303. **van Ommen, G. J.** 2005. Frequency of new copy number variation in humans. Nat Genet **37:**333-4.
304. **van Vliet, J., N. A. Oates, and E. Whitelaw.** 2007. Epigenetic mechanisms in the context of complex diseases. Cell Mol Life Sci **64:**1531-8.
305. **Vincent, T., P. Portales, V. Baillat, A. Eden, J. Clot, J. Reynes, and P. Corbeau.** 2006. The immunological response to highly active antiretroviral therapy is linked to CD4+ T-cell surface CCR5 density. J Acquir Immune Defic Syndr **43:**377-8.
306. **Volik, S., B. J. Raphael, G. Huang, M. R. Stratton, G. Bignel, J. Murnane, J. H. Brebner, K. Bajsarowicz, P. L. Paris, Q. Tao, D. Kowbel, A. Lapuk, D. A. Shagin, I. A. Shagina, J. W. Gray, J. F. Cheng, P. J. de Jong, P. Pevzner, and C. Collins.** 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome. Genome Res **16:**394-404.
307. **Waddington, C. H.** 1942. The Epigenotype. Endeavour. **1**:18-20
308. **Walley, N. M., B. Julg, S. P. Dickson, J. Fellay, D. Ge, B. D. Walker, M. Carrington, M. S. Cohen, P. I. de Bakker, D. B. Goldstein, K. V. Shianna, B. F. Haynes, N. L. Letvin, A. J.**

McMichael, N. L. Michael, and A. C. Weintrob. 2009. The Duffy antigen receptor for chemokines null promoter variant does not influence HIV-1 acquisition or disease progression. Cell Host Microbe **5:**408-10; author reply 418-9.

309. **Wang, W., H. Soto, E. R. Oldham, M. E. Buchanan, B. Homey, D. Catron, N. Jenkins, N. G. Copeland, D. J. Gilbert, N. Nguyen, J. Abrams, D. Kershenovich, K. Smith, T. McClanahan, A. P. Vicari, and A. Zlotnik.** 2000. Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). J Biol Chem **275:**22313-23.
310. **Wawer, M. J., R. H. Gray, N. K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, F. Nalugoda, F. Wabwire-Mangen, M. P. Meehan, and T. C. Quinn.** 2005. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis **191:**1403-9.
311. **Weinmann, A. S., P. S. Yan, M. J. Oberley, T. H. Huang, and P. J. Farnham.** 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev **16:**235-44.
312. **Weissman, D., M. Dybul, M. B. Daucher, R. T. Davey, Jr., R. E. Walker, and J. A. Kovacs.** 2000. Interleukin-2 up-regulates expression of the human immunodeficiency virus fusion coreceptor CCR5 by CD4+ lymphocytes in vivo. J Infect Dis **181:**933-8.
313. **Wigler, M., D. Levy, and M. Perucho.** 1981. The somatic replication of DNA methylation. Cell **24:**33-40.
314. **Willcocks, L. C., P. A. Lyons, M. R. Clatworthy, J. I. Robinson, W. Yang, S. A. Newland, V. Plagnol, N. N. McGovern, A. M. Condliffe, E. R. Chilvers, D. Adu, E. C. Jolly, R. Watts, Y. L. Lau, A. W. Morgan, G. Nash, and K. G. Smith.** 2008. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. J Exp Med **205:**1573-82.
315. **Wilson, A. G.** 2008. Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. J Periodontol **79:**1514-9.
316. **Winkler, C. A., P. An, R. Johnson, G. W. Nelson, and G. Kirk.** 2009. Expression of Duffy antigen receptor for chemokines (DARC) has no effect on HIV-1 acquisition or progression to AIDS in African Americans. Cell Host Microbe **5:**411-3; author reply 418-9.
317. **Wong, A. H., Gottesman, II, and A. Petronis.** 2005. Phenotypic differences in genetically identical organisms: the epigenetic perspective. Hum Mol Genet **14 Spec No 1:**R11-8.
318. **Wu, L., W. A. Paxton, N. Kassam, N. Ruffing, J. B. Rottman, N. Sullivan, H. Choe, J. Sodroski, W. Newman, R. A. Koup, and C. R. Mackay.** 1997. CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. J Exp Med **185:**1681-91.
319. **Yamashita, T. E., J. P. Phair, A. Munoz, J. B. Margolick, R. Detels, S. J. O'Brien, J. W. Mellors, S. M. Wolinsky, and L. P. Jacobson.** 2001. Immunologic and virologic response to highly active antiretroviral therapy in the Multicenter AIDS Cohort Study. Aids **15:**735-46.
320. **Yang, Y., E. K. Chung, Y. L. Wu, S. L. Savelli, H. N. Nagaraja, B. Zhou, M. Hebert, K. N. Jones, Y. Shu, K. Kitzmiller, C. A. Blanchong, K. L. McBride, G. C. Higgins, R. M. Rennebohm, R. R. Rice, K. V. Hackshaw, R. A. Roubey, J. M. Grossman, B. P. Tsao, D. J. Birmingham, B. H. Rovin, L. A. Hebert, and C. Y. Yu.** 2007. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet **80:**1037-54.
321. **Yang, Y. F., M. Tomura, M. Iwasaki, T. Mukai, P. Gao, S. Ono, J. P. Zou, G. M. Shearer, H. Fujiwara, and T. Hamaoka.** 2001. IL-12 as well as IL-2 upregulates CCR5 expression on T cell receptor-triggered human CD4+ and CD8+ T cells. J Clin Immunol **21:**116-25.
322. **Zardo, G., M. I. Tiirikainen, C. Hong, A. Misra, B. G. Feuerstein, S. Volik, C. C. Collins, K. R. Lamborn, A. Bollen, D. Pinkel, D. G. Albertson, and J. F. Costello.** 2002. Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. Nat Genet **32:**453-8.

323. **Zhang, F., W. Gu, M. E. Hurles, and J. R. Lupski.** 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet **10:**451-81.
324. **Zhang, J., and D. M. Webb.** 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet **13:**1785-91.
325. **Zhang, Z. N., H. Shang, Y. J. Jiang, J. Liu, D. Dai, Y. Y. Diao, W. Q. Geng, X. Jin, and Y. N. Wang.** 2006. Activation and coreceptor expression of T lymphocytes induced by highly active antiretroviral therapy in Chinese HIV/AIDS patients. Chin Med J (Engl) **119:**1966-71.
326. **Zhou, H., M. Xu, Q. Huang, A. T. Gates, X. D. Zhang, J. C. Castle, E. Stec, M. Ferrer, B. Strulovici, D. J. Hazuda, and A. S. Espeseth.** 2008. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe **4:**495-504.
327. **Zou, W., A. Foussat, S. Houhou, I. Durand-Gasselin, A. Dulioust, L. Bouchet, P. Galanaud, Y. Levy, and D. Emilie.** 1999. Acute upregulation of CCR-5 expression by CD4+ T lymphocytes in HIV-infected patients treated with interleukin-2. ANRS 048 IL-2 Study Group. Aids **13:**455-63.