Association of Genetic Polymorphisms in Select HIV-1 Replication Cofactors with Susceptibility to HIV-1 Infection and Disease Progression

by

PARADISE Z. MADLALA

B.Sc. Hons. (University of Natal), MSc. (University of KwaZulu-Natal)

Submitted in fulfillment of the academic requirements for the degree of Doctor of Philosophy (PhD) in the School of Biochemistry, Genetics and Microbiology

University of KwaZulu-Natal

2011

PREFACE

The experimental work described in this dissertation was conducted in the HIV Pathogenesis Programme (HPP), Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, January 2007 to June 2010, under the direct supervision of Professor Thumbi Ndung'u. Part of DNA sequencing was conducted at the Department of Molecular Medicine, Katholieke Universiteit Leuven, Belgium. Part of DNA genotyping was conducted at the National Cancer Institute in Frederick, Maryland, USA.

I hereby declare that, unless specifically indicated to the contrary in the text, this thesis is my own original work and has not been submitted to any university for a similar or any other degree.

_____

P. Z. Madlala

_____

Prof. T. Ndung'u (Supervisor)

_____

Prof. E. Kormuth (Co-Supervisor)

PUBLICATIONS AND PRESENTATIONS

Publications

1) **Paradise Madlala**, Rik Gijsbers, Annaleen Hombrouck, Lise Werner, Koleka Milisana, Ping An, Salim Abdool Karim, Cheryl Winkler, Zeger Debyser and Thumbi Ndung'u. Association of polymorphisms in the LEDGF/p75 gene (*PSIP1*) with susceptibility to HIV-1 infection and disease progression. *AIDS* 2011; 25 (14): 1711-9.

2) **Paradise Madlala**, Ping An, Frauke Christ, Lise Werner, Sengeziwe Sibeko, Salim S. Abdool Karim, Cheryl A. Winkler, Zeger Debyser and Thumbi Ndung'u. Association of genetic variation in the Transportin-SR2 gene (*TNPO3*) with susceptibility to HIV-1 infection and disease progression **(Manuscript in preparation for submission, review and publication December 2011)**

3) **Paradise Madlala**, Ravesh Singh, Lise Werner, Koleka Miisana, Ping An, Salim S. Abdool Karim, Cheryl A. Winkler, and Thumbi Ndung'u. Association of genetic variant A1650G in the Cyclophilin A gene (*PPIA*) with disease progression and mRNA expression levels **(Manuscript in preparation for submission, review and publication December 2011)**

Conference Presentations

1) Association of genetic variation in the Transportin-SR2 gene (*TNPO3*) with susceptibility to HIV-1 infection and disease progression. **Paradise Madlala**, Ping An, Frauke Christ, Lise Werner, Sengeziwe Sibeko, Salim S. Abdool Karim, Cheryl A. Winkler, Zeger Debyser and Thumbi Ndung'u **(4th International Meeting on Retroviral Integration in Siena, Italy in October 2011)**

2) Association of polymorphisms in the LEDGF/p75 gene (*PSIP1*) with susceptibility to HIV-1 infection and disease progression. **Paradise Madlala**, Rik Gijsbers, Annaleen Hombrouck, Lise Werner, Koleka Milisana, Ping An, Salim Abdool Karim, Cheryl Winkler, Zeger Debyser and Thumbi Ndung'u **(XVIII International AIDS2010 Conference in Vienna, Austria in July 2010)**

3) Was a delegate at the Third General Assembly of the Targeting HIV Integrase Co-factors (THINC) Consortium in Pisa, Italy in February 2010

4) Association of polymorphisms in the LEDGF/p75 gene (*PSIP1*) with susceptibility to HIV-1 infection and disease progression. **Paradise Madlala**, Rik Gijsbers, Annaleen Hombrouck, Lise Werner, Koleka Milisana, Salim Abdool Karim, Zeger Debyser and Thumbi Ndung'u 1st **(General Assembly of Targeting HIV Integrase Co-factors (THINC) Consortium in Leuven, Belgium in February 2010)**

5) Association of polymorphisms in the LEDGF/p75 gene (*PSIP1*) with susceptibility to HIV-1 infection and disease progression. **Paradise Madlala**, Rik Gijsbers, Annaleen Hombrouck, Lise Werner, Koleka Milisana, Ping An, Salim Abdool Karim, Cheryl Winkler, Zeger Debyser and Thumbi Ndung'u **(I was an invited Speaker and Session Facilitator at the Academy of Science of South Africa (ASSAf) Department of Science and Technology (DST) National Research Foundation (NRF) Young Scientists Conference, Pretoria, South Africa in September 2011)**

6) Association of genetic variation in the Transportin-SR2 gene (*TNPO3*) with susceptibility to HIV-1 infection and disease progression. **Paradise Madlala**, Ping An, Frauke Christ, Lise Werner, Sengeziwe Sibeko, Salim S. Abdool Karim, Cheryl A. Winkler, Zeger Debyser and Thumbi Ndung'u **(College of Health Sciences Research Symposium, Durban in September 2011)**

ABBREVIATIONS

CypA:	Cyclophilin A
HIV-1:	Human immunodeficiency virus type 1
IN:	Integrase
LEDGF/p5:	Lens epithelium derived growth factor p75
PPIA:	Peptidyl propyl isomerase A
PSIP1:	proprotein convertase 4 (PC4) or splicing factor, arginine/serine-rich 1 (SFRS1) interacting protein 1 (*PSIP1*)
SN:	Seronegative
SNP:	Single nucleotide polymorphism
SNP number:	Since rs numbers are too long, I chose to identify the SNPs analyzed here as SNP numbers so that they can fit in the figures
TRN-SR2:	Transportin-SR2
TNPO3:	Transportin 3
SP:	Seropositive
rs number:	A conventional method in scientific literature to refer to SNPs

ABSTRACT

Objective

Humans differ substantially with respect to susceptibility to human immunodeficiency virus type 1 (HIV-1) infection and disease progression. This heterogeneity is attributed to the interplay between the environment, viral diversity, immune response and host genetics. This study focused on host genetics. We studied the association of single nucleotide polymorphisms (SNPs) in peptidyl prolyl isomerase A (*PPIA*), transportin 3 (*TNPO3*) and PC4 or SFRS1 interacting protein 1 (*PSIP1*) genes with HIV-1 infection and disease progression. These genes code for Cyclophilin A (CypA), Transportin-SR2 (TRN-SR2) and Lens epithelium derived growth factor/p75 (LEDGF/p75) proteins respectively, which are all validated HIV replication cofactors *in vitro*.

Methods

One SNP A1650G in the *PPIA* gene was genotyped in 168 HIV-1 negative and 47 acutely infected individuals using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). 6 intronic and 2 exonic haplotype tagging (ht) SNPs (rs13242262; rs2305325; rs11768572; rs1154330; rs35060568; rs8043; rs6957529; rs10229001) in the *TNPO3* gene, 4 intronic ht SNPs (rs2277191, rs1033056, rs12339417 and rs10283923) and 1 exonic SNP (rs61744944, Q472L) in the *PSIP1* gene were genotyped in 195 HIV-1 negative and 52 acutely infected individuals using

TaqMan assays. The rs1154330, rs2277191, rs12339417 and rs61744944 were further genotyped in 403 chronically infected individuals. CypA and LEDGF/p75 messenger RNA (mRNA) expression levels in peripheral blood mononuclear cells (PBMCs) were quantified by real-time reverse transcriptase polymerase chain reaction (RT-PCR). The impact of the Q472L mutation on the interaction of LEDGF/p75 with HIV-1 integrase (IN) was measured by AlphaScreen.

Results

The minor allele (G) of SNP A1650G (1650G) in the promoter region of *PPIA* was significantly associated with higher viral load ($p<0.01$), lower $CD4^+$ T cell counts ($p<0.01$) and showed a possible association with rapid $CD4^+$ T cell decline ($p=0.05$). The 1650G was further associated with higher CypA expression post HIV-1 infection. The minor allele (G) of rs1154330 in the intron region of *TNPO3* was associated with faster HIV-1 acquisition ($p<0.01$), lower $CD4^+$ T cell counts, higher viral load during primary infection ($p<0.05$) and rapid $CD4^+$ T cells decline ($p<0.01$). The minor allele (A) of rs2277191 (rs2277191A) in the intron region of *PSIP1* was more frequent among seropositives ($p=0.06$). Among individuals followed longitudinally, rs2277191A was associated with higher likelihood of HIV-1 acquisition ($p=0.08$) and rapid $CD4^+$T cell decline ($p=0.04$) in the recently infected (primary infection) cohort. In contrast, the minor allele (C) of rs12339417 (rs12339417C) also in the intron region of *PSIP1* was associated with higher $CD4^+$ T cell counts during primary infection. The rs12339417C was also associated with slower rate of $CD4^+$ T cell decline ($p=0.02$) and lower mRNA

levels of LEDGF/p75 ($p<0.01$). Seroconverters had higher preinfection mRNA levels of LEDGF/p75 compared to nonseroconverters ($p<0.01$) and these levels decreased after HIV-1 infection ($p=0.02$). The Q472L mutation showed approximately 2-fold decrease in the association constant (K_d), suggesting stronger binding to HIV-1 integrase. Our findings demonstrate, for the first time, that genetic polymorphisms in the *TNPO3* and *PSIP1* genes may be associated with susceptibility to HIV-1 infection and the disease progression. These data provide *in vivo* evidence that TRN-SR2 and LEDGF/p75 are important host cofactors for HIV-1 replication. This is also the first study to show the association of genetic polymorphisms in the *PPIA* gene with disease outcome in a population (South African) with high burden of HIV-1 infection.

Conclusions

Genetic variation in HIV-1 replication cofactors may be associated with disease outcome in a South African population. These data strongly support the role of these HIV replication cofactors in disease pathogenesis *in vivo* and suggest that these factors are possible targets for therapeutic interventions. However, these data will need to be replicated in larger cohorts to confirm the effect of these genetic variants. Further studies on how to target these factors in antiviral strategies are needed.

AKNOWLEDGEMENTS

Special thanks to Professor Thumbi Ndung'u for his mentorship, support, supervision, guidance, advice, confidence and belief in me during this PhD research project. I also extend my heartfelt appreciation to Professor Kormuth for co-supervising this project.

This study was funded by the seventh framework programme (FP7) of the European commission (THINC, HEALTH-F3-2008-201032). Additional funding was provided by the South African Department of Science and Technology/National Research Foundation Research Chairs Initiative. This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health (NIH), under contract HHSN261200800001E.

We thank Taryn Green, Yuchun Zhou, Beth Binns-Roemer, Sofie Vets and Nam Joo Van der Veken for excellent technical support. The *in vitro* work on CypA was done in collaboration with Ravesh Singh who was investigating the effect of CypA on HIV-1 replication *in vitro* and CypA expression. We acknowledge the participants and their clinicians who participated in the CAPRISA AI 002 and Sinikithemba studies. The CAPRISA AI 002 study was supported by National Institute of Allergy and Infectious Disease (NIAID), NIH, US Department of Health and Human Services (grant U19 AI 51794). The Sinikithemba cohort was supported by a grant from the NIH (grant R01-AI

067073, contract NOI-AI 15422). I would also like to thank my colleagues in the Hasso Plattner and HIV Pathogenesis Programme laboratories at the Nelson R. Mandela School of Medicine, University of KwaZulu-Natal who always gave me moral support and assistance when needed.

I express my appreciation to my family for the endless love, support, and inspiration during this study. Members of my family are: Sipho, Thulisile, Thokozani, Zanele, Bongiwe, and Thembeka. I also thank Sibusiso (Jobe) Sithole for encouraging me.

I give special thanks to Pastor Neli Mkhize of the Soul Harvest and Life Changing Centre of the Apostolic Faith Mission of SA. She selflessly provided for my rehabilitation following a motor vehicle accident in 2009 where I sustained severe head trauma amongst other injuries.

DEDICATION

I would like to dedicate this thesis to my late father Mpopoli Timoty Madlala, late mother Thandi Flora Madlala and late brothers, Mandla Madlala, Eric Madlala, Bonga Madlala, Bongani Madlala and late sister Winnie Madlala. May their souls rest in peace.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1. INTRODUCTION

In1981, the Centers for Disease Control and Prevention (CDC) reported that 3 hospitals in Los Angeles, California in the United States had been treating 5 active homosexual young men for *Pneumocystis carinii* pneumonia, for the period October 1980 – May 1981. These patients also had a laboratory confirmed cytomegalovirus (CMV) infection and Candida mucosal infection. These infections are known as opportunistic infections since they occur in immunocompromised patients. At about the same time, another CDC report mentioned 26 cases of Kaposi's sarcoma among homosexual males, and that eight had died within 2 years of diagnosis [1]. The term "acquired immune deficiency syndrome" (AIDS) was used for the first time in 1982 to describe this disease, which was suspected to have been caused by an infectious agent transmitted sexually or through contaminated blood or blood products. A year later (1983), Luc Montagnier's group from the Institut Pasteur, Paris, France, discovered the human immunodeficiency virus type 1 (HIV-1) as the infectious agent that causes AIDS [2] and this was confirmed by two other independent research groups [3, 4].

More than 25 million people have died of AIDS since 1981 and there are still increasing numbers of people living with HIV. 33.4 million people were estimated to be living with HIV in 2008, which was 20% higher than the number in 2000 and a prevalence of approximately threefold higher than in 1990 (http://www.unaids.org). This report estimated that 2.7 million new infections occurred in 2008. The increasing number of people living with HIV could be explained by a combined effect of continued high rates of new HIV infections in addition to the beneficial impact of antiretroviral therapy.

Interestingly, UNAIDS 2010 epidemiological data reported that the number of new infections had fallen by 19% since 1999, the year in which it is thought that the epidemic peaked, globally (http://www.unaids.org). This report also states that of the estimated 15 million people living with HIV in low- and middle- income countries who need treatment today, over one third of them have access–translating into fewer AIDS deaths. Although the epidemiological estimates appear to suggest that the HIV epidemic has stabilized, sub-Saharan Africa remains the region most heavily affected by HIV, accounting for most new HIV infections and world's AIDS-related deaths in 2010.

Females continue to be affected disproportionately by HIV in sub-Saharan Africa, accounting for approximately 60% of estimated HIV infections in this region [5]. Social, legal and economic disadvantages contribute towards this high HIV infection rate observed in females. The world's response to HIV/AIDS has grown and improved considerably over the past three decades. However, universal access to antiretroviral treatment, prevention and continuous HIV-1 research is required to gain the upper hand against the AIDS epidemic (http://www.unaids.org).

1.1 An overview of HIV-1 infection

HIV-1 is a retrovirus belonging to the lentiviral subfamily. Lentiviruses consist of non-oncogenic retroviruses that are characterized by long incubation periods and persistent infection. Like all retroviruses, HIV-1 is an enveloped positive strand RNA virus that replicates through inserting its DNA intermediate into the host cell genome [6].

HIV-1 envelope is composed of a lipid bilayer of the host-cell membrane origin, gained during the budding process, and viral envelope glycoprotein (Env). The glycoprotein is formed of heterodimers, each containing a gp120 subunit on the outside of the membrane and a transmembrane gp41 subunit. Together, the two Env subunits mediate viral entry with the gp120 subunit being responsible for binding to the receptor (CD4) and coreceptor (CCR5 or CXCR4) on the host cell, and gp41 is needed for subsequent fusion of the viral and cellular membranes. The inside of the HIV-1 envelope is lined by units of matrix (MA) protein, helping to anchor the gp41 and gp120. HIV-1 is composed of two copies of single-stranded RNA enclosed by a conical capsid protein (CA) comprising the viral protein p24 [7].

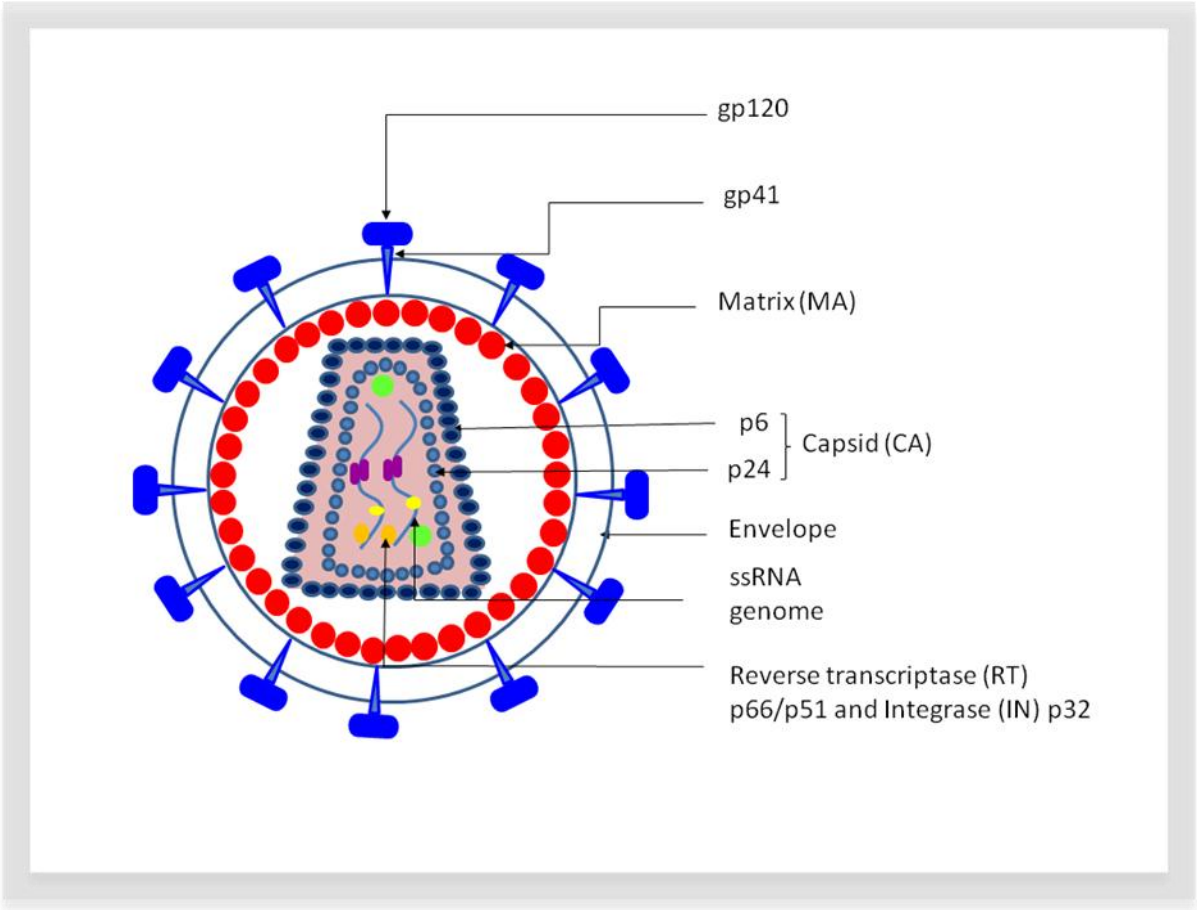

Figure 1.1. Cross-sectional diagram of the HIV-1 virion organization.

These two RNA copies are associated with the nucleocapsid (NC) proteins (p7 and p6) that protect the RNA from digestion by nucleases. Viral enzymes reverse transcriptase (RT), protease (PR) and integrase (IN) are also found within the CA (fig. 1.1). In addition, HIV-1 contains accessory proteins: viral protein R (Vpr); negative regulatory factor (Nef); and virion infectivity factor (Vif) (fig. 1.1.1) [8].

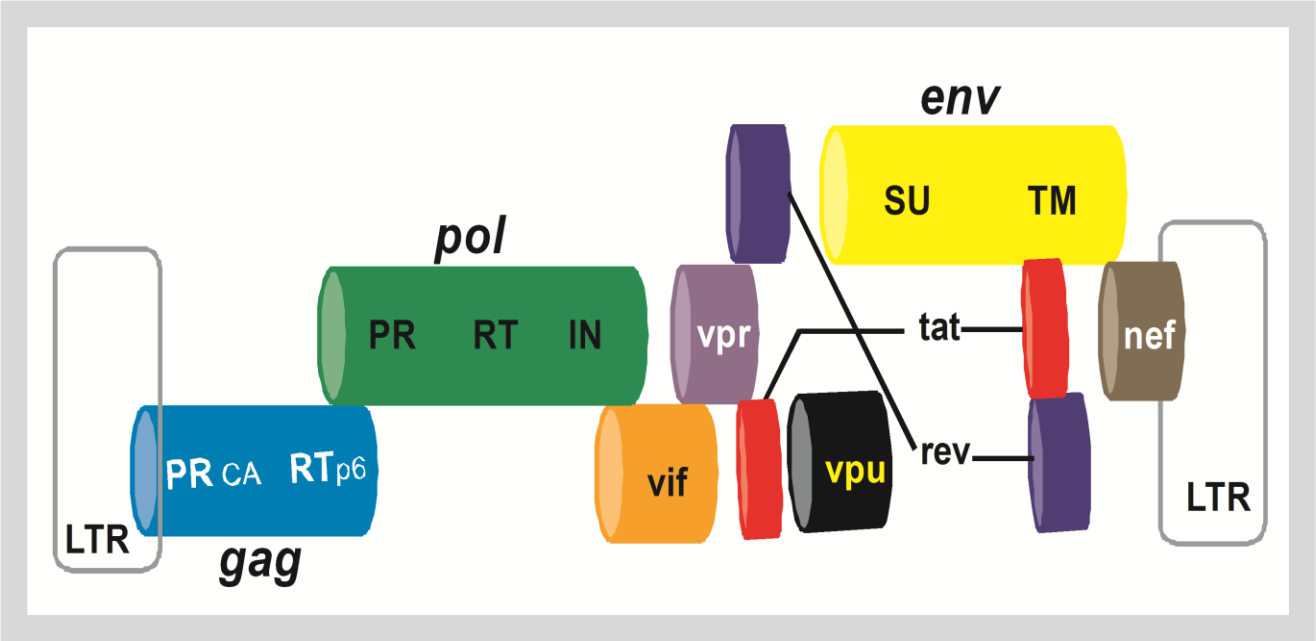

Figure 1.1.1. HIV-1 genome organization.

The largest three genes of HIV-1 transcribed are the *gag*, *pol*, and *env*. Gag is processed into proteins that make up the core of the viral particles, which are MA, CA, NC, and p6 proteins. Envelope proteins, gp120 and gp41, are derived from the Env polyprotein, whereas the Pol polyprotein is processed into viral enzymes, PR, RT and IN that are encapsulated into the core of a virion particle. The other genes encode accessory proteins Vif, Vpr and Nef which are also found in the virion core (fig. 1.1.1). Several cellular proteins have also been detected in the virion [9].

HIV-1 accessory proteins play important role to ensure efficient and productive HIV-1 infection. The function of HIV-1 Vif is to suppress a potent antiviral host mechanism mediated by apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G (APOBEC3G) [10].

Molecular block or absence of Vif impairs replicative capacity of HIV-1 meaning that HIV-1 requires Vif for effective viral spread in host cells, particularly in primary $CD4^+$ T cells, the natural targets of HIV-1 infection [11]. Nef is expressed abundantly during the early stages of HIV-1 infection and is important for replication and progression to AIDS. Nef has three main functions: (1) Nef changes the signalling pathways of host cells by interacting with tyrosine and serine/threonine kinases, (2) Nef increases the infectivity of the virus after cell entry and, (3) Nef decreases the expression of CD4 and major histocompatibility complex class I (MHC I) antigens on the surface of infected cells by interacting with the components of endocytic machinery [8]. The molecular mechanism of Nef-mediated down regulation of CD4 may be important for optimal virus replication and may facilitate the release of virions.

Viral protein R (Vpr) plays an important part in two unusual aspects of the interaction between HIV-1 and the host cell. Vpr prevents infected cells from proliferation, by interfering with normal cell-cycle control, and it collaborates with the MA to enable HIV-1 to enter the nucleus of nondividing cells [12, 13]. The induction of G2 cell cycle arrest by Vpr is thought to indirectly enhance viral replication by increasing transcription from the long terminal repeat (LTR).

Tat is the transactivating protein of HIV-1 LTR for the retroviral replication. In the presence of Tat for transactivation, the efficiency of HIV-1 genome transcription is a hundred-fold higher than that of the non-Tat-activated ones [4]. The Tat protein has also been shown to play a crucial role in AIDS pathogenesis by inducing neurotoxicity in AIDS patients contributing to

the development of HIV-associated dementia [14]. HIV-1 Tat also deregulates cytokines expression in the immune cells, induces apoptosis in neuronal cells, and interacts with other cellular factors to favour its own survival [15]. These processes may inadvertently develop a favourable environment for opportunistic infections [16]. Rev stabilizes the viral unspliced and single spliced RNA and promotes its export from the nucleus by binding to the Rev responsive element (RRE), a specific RNA structure found in the *env* coding region [17]. Vpu, the smallest of the proteins encoded by HIV-1, is the transmembrane protein that interacts with the CD4 molecule in the rough-endoplasmic reticulum (RER), resulting in its degradation via the proteasome pathway [18].

1.2 HIV-1 Subtypes

HIV-1 is classified into M, N, O and P groups. M group represents the major group of HIV-1 strains and it accounts for more than 90 percent of HIV-1 infections worldwide [19]. Groups N, O and P are not spread worldwide but mainly confined to West and Central Africa [20-22]. A hallmark of lentiviruses is their extensive genetic variability due to the high error rate of the reverse transcriptase enzyme, the recombinogenic properties of the diploid viral genome, and the fast turnover of the virions in HIV-1 infected individuals. This characteristic of lentiviruses culminates in many strains and subtypes within the same type (reviewed in [23]).

The subtypes of group M are classified into: A-D, F-H, J-K and circulating recombinant forms (CRFs). Subtype B of HIV-1 is most common in Europe, America, Japan, Australia and some Asian countries, while subtype A, C and D are more common in Africa. The other subtypes F,

G, H, J and K are rare but spread through Africa, South America, and some parts of Europe. When a naïve cell is coinfected with two viruses of different subtypes, the genetic material of both viruses may mix during the replication process resulting in a new viral strain called a circulating recombinant form (CRF) [21]. Virus mutation and replication are likely to result in new HIV genetic subtypes and CRFs.

1.3 The overview of the HIV-1 Replication Cycle.

HIV-1 infect and replicate in $CD4^+$ T cells, dendritic cells and macrophages thereby functionally impairing immune cells resulting in immune deficiency [24]. The replication process of HIV-1 starts with the entry of the virus into the host cell through the attachment of the viral gp120 to the cellular CD4 receptor. This interaction induces conformational change of gp120, enabling it to bind to a co-receptor, CCR5 or CXCR4. This binding triggers another conformational change in the gp140/gp120/co-receptor complex, enabling the fusion between the virus and the host membrane resulting in the HIV-1 core being released into the cytoplasm (reviewed in [25]). This is followed by the synthesis of viral copy DNA (cDNA) from the viral RNA genome, through reverse transcription.

Reverse transcription takes place within the reverse transcription complex (RTC), constituted by infecting virion core proteins, cellular proteins and the viral RNA genome [26, 27]. Following the completion of viral cDNA synthesis, the RTC matures into the pre-integration complex (PIC) [28]. The PIC then migrates towards the nucleus, probably by using cellular motors such as dynein to move along the cell cytoskeleton [26, 29].

At the nuclear membrane, the viral PIC has to transverse through the nuclear pores which allow the diffusion of ions and molecules smaller than 9 nm in diameter. However, the diameter of the PIC is 56 nm [30] which grossly exceeds the diffusion limit (reviewed in [31]). Surprisingly, HIV-1 PIC is actively transported during interphase, reflecting its ability to efficiently infect non-dividing cells [32, 33]. Although a number of HIV-1 proteins that constitute the PIC have been implicated in nuclear import because they carry sequences that may serve as nuclear localization signals (NLS) [34, 35], this field of HIV-1 biology remains relatively muddled.

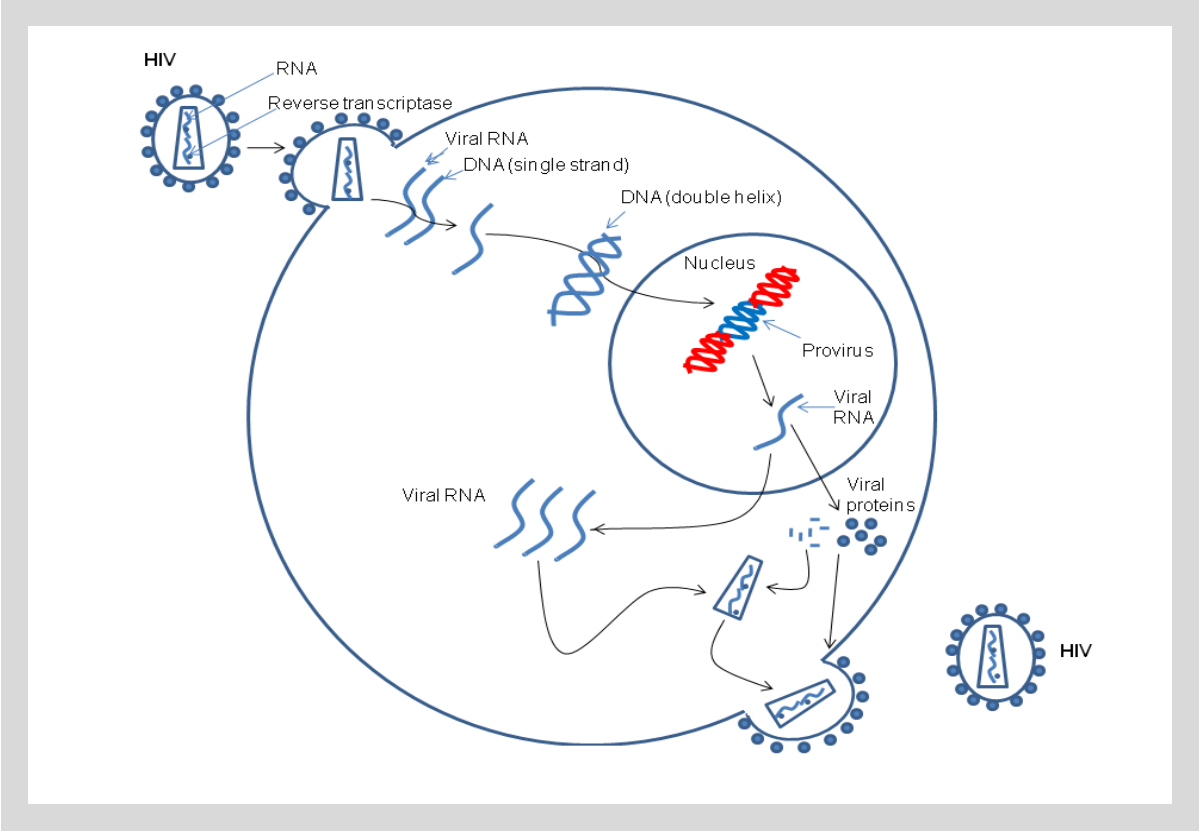

Figure 1.3. A schematic diagram showing HIV-1 replication cycle.

The PIC associates with the host chromosomes upon nuclear entry, at which point the viral cDNA is integrated. From this point on, the provirus remains part of the cellular genome as long as the cell survives. The provirus resides latently in the cell, forming a viral reservoir; alternatively, activation of transcription of viral genes can initiate the formation of the new virions. When transcription starts, multiple spliced mRNA encoding Tat, Rev, and Nef proteins necessary for enhanced and efficient HIV-1 replication get produced. Binding of Tat to the LTR boosts HIV transcription due to the assembly of processive transcription complexes of the LTR promoter. Rev allows nuclear export of full-length genomic RNA, as well as singly spliced mRNA, resulting in the production of Gag, Pol and Env.

The viral proteins and genomic RNA assemble into newly formed particles at the cellular membrane which contains the expressed envelope proteins. Immature viral particles bud from the cellular membrane or in endosomes that release the viral particles upon fusion of the endosome with the cellular membrane. The viral particles maturate outside of the cell which involves proteolytic processing of the viral Gag, and Gag-pol polyproteins by the viral protease (reviewed in [36]).

Lentiviral infections persist lifelong as a function both of their ability to integrate their genome into the host chromosome and their ability to evade host immunity [36]. Their ability to evade host immunity may be due both to high mutation rates, and to their ability to infect the immune system, such as macrophages and T lymphocyte cells.

1.4 Natural History of HIV Infection

HIV-1 infection proceeds through different phases (fig. 1.4) starting with the acute phase of infection, which is characterized by rapid viral replication and spreading, which is usually accompanied by a symptomatic period of disease. Following the acute phase of infection is a latent period (viral set point), during which the virus is brought into equilibrium of replication and relative immune control and no disease occurs. The last phase of infection is the chronic phase. It is during this phase where high levels of viral replication resume at some later time, resulting in disease [37].

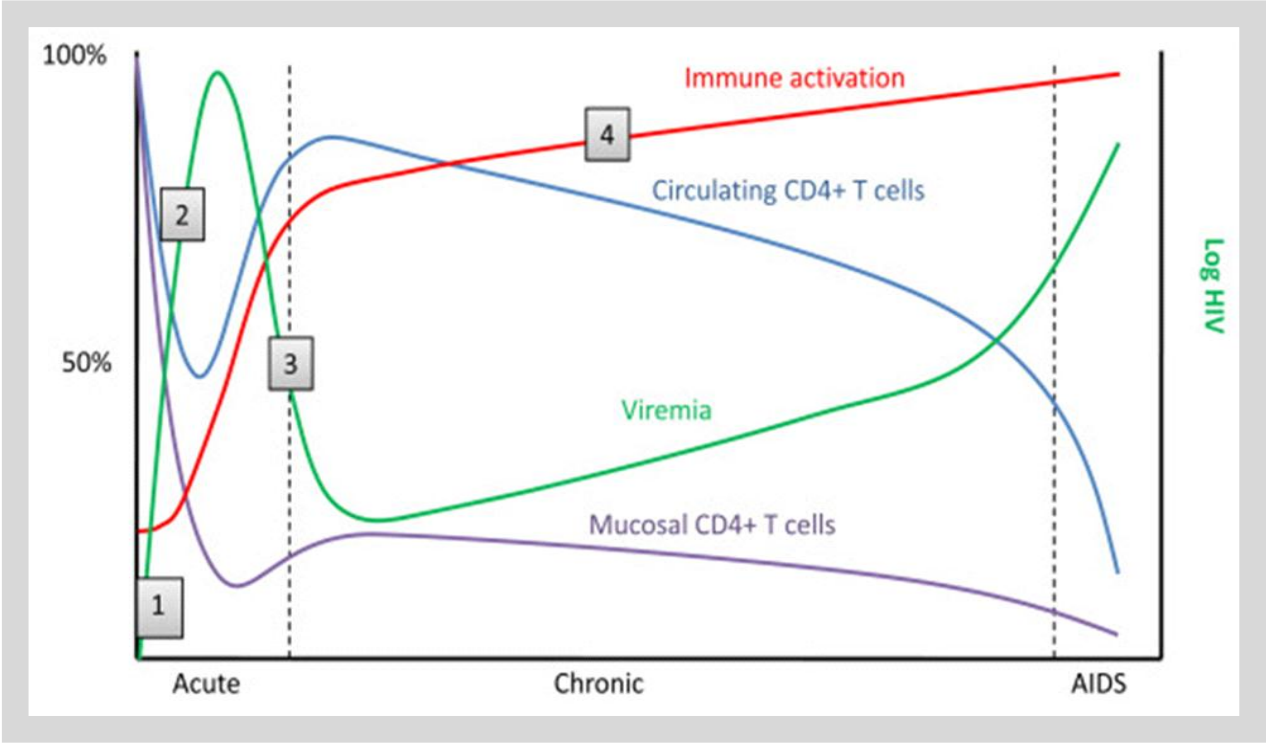

Figure 1.4. Dynamics of peripheral blood $CD4^+$ T cell counts and plasma viral load during a typical course of HIV infection. The three major phases of infection are shown: acute, chronic, and AIDS [38].

Damages caused by lentiviruses accumulate slowly over time and lead to the destruction of the immune system, making victims vulnerable to opportunistic infections. This is the AIDS stage of the disease that ultimately leads to death of the infected person. Once these people reach a CD4 counts below 200 cells/mm^3 of blood they are then diagnosed as having AIDS and are at risk of developing serious opportunistic infections such as *Pneumocystis carinii* pneumonia (PCP), *Mycobacterium avium* complex (MAC), and cytomegalovirus (CMV) [3, 37].

1.5 Interindividual variability in susceptibility to HIV-1 infection and disease progression

HIV-1 has been the most thoroughly studied virus since its discovery, but the accumulated knowledge has not afforded researchers the ability to design therapeutic or prophylactic interventions that will halt the AIDS pandemic. The intensive study of HIV-1/AIDS has found differences between individuals in susceptibility to HIV-1 infection and disease progression (reviewed in [39]). This variability is attributed to the complex interplay between the virus, host genetics, immune response and environment (fig. 1.5).

HIV-1 has a small genome of only 9 genes (fig.1.1.1), which encode 15 proteins as a result of post-translational cleavage. Due to this limited genome, the virus relies on human proteins referred to as HIV-1 replication cofactors for productive infection [40-42]. Previous research on HIV-1 and host genetics has revealed numerous human gene variants that modulate the host response to HIV-1 exposure [43-45]. Human genetic variation offers a partial explanation why some HIV-1 infected individuals maintain undetectable levels of viral load

(less than 50 copies viral RNA per ml) and experience a prolonged course of HIV-1 infection with slow progression to AIDS.

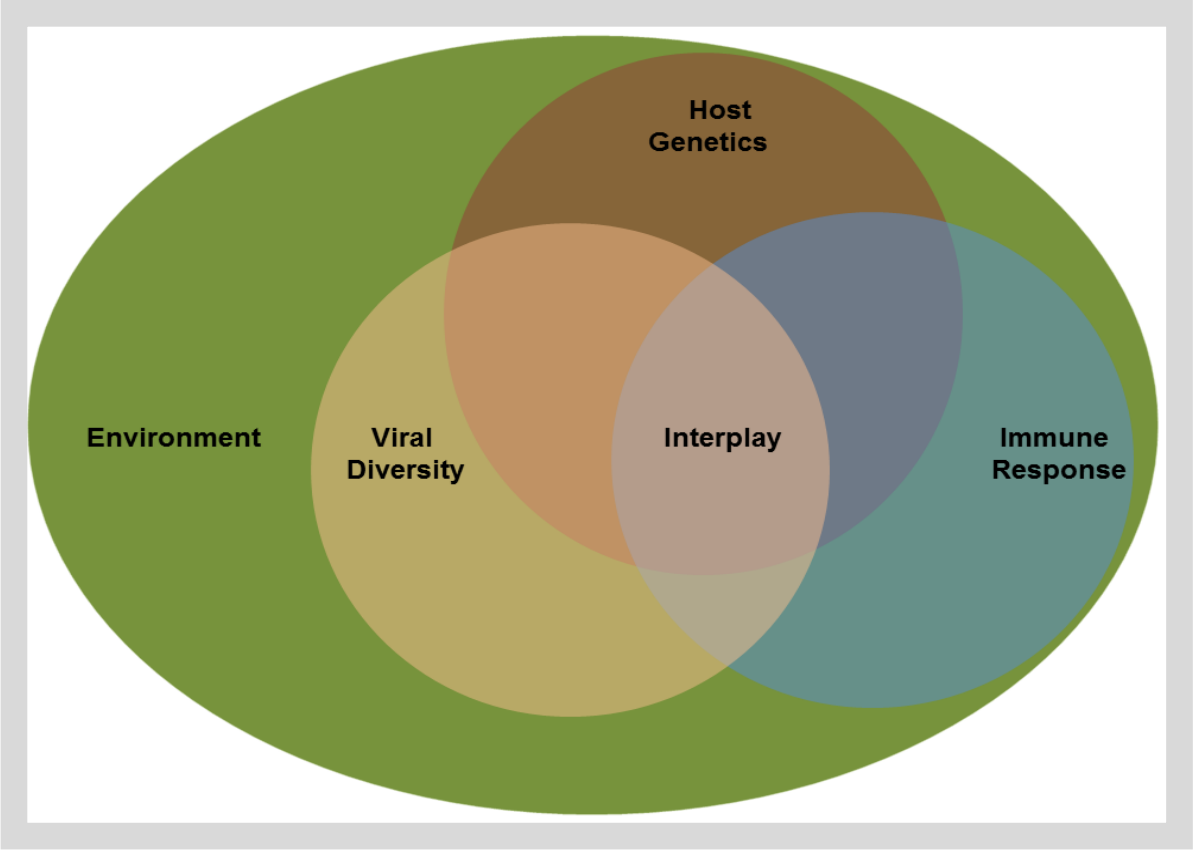

Figure 1.5. Complex interplay between virus, host (genetics and immune response) and the environment.

The availability of genome-wide approaches represents a change in paradigm for complex genetic traits; however, most of the information on host genetic factors that modulate HIV-1 has been obtained from candidate gene studies. The candidate gene study approach has been used to analyze allelic variants in genes that are known or suspected to be involved in HIV-1 pathogenesis and immune response. Therefore most genetic markers relevant to HIV-

1 disease identified using this approach are related to genes that are classified into one of the two categories: (1) HIV-1 replication cofactors, host genes coding proteins that are implicated in HIV-1 replication cycle and (2) restriction factors, immune related genes coding for innate and adaptive immune response factors as well as proteins involved in immune regulation and in specific anti-retroviral defense mechanism (reviewed in [46]).

1.5.1 HIV-1 cell entry: chemokine receptors

Protection against HIV-1 infection is provided by genetic variation in the CC chemokine receptor 5 (CCR5) gene, which encodes the coreceptor for macrophage-tropic (R5) strains of HIV-1 expressed on $CD4^+$ T-cells. Deletion of 32 base pairs (Δ32) in the coding region of *CCR5* gene confers protection against infection by R5 strains in homozygous and delayed disease progression in heterozygous individuals for this allele [47-50].

Chemokines are natural ligands for the same receptors hijacked by HIV-1 for cell entry. Therefore chemokines can have an effect on HIV-1 entry into the target cell by either competing with the virus for coreceptor binding or reducing the expression of the coreceptor on the cell surface by inducing their internalization. A number of chemokines have been reported to contain genetic polymorphisms that play a role in differential susceptibility to HIV-1 infection and disease progression.

Polymorphisms in the *CCL5* gene that codes for RANTES, the CC chemokine, have been shown to inhibit HIV-1 replication *in vitro*, modulate the expression levels of RANTES and reduced expression was shown to accelerate disease progression [51, 52]. Other polymorphisms in the regulatory regions of *CCL5* have been shown to affect susceptibility to HIV-1 infection and disease progression [53-57]. Polymorphisms in the *CCL*3 gene, macrophage inflammatory protein 1 α (MIP1α) were shown to be associated with susceptibility to HIV-1 infection and disease progression [55, 58]. Low copy numbers of *CCL31* gene (MIP1αP) were shown to have detrimental effect on HIV-1 outcome [59, 60]. Stroma-derived factor 1 (SDF-1) on the other hand is the natural ligand for CXCR4, the coreceptor hijacked by CXCR4-using HIV-1 strains (X4) for entry into the target cells. Although polymorphisms in the 3'-untranslated region of the *SDF-1* gene have been reported to have various effects on HIV-1 disease outcome [61-65], however, these data could not be replicated [66].

1.5.2 Intracellular viral and host protein interaction

Post cell entry, HIV-1 continues to interact with numerous host proteins with some proteins acting as antiviral (restriction) factors, restricting HIV-1 replication while HIV-1 replication cofactors enhance productive HIV-1 infection (fig. 1.5.2). HIV-1 exploits numerous human proteins in order to successfully complete its replication cycle and have a productive infection, as demonstrated by small interfering RNA screens [67-69]. Human genes that code for HIV-1 replication factors are ideal candidate genes for genetic variants that modulate HIV infection and many of these genes have been investigated in host genetic studies using candidate gene approaches.

Host genetic studies investigating the association between specific gene variants and HIV-1 outcomes have been conducted predominantly in the populations from the developed countries (United States of America and Europe), mainly Caucasian populations. The focus of HIV-1 host genetic studies over the past 15 years, have revealed that genetic variants of specific genes modulate HIV-1 pathogenesis. Two of the genes whose variants have been confirmed to modulate HIV-1 pathogenesis include peptidyl propyl isomerise A gene (*PPIA*) which encodes the Cyclophilin A protein (CypA) and tumor susceptibility gene 101 (*TSG101*).

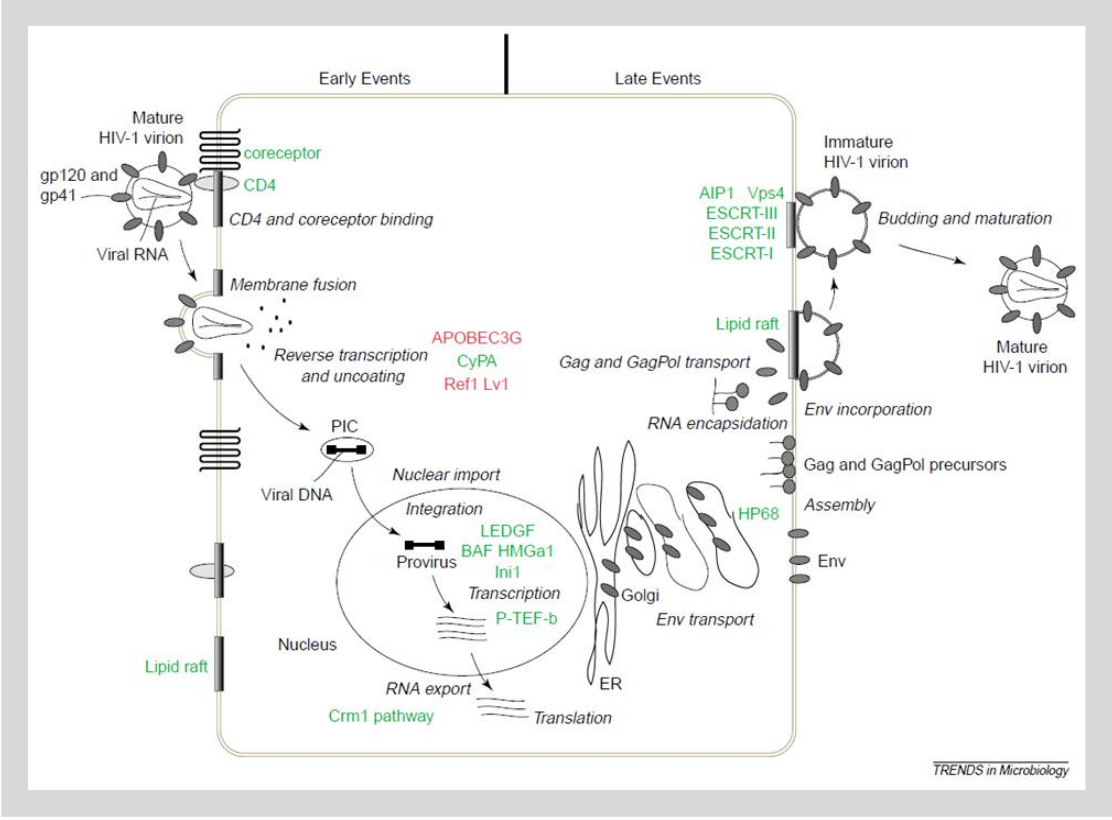

Figure 1.5.2. The interplay between HIV-1 and cellular proteins during HIV-1 life cycle. Cellular proteins that promote HIV-1 replication in the host cell are indicated in green and cellular proteins that restrict HIV-1 replication is host cells are in red [70].

TSG101 protein interacts with HIV-1 Gag, P6 protein specifically, and plays a pivotal role in the budding process of new viral particles from the plasma membrane of infected cells. Two polymorphisms (-T183C and +A181C) in the regulatory region of the *TSG101* gene were shown to be associated with viral load and disease progression measured by $CD4^+$ T cell decline [61, 71]. CypA interacts with HIV-1 capsid protein and as a result gets incorporated into the virion particle. Although the exact mode of action which CypA uses to promote HIV-1 infection is not well elucidated, one theory postulated that CypA is involved in the uncoating process of the viral core [72]. Variants in the regulatory region of the *PPIA* gene have been shown to influence $CD4^+$ T-cell depletion and possibly susceptibility to infection [71, 73, 74]. Genetic variants in restriction factors such as APOBEC3G and TRIM5α that potently inhibit HIV-1 replication have been shown to influence disease progression or susceptibility to HIV infection [75-82].

1.5.3 Immunity

It is important to note that bulk of the information on the association between host genetics and HIV-1 disease outcome have been obtained from studies of immune response genes. The most consistent and prominent association identified in HIV-1 host genetic studies are those between human leukocyte antigen (HLA) genes and disease outcome. HLA molecules are expressed at the cell surface where they present antigenic epitopes, viral epitopes to $CD8^+$ T cells, thereby initiating the cytotoxic T cell response. Three genes (*HLA*-A, *HLA*-B and *HLA*-C) encode HLA class I proteins.

The HLA genes are the most polymorphic genes in the human genome, with HLA class I genes presenting an extreme allelic diversity. The potency of elicited immune response in an individual infected by HIV-1 is determined by the viral epitopes that their HLA alleles are able to present to $CD8^+$ T cells. Some HLA alleles such as HLA-B*57, and to a lesser extent HLA-B*5801, have been reported to control HIV-1 and be associated with slower progression to AIDS in several studies [83-85]. In addition, genome-wide association study reported an association between HLA-B*57 and a lower viral set point [54]. Some HLA alleles are thought to be population specific, with B*5701 observed almost exclusively in the Caucasian and B*5703 being common in individuals of African ancestry.

It is important to note that a SNP that is a surrogate for HLA-B*5701 showed the strongest association with HIV-1 viral control or long term non-progression [86-88]. HLA-B*27 has also been reported to restrict HIV-1 similar to HLA-B*57 [89, 90]. However, other HLA-B alleles such as HLA-B*35Px, are associated with rapid progression to AIDS [91, 92]. The associations observed between HIV-1 control and haplotypes in the major histocompatibility complex (MHC) and HLA supertypes [93-97] could be due to individual alleles that are included in these groups and to the linkage disequilibrium structure of the MHC region. Hence, HLA molecules can present different HIV-1 epitopes that result in immune response with differential restriction of HIV-1 by $CD8^+$ T cells.

Previous studies have shown that the homozygosity in *HLA-A*, *HLA-B* and/or *HLA-C* gene results in restricted (or fewer) epitopes recognized by cytotoxic T lymphocytes and

accelerated disease progression [91, 98]. Genetic variation in *HLA-C* gene has also been shown to be associated with viral control and gene expression [87, 99], suggesting that HLA class I molecules play a central role in efficacy of the immune response in certain individuals.

1.5.3.1 Killer cell immunoglobulin-like receptors

In addition to playing an important role in acquired immunity processes, HLA molecules are also ligands for killer cell immunoglobulin-like receptors (KIRs). KIRs are expressed on the surface of natural killer (NK) cells and regulate the activation of NK cells through inhibitory or activating signaling. NK cells constitute an integral part of the innate immune defense mechanism against viruses by directly killing virus infected cells and/or producing cytokines.

Interaction between some HLA molecules and KIRs commonly referred to as HLA-KIR combination, influence HIV-1 clinical outcomes [100]. For examples KIR3DL1 and KIR3DS1 expressed as allelic variants of the same locus on chromosome 19, both protect against disease progression when existing in combination with HLA-B molecules that have a Bw4 serological specificity. Various combinations of inhibitory KIR3DL1 alleles and HLA Bw4 molecules have been associated with lower HIV-1 viraemia and slow disease progression [101]. Likewise, the activating allele KIR2DS1 has been associated with lower viraemia and a delayed progression to AIDS when found alone [102], or in combination with HLA Bw4 molecules that have an isoleucine at position 80 (Bw4-80I) [103, 104]. Functionally, KIR3DS1 has been shown to correlate with strong inhibition of HIV-1 replication [105] and with higher NK cell effector functions in early HIV-1 infection phase [106].

Association studies of the influence of human genetic variation on HIV-1 replication may reveal the essential *in vivo* host factors that interact with HIV-1 and their epidemiologic importance at the population level. However, this approach has been used mostly in studies conducted in populations from developed countries and of European origin and yet there are genetic differences between populations of different origin [107].

1.6 Genes that influence HIV-1 clinical outcomes differ according to populations

Modern humans originated from Africa about 100,000-200,000 years ago, subsequently migrated out of Africa to the rest of the world about 50,000 years ago and this human dispersal shaped the human genome by evolutionary and historical forces [108-113]. Therefore, allele and haplotype diversity in the non-African populations is believed to have been caused by migration out of Africa, followed by rapid expansion of human populations [114-116]. In addition, periodic outbreaks of deadly infectious agents and regional environmental pressures have modified the genetic architecture of disease gene allelic variation in local human populations [117, 118]. Genes required for reproductive and housekeeping functions remained conserved, whereas genes required for immunity and homeostasis display different alleles among different ethnicities. Allele diversity in the genes involved in immunity may be caused by the fact that these genes encode factors required by pathogens for completion of their lifecycle. Therefore these genesunder selective pressures from microbial or ecological conditions that occur on different continents [119].

The best evidence for the influence of human pathogens on natural selection comes from studies of host genetic resistance to Malaria. Malaria became endemic 6000–10 000 years ago while AIDS emerged within the last 30 years [117, 118, 120-123]. Malaria is responsible for most deaths among children and pregnant women. The selective pressure exerted by Malaria over 300–500 generations has resulted in adaptive shifts in the allele frequency of several genes with a role in Malaria resistance [117, 124-126]. These genes show geographical differences in allele frequencies correlated with the occurrence of Malaria. Given that HIV-1/AIDS is too recent, only three decades old, to have caused an adaptive shift in the allele frequency of genes involved in HIV-1 infection, Slatkin *et al*. modeled the effects of HIV-1-mediated selection for resistant CCR2 and CCR5 genotypes, predicting a modest 2–4 year increase or decrease in survival for South Africa, a country of high HIV-1 prevalence [127, 128]. Their model projected that within 100 years, resistance genotype frequencies will increase from 40 to 53% while the susceptible genotypes will decrease from 20 to 10%, leading to a mean increase in AIDS-free survival from 7.8 to 8.8 years.

HIV-1 replication cofactors are increasingly being targeted for antiretroviral therapy, yet there is very little information on genetic variation in these cofactors and their impact if any on HIV. Studies on human genetic factors that affect HIV/AIDS pathogenesis have focused mainly on host immune factors as discussed above (in section 1.5.3). Therefore there is a need to explore other biomedical interventions that may be pursued. This study examined the effect of genetic polymorphisms in select validated HIV-1 replication cofactors on susceptibility to HIV-1 infection and disease progression, shifting the focus from immune factors. An overview of HIV-1 replication cofactors Cyclophilin (CypA), Transportin-SR2 (TRN-SR2) and Lens

epithelium derived growth factor p75 (LEDGF/p75), which were the focus of this study, is given below.

1.6.1 Cyclophilin A (CypA) modulate early post-entry HIV-1 replication outcomes

CypA, also known as peptidyl prolyl isomerase A (PPIA) belongs to a large family of proteins known as the cyclophilins. By convention, we refer to the protein as CypA and the gene as *PPIA*. This family of proteins is defined by a conserved sequence of 150 amino acids that form an eight-stranded β-barrel with a hydrophobic pocket that serves as the binding site for cyclosporin and HIV-1 Gag [129-131]. Although the exact biochemical function in cells of the core cyclophilin domain is unknown, CypA is presumed to play a role in maintaining proper protein conformation. This function is inferred from the fact that cyclophilins catalyze the cis-trans interconversion of peptide bonds N-terminal to proline. This activity has been shown to stimulate the rate of refolding of model proteins *in vitro* [132, 133].

CypA is a cytosolic protein that binds HIV-1 CA and this interaction results in the incorporation of CypA into virion particles [134-137]. Previous studies have demonstrated the CypA promotes HIV-1 infectivity in target cells classifying this protein as HIV-1 replication cofactor [138, 139]. Given the role of cyclophilins in a cell, it is suspected that CypA regulates the conformation of HIV-1 Gag. Further studies were conducted to elucidate the functional role of CypA in HIV-1 infection [140-144]. These studies relied largely on the use of mutations in *gag* or competitive inhibitors such as cyclosporin to block the Gag-cyclophilin interaction. However, neither of these experimental conditions abrogates the interaction completely and both potentially can cause pleiotropic effects [145]. A third confounding issue

is the number and abundance of cyclophilins in mammalian cells; at present there are 15 known human cyclophilins and nearly all have the capability to bind HIV-1 Gag [134, 135].

Braaten and Luban produced cell lines depleted for *PPIA* ($PPIA^{-/-}$ Jurkat T-cell lines), which helped in demonstrating that CypA is required for wild-type replication kinetics of HIV-1 and, more specifically, for the infectivity of HIV-1 virions [145]. However, none of the 14 other known cyclophilins substituted functionally for CypA in $PPIA^{-/-}$ Jurkat cells [145]. These data demonstrated that CypA is important for HIV-1 replication kinetics *in vitro*.

The $PPIA^{-/-}$ Jurkat cells did not produce virions that had significant biochemical abnormalities, which was in keeping with previous studies where the Gag-cyclophilin interaction was disrupted using Gag mutations or cyclosporin [134, 137, 141, 142]. Noticeably, virions produced by $PPIA^{-/-}$ cells were defective at an early stage of the virus life cycle. However, these data did not pinpoint the exact stage of HIV-1 replication at which CypA acts. Virions produced under conditions that blocked the Gag-CypA interaction exhibited normal endogenous reverse transcriptase activity *in vitro* [141]. This could suggest that virion-associated CypA might be required during the process of virion uncoating [146], or virion binding or fusion to target cells [142]. However, these data do not rule out the possibility that the defect observed before the start of reverse transcription is a consequence of CypA deficiency during the preceding virion assembly process.

Current data on the mode of action of CypA in the HIV-1 life cycle suggests two possible roles that CypA may play in HIV-1 infection. Some data points to CypA playing a crucial role

post cell entry, during the uncoating process while the other data suggests its involvement in cell attachment. These two theories are discussed in the following sections.

1.6.1.1 The role of CypA in the HIV-1 replication cycle

Cyclophilin A is incorporated into nascent virions by specifically binding to the CA domain of the Gag precursor during HIV-1 assembly [129, 134]. Preventing CypA-packaging, either by the addition of cyclosporin A (CsA) to producer cells or by the introduction of mutations in the binding region of CA, inhibits virus infectivity, demonstrating a strict requirement for CypA in HIV-1 replication [134, 137]. As aforementioned, these CypA-deficient viruses are identical to wild-type (WT) particles and the observed block of infectivity of CypA-deficient virus has been mapped prior to the initiation of the viral reverse transcription into target cells [141].

CA dissociates from the viral nucleoprotein complex shortly after entry [147], suggesting that unfolding (or uncoating) of the CA is necessary for productive infection. Given the cellular function of CypA, researchers have proposed that CypA induces HIV-1 uncoating in a manner similar to that of the chaperone heat shock protein in the uncoating of clathrin from coated vesicles [146]. It is suggested that CypA facilitates the proper disassembly of the shell of CA molecules which protects the viral genome. However, anti-CypA antibodies block HIV-1 entry suggesting that CypA does not exclusively serve to uncoat the viral genome, it also play a role during HIV-1 entry [143].

In order to delineate the role of CypA in HIV-1 entry, researchers used viruses that lack CypA to establish whether these viruses fail to enter target cells. Interestingly, a preponderance of

evidence directly implicated CypA in HIV-1 entry [142]. Viruses that lacked CypA were unable to attach to target cells [148]. The data that demonstrated that anti-CypA antibodies prevent HIV-1 replication by blocking the initial step of infection, viral adsorption, further confirmed the role of CypA during cell entry.

The latter finding suggests that CypA incorporated into the virion localize at the viral surface. Corroborating this hypothesis, anti-CypA antibodies immobilized to a solid phase can capture WT viruses but fail to capture CypA deficient viruses [142]. Given that host CypA is specifically incorporated into virus particles via an interior viral protein (Gag), data from Saphire *et al.* suggests that it must subsequently relocate to the surface of the virus [142]. This exposure of CypA at the viral surface is consistent with its suspected role in mediating HIV-1 attachment to the cell surface.

1.6.1.2 The mode of action of CypA in the HIV-1 life cycle during attachment

Based on its peptidyl prolyl *cis-trans* isomerase activity, CypA was initially thought to act as an uncoating factor [146]. In this model, CypA is proposed to interfere with CA-CA interactions thereby mediating the disassembly of the viral core within the cytosol of target cells after cell entry. However, the involvement of CypA in the uncoating process could not be confirmed by other studies [9, 149].

A subsequent study demonstrated that CypA rather, acts prior to the uncoating step by mediating the initial attachment of HIV-1 to target cells [150]. This group, therefore proposed that HIV-1 attaches to target cells via an interaction between CypA and cell surface heparins

(fig. 1.6.1.2), as the first step. In a second step, once the envelope glycoprotein (gp120/gp41) is in close proximity to the plasma membrane, it then interacts with CD4 and chemokine receptors that trigger the fusion process. It is advantageous for HIV-1 to use a host protein, such as CypA, for viral attachment because it is a self-antigen. This may help HIV-1 escape a strong host immune response.

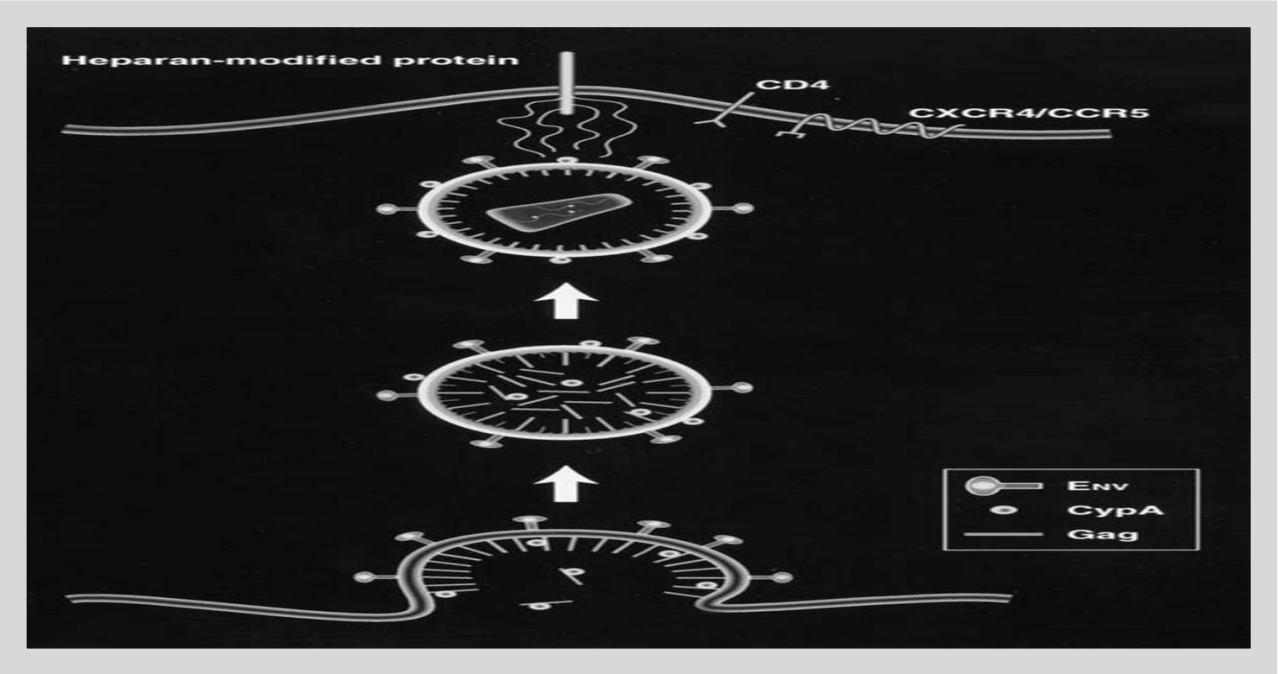

Figure 1.6.1.2. Proposed model for the role of CypA in HIV-1 attachment. Host CypA is incorporated into nascent viruses via HIV-1 Gag. Viral maturation triggers the release of CypA from Gag, permitting the redistribution of CypA at the viral surface. By interacting with cell-surface heparans, exposed CypA mediates HIV-1 attachment to target cells. The initial CypA-mediated attachment permits the subsequent binding between HIV-1 envelope glycoproteins (gp120/gp41) and CD4 and chemokine receptors. This envelope/CD4/chemokine receptor complex triggers the fusion between viral and cellular membranes, allowing the delivery of the viral genome into the cytosol of target cells [150].

This theory suggests that CypA interacts with heparans which are ubiquitously expressed on human cells. This observation was consistent with other studies that had already demonstrated that a number of viruses employ heparans for their initial step of infection [151-156].

1.6.1.3 The role of CypA in HIV-1 replication *in vivo*

Although a consensus has not been reached in terms of at what step of HIV-1 replication does CypA act, it is very clear that CypA promotes HIV-1 replication *in vitro*. The *in vitro* data demonstrated that HIV-1 replication was inhibited in CypA-null human $CD4^+$ T cells (*PPIA*$^{-/-}$ Jurkat cells) [145]. This data demonstrated that CypA is an important cellular cofactor that promotes HIV-1 replication.

Host genetic studies demonstrated an association between a *PPIA* polymorphism and HIV-1 disease outcome [61, 73, 74], thereby validating the *in vitro* data. Bleiber *et al*. demonstrated that the minor variant (G) of the SNP A1650G (1650G), in the *PPIA* promoter region, was associated with higher *ex vivo* virus replication and rapid disease progression in a Swiss Caucasian HIV-1 cohort [61]. In the subsequent study [73], the 1650G was suggested to be associated with increased susceptibility to HIV-1 infection since it was found to be more frequent among HIV-1 positive individuals (the seroconverters (SC)) compared to high-risk exposed uninfected (HREU) cohort. However, the group did not find significant association between the 1650G and disease progression both in the European Americans (EA) and African Americans (AA). Interestingly, they found the minor allele (G) of SNP C1604G

(1604G) which is also in the promoter region of the *PPIA*, to be significantly associated with rapid disease progression both in EA and AA. The latest study on the role of the A1650G and C1604G SNPs upon exposure to HIV-1 infection [74], demonstrated that the frequency of the 1650G was significantly increased in HREU compared to seropositive (SP) group, suggesting that the 1650G may be associated with reduced susceptibility to HIV-1 infection in participants of the Amsterdam Cohort studies (ACS). Lastly, Rits *et al*. found that the 1604G was significantly associated with reduced levels of CypA mRNA expression suggesting that the 1604G may down regulate the expression levels of CypA mRNA [74]. The association of the genetic variation in the *PPIA* with susceptibility to HIV-1 infection remains inconclusive as previous studies found contradictory results [73, 74], which could possibly be attributed to the fact that these studies used populations of different origins. Secondly, the influence of the *PPIA* polymorphisms on HIV-1 clinical outcomes in populations that bear the heaviest burden of HIV infection has not been studied. Therefore, this study investigated the association of genetic variation in the *PPIA* with HIV-1 clinical outcomes in South African HIV-1 study cohorts.

1.6.2 The role of Transportin SR-2 in HIV-1 replication

Transportin-SR2 (TRN-SR2), also known as transportin 3 (TNO3) is a cellular protein that shuttles essential splicing factors–such as serine/arginine-rich proteins–into the nucleus [157, 158]. By convention, the gene and its protein are referred to as *TNPO3* and TRN-SR2, respectively [126]. TRN-SR2 has been shown to play a pivotal role in HIV-1 replication [32, 67, 159].

TRN-SR2 was initially identified as a host factor required by HIV-1 for infection [67, 68]. Subsequently,TRN-SR2 was identified as an HIV-1 IN binding protein and shown to mediate HIV-1 nuclear import [160]. Knockdown of the *TNPO3* yielded significant inhibition of HIV-1 replication indicating that TRN-SR2 is an important HIV-1 replication co-factor [67, 68, 160]. Interestingly, the block to HIV-1 infection was pinpointed at HIV-1 nuclear import, which suggested the role of TRN-SR2 in nuclear import [160]. However, Krishman *et al*. could not confirm the role of TRN-SR2 in HIV-1 nuclear import [32]. This study showed that TRN-SR2 interacts with CA, instead of IN but these results have not been replicated. The most recent study demonstrated that depletion of TRN-SR2 altered the selection of integration sites on the chromosome [161]. Although the exact mode of action for TRN-SR2 in HIV-1 replication cycle is misunderstood, taken together these studies clearly demonstrate that TRN-SR2 is an important HIV-1 replication co-factor.

Studies on the role of TRN-SR2 on HIV-1 replication have been done *in vitro*, meaning that these findings need to be confirmed *in vivo*. At the time of this PhD project, there was no *in vivo* data published on the role that TRN-SR2 may play upon HIV-1 exposure or infection. This project undertook to identify genetic variants in the *TNPO3* gene that may be associated with clinical outcomes upon HIV-1 exposure.

1.6.3 Host factors enhancing HIV-1 integration

Purified pre-integration complexes (PICs) from infected cells preferentially integrate viral DNA into a target cell DNA intermolecularly thereby avoiding suicidal intramolecular autointegration [162]. Secondly they insert both viral DNA ends into a target cell DNA

efficiently, in a concerted manner [162, 163]. Although purified recombinant IN is sufficient to perform 3' processing and strand transfer reactions *in vitro*, it can only insert a single viral DNA end in a single strand of the duplex target DNA [164, 165].

HIV-1 PICs are constituted by viral cDNA, proteins and cellular proteins. Therefore, results obtained from *in vitro* reactions using purified PICs would closely resemble the *in vivo* situation. High salt treated PICs are defective for integration and their activity can be restored upon addition of host cell cytoplasmic extracts [166]. This suggests that the presence of cellular factors in the PIC aids retroviral DNA integration. Amongst the cellular proteins reported to play a role in the HIV-1 integration process are Barrier-to-autointegration factor (BAF), High mobility group chromosomal protein A1 (HMGA1), Integrase interactor 1 (INI1) and Lens epithelium derived growth factor p75 (LEDGF/p75).

BAF forms part of the PIC and directs the integration of viral cDNA into the target DNA, thereby preventing suicidal autointegration [165, 167]. HMGA1 restores the PIC activity after salt-stripping [134, 136]. However, there is no evidence for the requirement of HMGA1 in retroviral integration *in vivo* [166, 168]. Integrase interaction protein 1 (INI1) interacts with HIV-1 IN to enhance its strand transfer reaction [169] and has been shown to be important for HIV-1 replication [170]. LEDGF/p75 is a recently identified binding partner of HIV-1 [171] and is discussed below since it is one of the replication cofactors that were studied in this project.

1.6.3.1 Lens epithelium derived growth factor p75 (LEDGF/p75)

LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF) family. LEDGF/p75 was first identified as a binding partner of the transcriptional co-activator C4, suggesting its role in transcriptional regulation [172]. This protein derived its name 'lens epithelium derived growth factor' from its ability to stimulate and prolong cell survival of lens epithelium cells, cos7 cells, skin fibroblasts and keratinocytes when added to the culture medium [173].

LEDGF/p75 is a survival factor that is involved in promoting mammalian cell growth and protecting cells against stress-induced cell death [174]. LEDGF/p75 provides protection by transcriptionally activating stress-related or anti-apoptotic proteins [175] and is a DNA-binding protein [176]. In addition to its cellular functions, LEDGF/p75 has been identified as the binding partner for HIV-1 IN [171]. This interaction was mapped to the C-terminal domain of LEDGF/p75 [177].

LEDGF/p75 is 530 amino acids long (fig. 1.6.3.1.1). The N-terminal region of LEDGFp75 contains a PWWP (for Pro-Trp-Trp-Pro) domain that is responsible for protein–protein interaction and DNA-binding [178, 179]. The C-terminal region of LEDGF/p75 contains an evolutionary conserved integrase binding domain (IBD) that interacts with HIV-1 IN [180]. LEDGF/p75 also contains a functional nuclear localization signal (NLS) [181, 182]. It is important to note that LEDGF/p75 has a splice variant, LEDGF/p52 (p52), which shares a region of 325 residues of the N-terminus but lacks the C-terminus of LEDGF/p75 [183]. However, p52 does not interact with HIV-1 IN due to lack of the IBD.

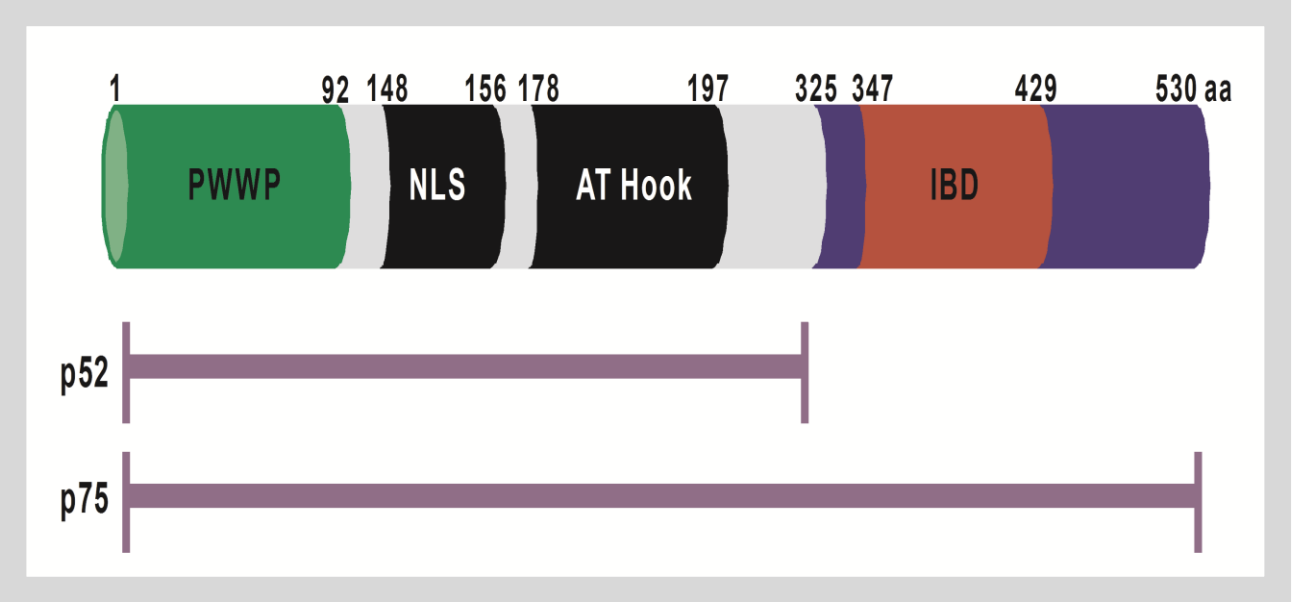

Figure 1.6.3.1.1. Domain structure of LEDGF/p75 [184]. The different domains of p75 and p52 splice variants are highlighted.

The IBD is a compact right-handed bundle composed of 5 α helices as was resolved by nuclear magnetic resonance and the amino acids Ile365, Asp366 and Phe406 were shown to be important for the interaction with HIV-1 IN [185]. In addition, this group reported a crystal structure of the dimeric catalytic core domain of HIV-1 IN complexed to the IBD [186].

The precise stoichiometry of the HIV-1 IN–LEDGF/p75 complex in the infected cell has not been elucidated and it suggests a symmetrical complex containing a pair of integrase tetramers and two subunits of LEDGF/p75 [156] (fig. 1.6.3.1.2). The nuclear distribution of HIV-1 integrase perfectly matches that of LEDGF/p75 [171, 187].

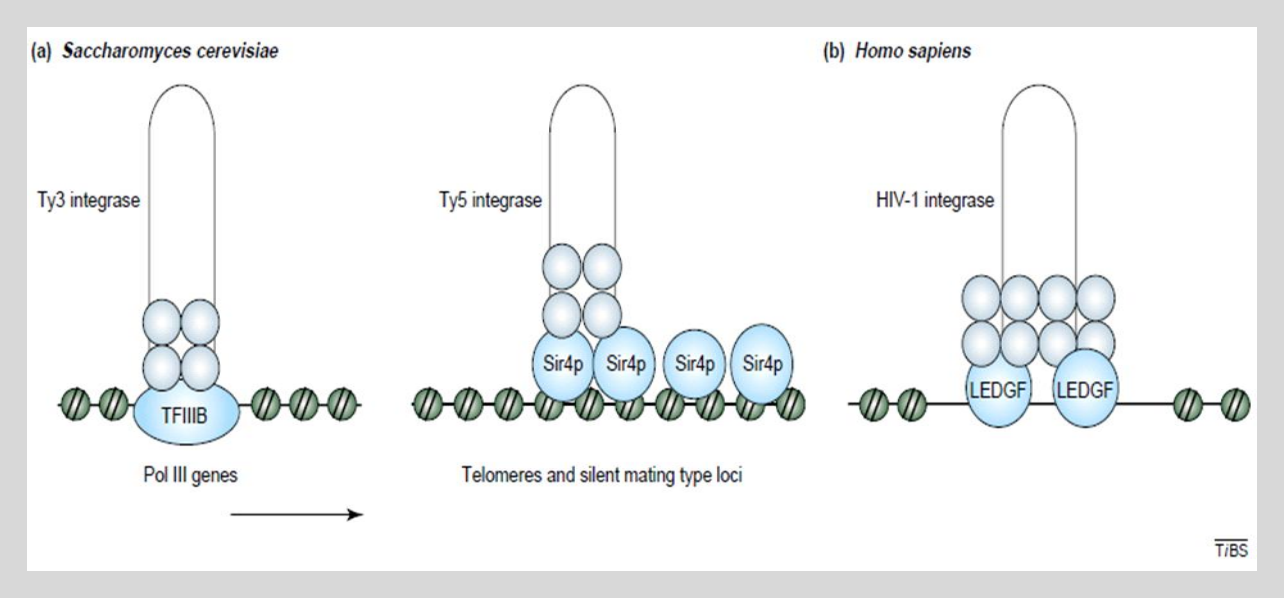

Figure 1.6.3.1.2. Tethering function of LEDGF/p75 in analogy with yeast LTR retrotransposons [184]. **(a)** Ty3 integrates within one or two nucleotides of polymerase III (pol III) transcription initiation sites, probably by interaction with the transcription factors TFIIIB and TFIIIC. Ty5 integrates into heterochromatic DNA of telomeres and silent mating loci (HML and HMR) by interaction with the C-terminal portion of Sir4p. **(b)** Interaction of lentiviral integrase with LEDGF/p75 targets integration into transcriptionally active regions.

Knock-down of endogenous LEDGF/p75 using small interfering RNA (siRNA) completely abolished the nuclear localization of HIV-1 IN and its association with chromosomes in cells [187], suggesting its involvement in HIV-1 nuclear import [188].

1.6.3.2 Role of LEDGF/p75 during lentiviral integration

It is clear from the data presented above that LEDGF/p75 is important for efficient HIV-1 replication. Now the next step was to elucidate the exact function of LEDGF/p75 in the HIV-1

replication life cycle. In direct nuclear-import assays, recombinant HIV-1 IN was still actively imported into the nucleus in the absence of LEDGF/p75 [177]. Surprisingly, mutant HIV-1 IN (Q168A) defective for interaction with LEDGF/p75 failed to associate with the mitotic chromosomes. Moreover, these mutant viruses (Q168A) yielded normal levels of 2-LTR circles but failed to integrate. The results obtained from this study suggested that LEDGF/p75 may not be a dominant factor for HIV-1 nuclear import but might be involved in the integration step of HIV-1 replication cycle.

Subsequently, LEDGF/p75 was found to stimulate the binding of HIV-1 IN to the chromosomal DNA [189]. These results suggested that LEDGF/p75 could be functioning as a tethering factor of HIV-1 IN to the chromosomes. This plausible function explained the nuclear accumulation of HIV-1 IN and its association with mitotic chromosomes, both of which were abolished by knock-down of LEDGF/p75 [177, 188, 190, 191].

HIV-1 is preferentially integrated into transcriptionally active regions, which allows efficient viral gene expression after integration [192]. Therefore, a plausible hypothesis at this time was that LEDGF/p75 played a role in the tethering of IN to the chromosomal DNA and targeting its integration to actively transcribed regions. LEDGF/p75-mediated site selection is reminiscent of target-site selection by yeast LTR transposons Ty3 and Ty5, which interact with transcription factor IIIB and Sir4p, respectively [193, 194].

1.6.3.3 The role of LEDGF/p75 in lentiviral integration targeting

Knockdowns of LEDGF/p75 showed a reduced integration frequency of HIV-1 within transcription units and an increase in the G/C content around sites of HIV-1 integration [195]. These data confirmed that LEDGF/p75 tethers HIV-1 to the chromatin and directs integration into active genes (fig. 1.6.3.3). In support of the tethering model, artificial fusion proteins in which the LEDGF/p75 IBD was fused to the sequence specific for DNA binding domain of phage lambda repressor favoured integration near repressor binding sites *in vitro* [196]. HIV-1 binds to the IBD which is in the C-terminal then the PWWP domain which is in the N-terminal region of LEDGF/p75 tethers HIV-1 to the host chromosome thereby promoting HIV-1 replication [197]. This suggests that LEDGF/p75 needs both ends of LEDGF/p75 to promote HIV-1 replication.

However, key questions still remain on the role of LEDGF/p75. In all the models studied, HIV continued to favor integration within active transcription units. This could either be because residual LEDGF/p75 remaining in the knockdown was sufficient for residual targeting activity, or because additional host cell factors also contributed independently to targeting HIV-1 integration.

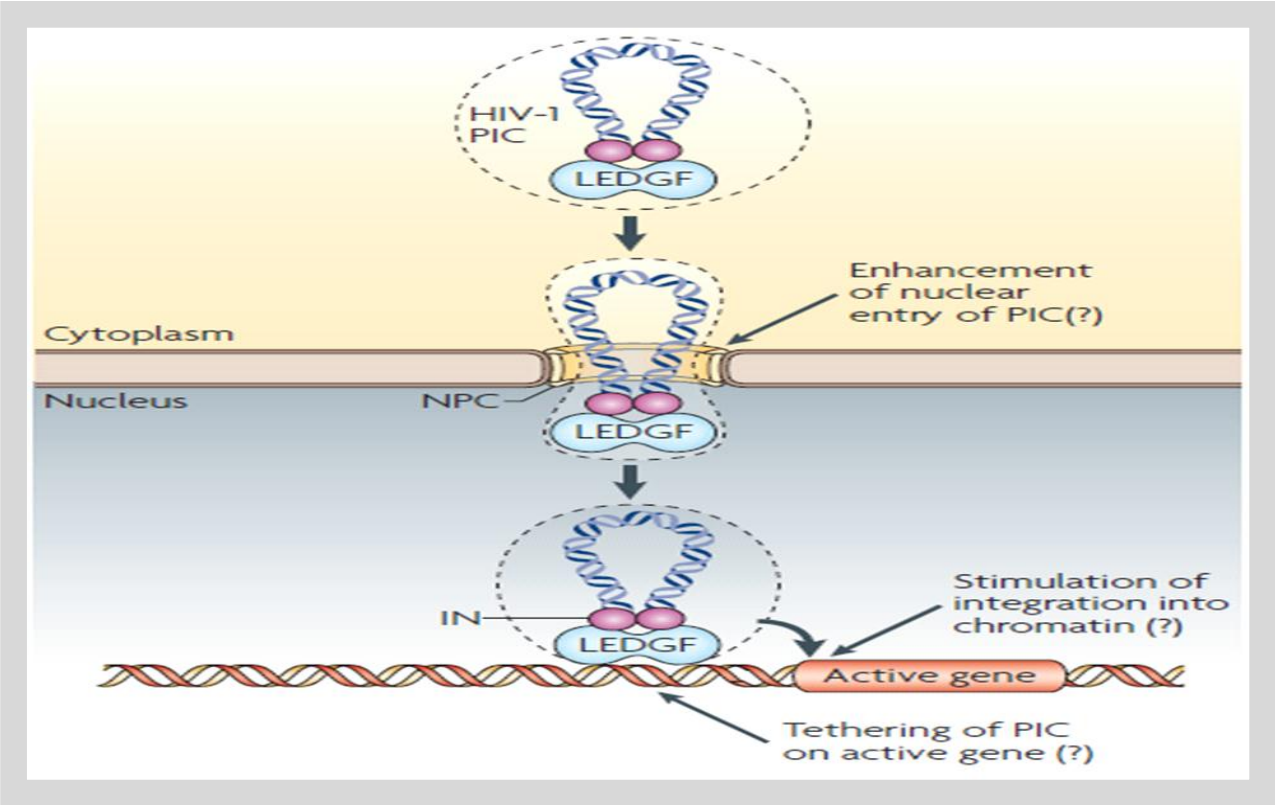

Figure 1.6.3.3. LEDGF/p75 and nuclear entry of PICs. LEDGF/p75 might regulate HIV-1 replication through the tethering the IN to the chromatin and targeting integration into active genes [198]. NPC→ nuclear pore complex; PIC→ pre-integration complex.

In an effort to address this issue, Shun *et al*. prepared a mouse strain that had part of the LEDGF/p75 locus flanked by *Cre* recombination sites [199], and the LEDGF/p75 exon regions deleted by exposure to *Cre* recombinase. Mouse embryonic fibroblasts were then studied for effects on infection with HIV-1 reporter viruses. These cells showed a 20-fold reduction in infectivity by HIV-1, and also a reduction in integration frequency in transcription units [195]. In addition, these mouse cells showed some new targeting features in the LEDGF/p75-depleted cells, including increased integration near CpG islands.

The role of LEDGF/p75 in tethering the IN to the host chromosome and targeting integration in transcription units was confirmed in human SupT1 T-cell line with intensified RNAi knockdown of LEDGF/p75 [197]. The mouse cells containing homozygous gene trap mutations at the LEDGF/p75 locus also confirmed the tethering effect of LEDGF/p75 [200]. Taken together these data demonstrated that LEDGF/p75 is a determinant of integration target site selection for HIV-1 and other lentiviruses in primary cells where LEDGF/p75 levels were not artificially reduced.

1.6.3.4 Retargeting lentiviral integration into heterochromatin using LEDGF constructs

Overexpression of the C-terminal fragment of LEDGF/p75 (amino acid 325–530; LEDGF325–530) or IBD alone did not only fail to mediate chromatin binding but also it relocated HIV-1-IN to the cytoplasm and blocked HIV-1 replication [201, 202]. The mechanism of chromatin association is believed to be facilitated by elements located in the N-terminal portion of LEDGF/p75. These include a PWWP domain, which contains a Pro-Trp-Trp-Pro signature, a NLS and two AT hooks [197, 203] (fig. 1.6.3.4). It appeared that PWWP domain of LEDGF/p75 was responsible for chromatin binding and integration targeting and therefore Lin *et al*. constructed fusions of IBD and the λ repressor DNA–binding domain to test this. Interestingly, this fusion protein showed increased *in vitro* strand transfer activity near λ repressor–binding sites [166], implying that indeed PWWP is responsible for tethering and integration targeting effect of LEDGF/p75.

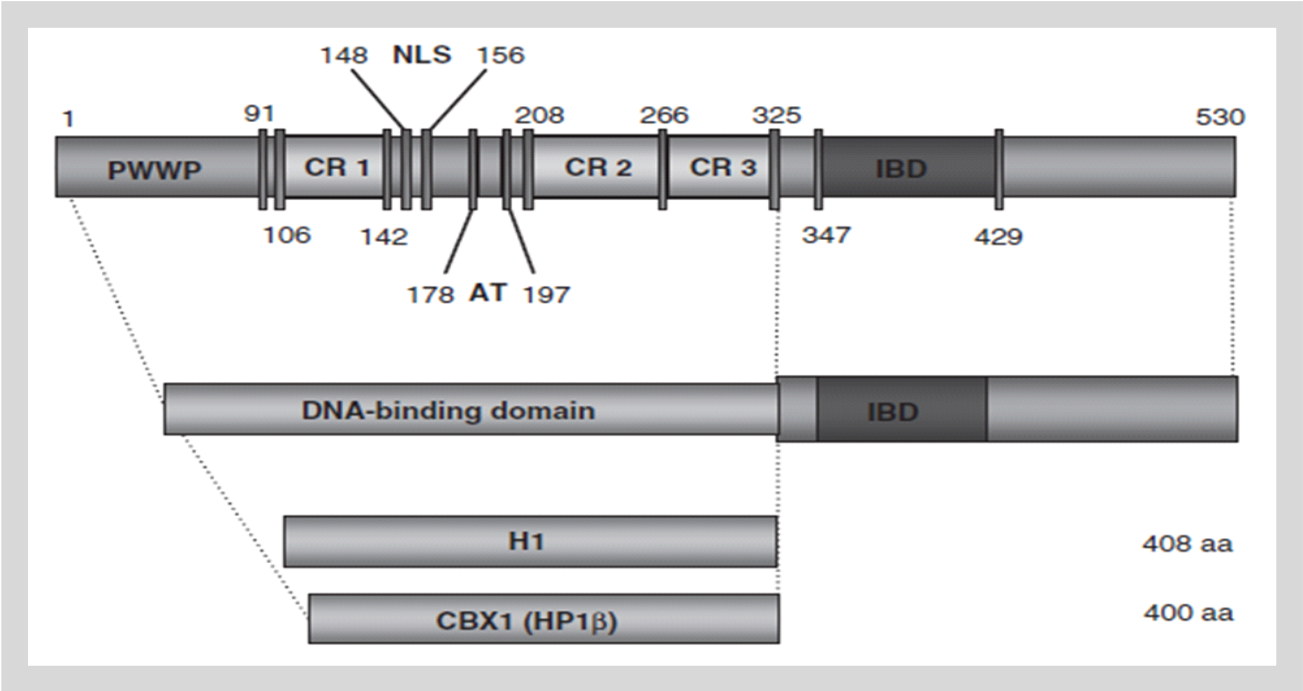

Figure 1.6.3.4. Domain structure of LEDGF/p75 and schematic representation of LEDGF325–530 fusions. In the lower panel the DNA binding domain fusions with LEDGF325–530 are depicted, H1-LEDGF325–530, and CBX1-LEDGF325–530, respectively. Numbers indicate amino acids of each domain [204].

Fusion of the IBD and alternative chromatin-binding proteins retargeted lentiviral integration to alternative regions of the genome, when it was expressed in LEDGF/p75-depleted cells [204]. These findings further confirmed that the PWWP domain of LEDGF/p75 is responsible for integration targeting. Gijsbers*et al*. developed fusion proteins (hybrids) between the IBD and different chromatin-binding proteins, with particular focus on domains with binding specificities that might be useful during human gene therapy [204].

A fusion protein where the PWWP domain of LEDGF/p75 was replaced with CBX1 rescued the infection block in LEDGF/p75-depleted cells but the integration was retargeted to genomic sites bound by CBX1 [204]. Normally, these regions are disfavoured for lentiviral integration because their expression levels are very low. Interestingly, transgene expression from the vector was nevertheless efficient. These findings open possibilities of using gene therapy vectors to retarget IN integration to gene-poor regions thereby inhibiting HIV-1 replication [204].

1.6.3.5 Small-molecule inhibitors (SMI) of the LEDGF/p75-integrase interaction and HIV replication

The data presented above support the theory that LEDGF/p75 tethers the HIV-1 preintegration complex (PIC) to cellular chromatin. Analysis of integration sites in human cells depleted for LEDGF/p75 by RNAi or in embryonic fibroblasts derived from LEDGF/p75 knockout mice, corroborated the role of LEDGF/p75 as tethering and integration targeting factor of HIV-1 [195, 205].

Since LEDGF/p75-IN interaction has been shown to be important for efficient and productive HIV-1 infection, this interaction could be explored as a potential antiviral target. Overexpression of the C-terminal fragment (Δ325) of LEDGF/p75 in human cells resulted in the inhibition of HIV-1 replication [201]. In this study, the Δ325 containing no PWWP domain outcompeted the endogenous LEDGF/p75 for binding to the IN and this inhibited HIV-1 replication to nearly undetectable levels (fig. 1.6.3.5). HIV-1 inhibition was due to the fact that

the Δ325 did not contain the PWWP domain and as a result was unable to tether HIV-1 to the host chromosome thereby inhibiting HIV-1 replication.

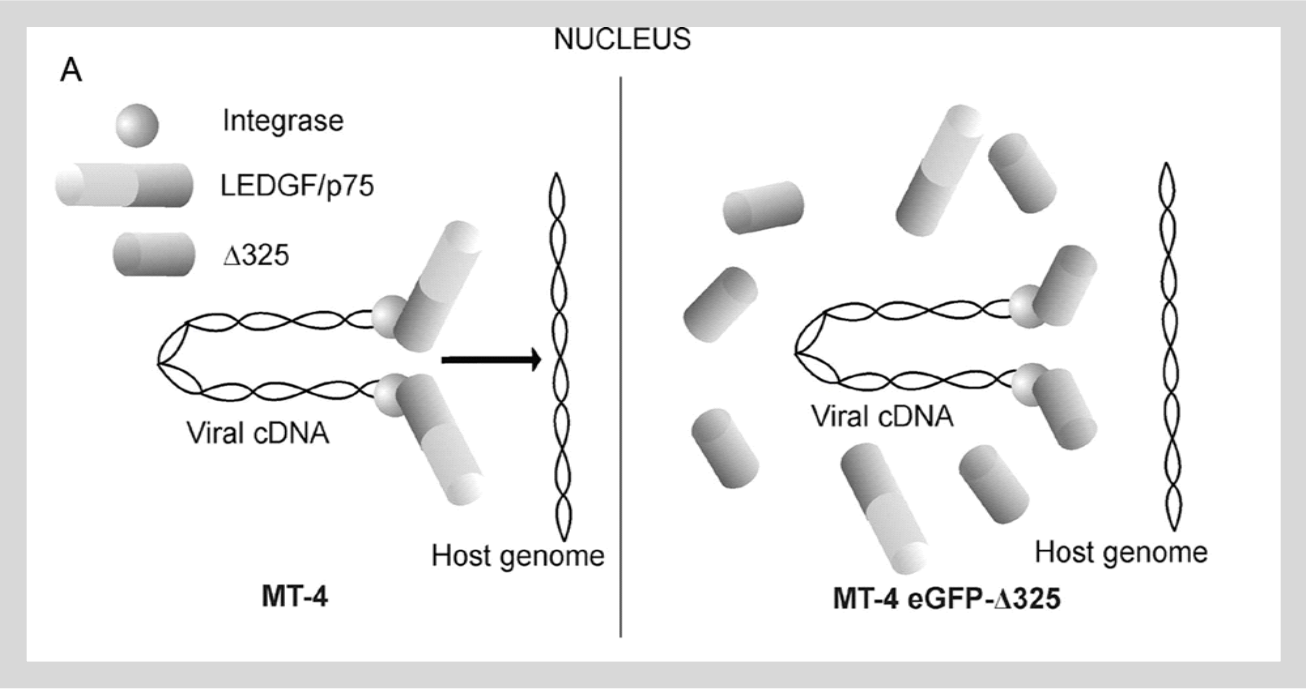

Figure 1.6.3.5. Overexpression of the C-terminal domain of LEDGF/p75 (Δ325) outcompetes the endogenous LEDGF/p75 for binding to HIV-1 IN thereby inhibiting HIV-1 replication because it lacks (PWWP) chromatin binding domain [206].

Another study repeatedly passaged HIV-1 in cells overexpressing this Δ325, thereby causing HIV-1 to develop resistance against this phenotype [202]. Notably, two IN mutations (A128T and E170G) that are crucial for IBD-IN interaction, rendered IN resistant [207]. Subsequently, Christ *et al.* embarked on a rational drug design program to discover LEDGINs, small

molecule inhibitors that targeted the LEDGF/p75-IN interaction in order to inhibit HIV replication [208].

Protein-protein interactions (PPIs) between HIV-1 and host proteins (replication cofactors) constitute a pool of potential new antiviral targets against which small-molecule protein-protein interaction inhibitors (SMPPIIs) can be designed. A new class of antiviral agents that effectively target the LEDGF/p75-IN interaction has been designed and synthesized, not only demonstrating the feasibility of exploiting PPIs as targets for drug discovery but also validating a new paradigm in anti-HIV research [208].

Christ *et al*. found 2-(quinolin-3-yl) acetic acid derivatives that blocked integration by inhibiting LEDGF/p75-IN interaction [209]. The 2-(quinolin-3-yl) acetic acid derivatives also moderately inhibited the enzymatic activity of HIV-1 IN. This did not come as a surprise since it is well known that LEDGF/p75 by itself act as an allosteric activator of IN activity. Although all compounds tested were more active on the cofactor binding than on the enzymatic activity, this allosteric function might add to the potent antiviral profile of this compound class.

1.7 Conclusion

AIDS has had a devastating effect on human health killing more 25 million people since its description in 1981 while there are about 33.4 million people living with HIV-1 worldwide (http://www.unaids.org).Despite the enormous efforts in developing new effective antiviral agents and the introduction of highly active antiretroviral therapy (HAART), the problem of HIV-1 infections persists. Although HAART is effective in chronically suppressing HIV-1

replication, antiretroviral treatment has been accompanied by development of resistance and toxicity in some patients. Therefore, the quest for developing new and safe antiviral agents to complement existing treatment strategies remains one of the main goals in HIV-1 drug discovery. Current antiviral drugs target the viral enzymes RT, PR and IN. For example, raltegravir (MK-518) targets the strand-transfer reaction of viral IN and inhibits the integration process thereby reducing viral replication [210, 211]. However, raltegravir resistance evolves readily in the clinic [212], necessitating the efforts to develop second-generation integrase inhibitors with beneficial resistance profiles.

HIV-1 relies on replication cofactors for completion of its replication cycle and productive infection. Therefore, the second generation of HIV-1 inhibitors should target the interaction between viral and cellular proteins. Disruption of the interaction between HIV-1 and cellular proteins could potentially reduce viral resistance and block HIV-1 replication.

HIV-1 replication cofactors are increasingly being targeted for antiretroviral therapy yet there are three fundamental challenges facing this approach: (1) there is very little information on genetic variation in the replication cofactors in humans and its impact if any upon HIV-1 exposure; (2) studies on human genetic factors that affect HIV/AIDS pathogenesis have focused mainly on host immune factors and the failure of vaccines and immunotherapies in HIV-1 and; (3) host genetic studies have been conducted using populations from developed countries and of European origin yet genetic milieu of populations differ according to geographic location.

These challenges necessitate the need to explore other biomedical interventions that may be pursued for therapeutic and prophylaxis intervention using populations from developing countries, heavily burdened by HIV-1 epidemic. The current study addresses the latter challenge by studying the influence of genetic variation in select HIV-1 replication cofactors–CypA, TRN-SR2 and LEDGF/p75–on susceptibility to HIV-1 infection and disease progression using a South African population. This study seeks to shift the focus from immune factors and look at genetic variation among South Africans who bear the greatest burden of infection as opposed to previous studies on mainly Caucasians or African Americans who may have a relatively low HIV/AIDS burden and different genetic milieu for disease genes.

CHAPTER TWO

2.1 Molecular Methods

Understanding the patterns of single nucleotide polymorphisms (SNPs) is essential for the studies of the disease [213-216], in preventive medicine [217, 218], personalized medicine [219], forensics [220] and evolution [221]. Therefore, many SNP genotyping methods have been developed [222-224]. However, in this thesis we are discussing only three genotyping methods: (1) DNA sequencing; (2) polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis and; (3) Taqman probe assay. In addition to genotyping methods, this thesis gives an overview of the AlphaScreen Assays since these techniques were used in this project to investigate genetic variation in relation to HIV pathogenesis and to measure protein-protein interactions.

2.2 DNA Sequencing

The introduction of polymerase chain reaction (PCR) advanced genetic research, leading to improved and faster DNA sequence analysis (DNA sequencing) and identification of different genes based on their DNA sequences. Two DNA sequencing techniques, Sanger (or dideoxy) method [225] and Maxam-Gilbert (chemical cleavage) method [226], had been developed in the late 1970's. The Sanger method is technically easier to apply, and, with the advent of PCR and automation of the technique, is easily applied to long strands of DNA including some entire genes. This technique is based on chain termination by dideoxy nucleotides during PCR elongation reactions.

The Maxam-Gilbert method is based on nucleotide-specific cleavage by chemicals and is best used to sequence oligonucleotides, usually smaller than 50 base-pairs in length. Gel

electrophoresis technique with the ability to separate DNA fragments that differ in size by as little as one base pair forms an integral part of DNA sequencing. The Sanger (or dideoxynucleotide chain termination) method was therefore the method of choice in this study, given its simplicity and applicability.

2.2.1 Sanger Method (Dideoxynucleotide chain termination)

The sequencing reaction of the Sanger method consists of DNA template, single oligonucleotide primer, deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) [183]. The sequencing reaction proceeds in the following steps: (1) denaturation of the DNA template; (2) annealing of oligonucleotide primer to single stranded DNA template; (3) DNA polymerase then starts extending the primer by incorporating dNTPs until, if by any chance, a ddNTP is incorporated instead of dNTP. This is a "chain termination" event, because ddNTPs lack the 3’ OH group to which the next dNTP of the growing DNA chain is added. Without the 3’ OH, no more nucleotides can be added, and DNA polymerase falls off [225] (fig. 2.2.1).

Higher concentrations of the ddNTPs in the reaction generate shorter sequencing products that are closer to the oligonucleotide primer. Lower concentrations of ddNTPs, on the other hand, result in longer products that are further away from the oligonucleotide primer [227].

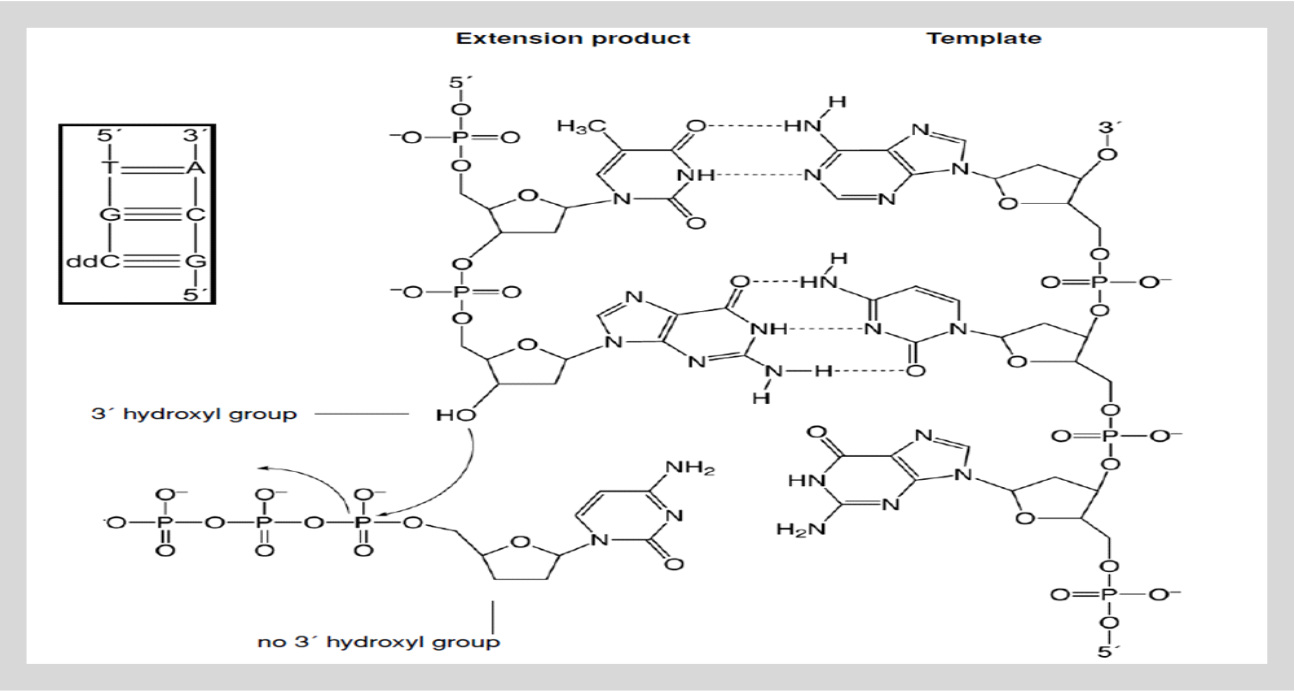

Figure 2.2.1. A schematic representation of sequencing reaction. DNA strand is synthesized by formation of phosphodiester bonds. In this example of the “C reaction”, the chain is terminated by the use of dideoxycytidine triphosphate (ddCTP) in place of deoxycytidine triphosphate (dCTP) (www.appliedbiosystems.com; 11 October 2010).

However, newly synthesized products are a mixture of lengths, depending on how long the chain was when a ddNTP was randomly incorporated. Since the newly synthesized DNA products are labeled (classically with 35S-dATP) then these products can be detected and distinguished from the template because template DNA is not labeled.

Before the invention of automated sequencing machines, these products were denatured into single stranded DNA molecules and run on a polyacrylamide/urea gel. Unlike the agarose gel, polyacrylamide gels allow resolution of DNA molecules that differ in size by only one

nucleotide [227]. The gel would be then dried onto chromatography and exposed to X-ray film. The template DNA is distinguished from synthesized products by the fact that the template strand is not radioactively labeled and therefore, does not generate a band on the X-ray film but newly synthesized products which are labeled strands do generate bands (fig. 2.2.1.1 A).

The chain termination events, as aforementioned, produce DNA or sequencing products of different lengths which results in individual bands on a gel as demonstrated using the "G" reaction, as an example (fig. 2.2.1.1B).

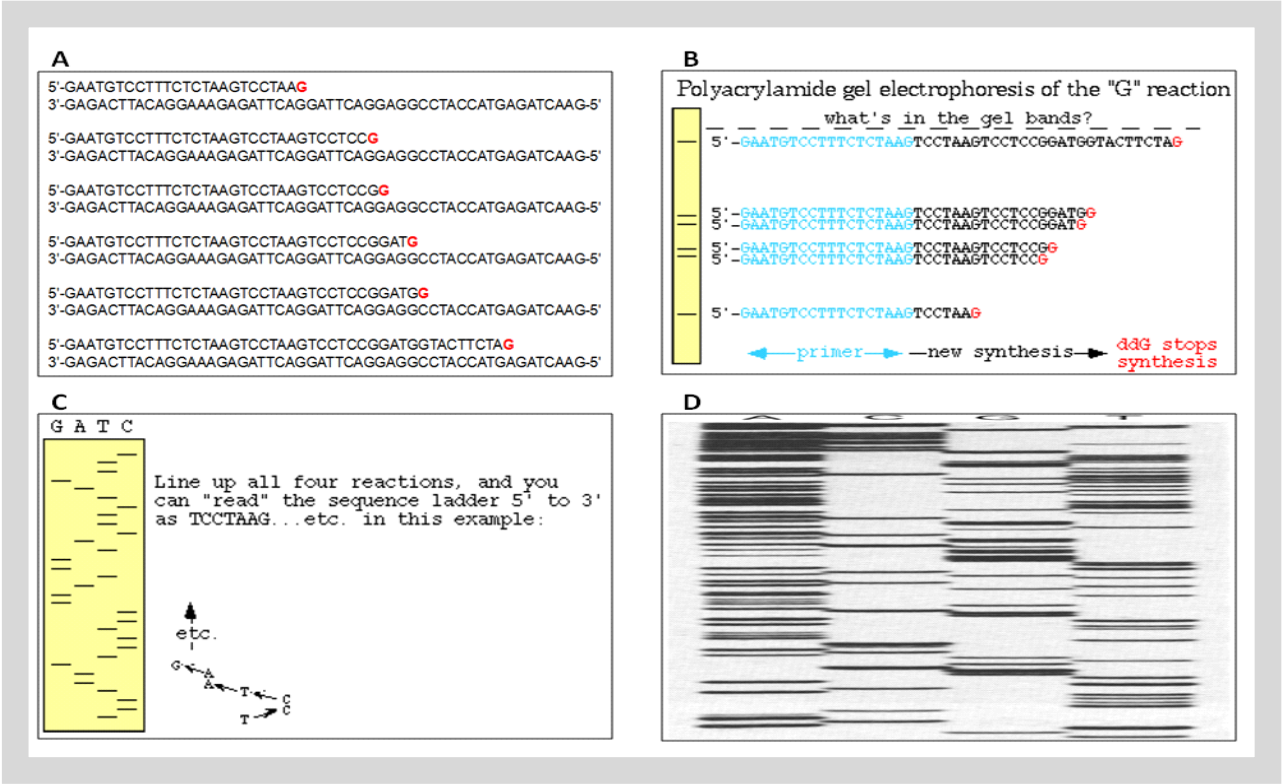

Figure 2.2.1.1. Sequencing products of different lengths. **A**, Mixture of products of different lengths synthesized for the "G" reaction. Each newly synthesized strand at some point had a ddGTP incorporated instead of dGTP resulting in chain termination. **B**, Polyacrylamide gel showing sequencing products of different lengths. Chain terminations closest to the primer generate the smallest DNA molecules that migrate the longest distance down the gel and chain terminations further from the primer generate larger DNA molecules that remain nearer to the top. **C**, Chain termination reactions for all four nucleotides are run next to each other to read the sequence of the DNA off of the "ladder" from bottom to top. **D**, Large polyacrylamide gel allows DNA molecules to migrate further and be better resolved (www.appliedbiosystem.com; 11 October 2010).

Different sizes of newly synthesized products (fragments) migrate different distances on a gel during gel electrophoresis which generate different banding patterns. The shortest (smallest DNA fragment) product migrates further down on the gel whereas the largest DNA molecule migrates slower on the gel during electrophoresis, remaining nearer to the top of the gel [228].

The chain termination reactions could be performed for all four nucleotides (A, C, G and T) concurrently, and when these four reactions are run next to each other on a gel, then the sequence of the DNA could be read off of the "ladder" of bands, 5' to 3' sequence being read from bottom to top (fig. 2.2.1.1C) [229].

The resolution of the gel electrophoresis is very important in DNA sequencing. DNA molecules that are 50, 100, or 200 bases in length must be separable from the DNA molecules that are 51, 101, or 201 bases in length, respectively. Large polyacrylamide gels are therefore used to separate DNA molecules that differ in length by either one or a few nucleotides because these gels allow the molecules to migrate further and be better resolved (fig. 2.2.1.1D) [230]. Firstly, samples must be denatured before they are loaded onto the gel, and secondly these polyacrylamide gels should contain a high concentration of urea (7 to 8 molar) to prevent folding of the molecules and formation of secondary structures by hydrogen (H) bonding that would alter the mobility of the molecule [231]. Formation of H bonds is further prevented by running these gels at higher temperature (about 50 ºC).

Increased ddNTP:dNTP ratio results in more sequences closer to the primer making it difficult to read sequences of 200 to 300 nucleotides further down, since most of the synthetic products would have terminated earlier.

2.2.2 Automated DNA sequencing: Dye termination sequencing

The Sanger method chain termination reactions are still used, but pouring, running, and reading polyacrylamide gels have been replaced by automated methods. Instead of labeling the products of all 4 sequencing (chain termination) reactions the same, each ddNTP is labeled with a different fluorescent marker (fig. 2.2.2A).

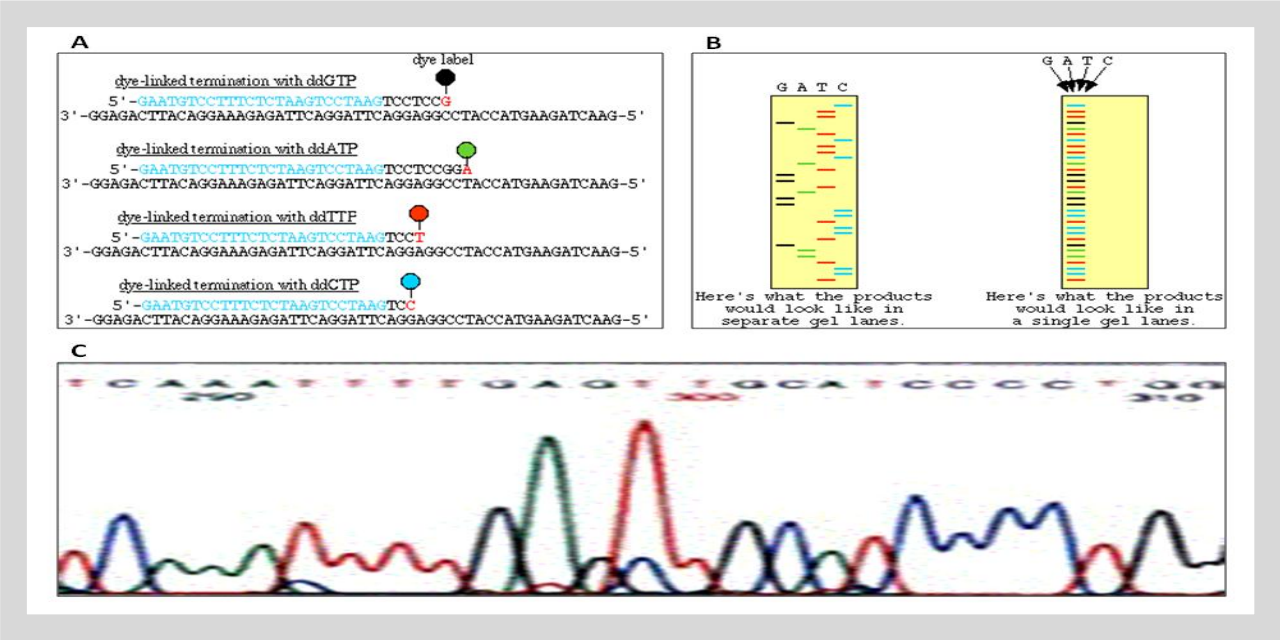

Figure 2.2.2. Automated DNA sequencing: Dye termination sequencing. **A,** ddNTP labeled with a different fluorescent marker, **B,** gel on the left shows chain termination reactions run individually and gel on the right shows four chain termination reactions that were performed in the same tube, and run on a single lane. **C,** chromatogram showing fluorescence intensity translated into data peaks (www.appliedbiosytems.com; 11 October 2011)

The laser beam of the automated DNA sequencing machine, excite four different kinds of products which get detected and the fluorescence intensity of these products is translated into a data "peak" [232].

Thus all four chain termination reactions can be performed in the same tube, and run on a single lane on a gel (fig. 2.2.2B). The sequencing machine scans the lane with a laser and the wavelength of fluorescence from the label conjugated to the ddNTPs can be interpreted by the sequencing machine as an indication of which reaction–ddG, ddA, ddT, or ddC–a particular DNA band came from. The fluorescence output is stored in the form of a chromatogram (fig. 2.2.2C). Although different automated DNA sequencing approaches exist, in this study we used the Applied Biosystems (ABI) (www.appliedbiosystems.com) approach to sequence DNA obtained from the participants of this study. This approach uses the ABI ready reaction mix to sequence the DNA.

2.2.3 ABI PRISM™ Dye Terminator Cycle Sequencing Kits

The ABI ready reaction mix used for sequencing (chain termination) reaction contains AmpliTaq® DNA polymerase, fluorescent sequencing (FS), labeled dye ddNTPs/primers, dNTPs, rT*th* pyrophoshatase, magnesium chloride and buffer.

AmpliTaq® DNA Polymerase, FS is a mutant form of *Thermus aquaticus* (*Taq*) DNA polymerase and contains a point mutation in the active site of *Taq* DNA polymerase, replacing phenylalanine with tyrosine at residue 667 (F667Y) [233]. This point mutation

(F667Y) reduces discrimination against dideoxynucleotides and leads to a much more even peak intensity pattern [234]. The N-terminal domain of AmpliTaq® DNA Polymerase, FS possess a point mutation that results in glycine to aspartate change at residue 46 (G46D), which removes almost all of the 5´→3´ nuclease activity. This eliminates artifacts that arise from the exonuclease activity. The enzyme has been formulated with a thermally stable inorganic pyrophosphatase that cleaves the inorganic pyrophosphate (Ppi) byproduct of the extension reaction and prevents its accumulation in the sequencing reaction.

In the presence of high concentrations of Ppi the polymerization reaction can be reversed [235], a reaction called pyrophosphorolysis. In this reaction, a nucleoside monophosphate is removed from the extension product with the addition of Ppi to form the nucleoside triphosphate. In a sequencing reaction, if a dideoxynucleotide is frequently removed at a particular position and replaced by a deoxynucleotide, eventually there is little or no chain termination at that location. This would result in a weak or missing peak in the sequence data [236]. With dye terminator labeling, each of the four dideoxy terminators (ddNTPs) is tagged with a different fluorescent dye. The growing chain is simultaneously terminated and labeled with the dye that corresponds to that base (fig. 2.2.3).

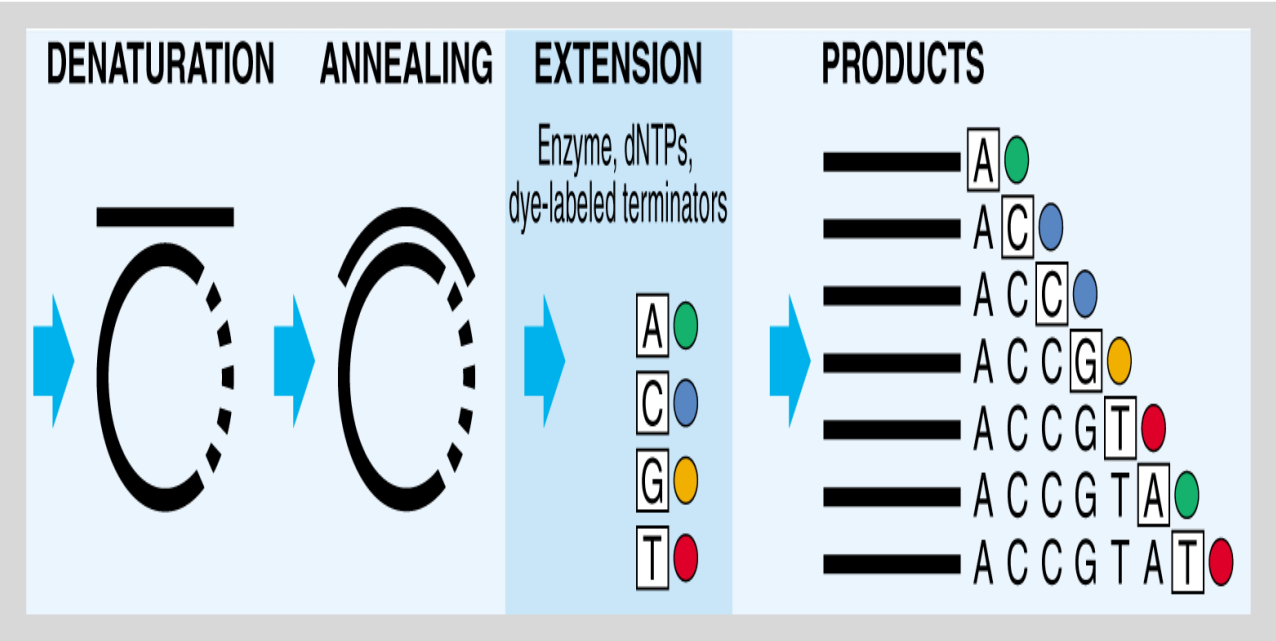

Figure 2.2.3. One cycle of dye terminator cycle sequencing. Features of Dye-labeled Terminator Reactions: An unlabeled primer can be used. Dye terminator reactions are performed in a single tube. They require fewer pipetting steps than dye primer reactions. Four-color dye labeled reactions is loaded in a single gel lane or capillary injection. False stops, *i.e.*, fragments that are not terminated by a dideoxynucleotide, go undetected because no dye is attached (www.biosystems.com, May 2011).

Dye-labeled primer reactions produce more even signal intensities than dye terminator chemistries. Labeled primers are available for common priming sites and customer designed primers can also be labeled. Four-color dye-labeled reactions are loaded onto a single lane or capillary injection.

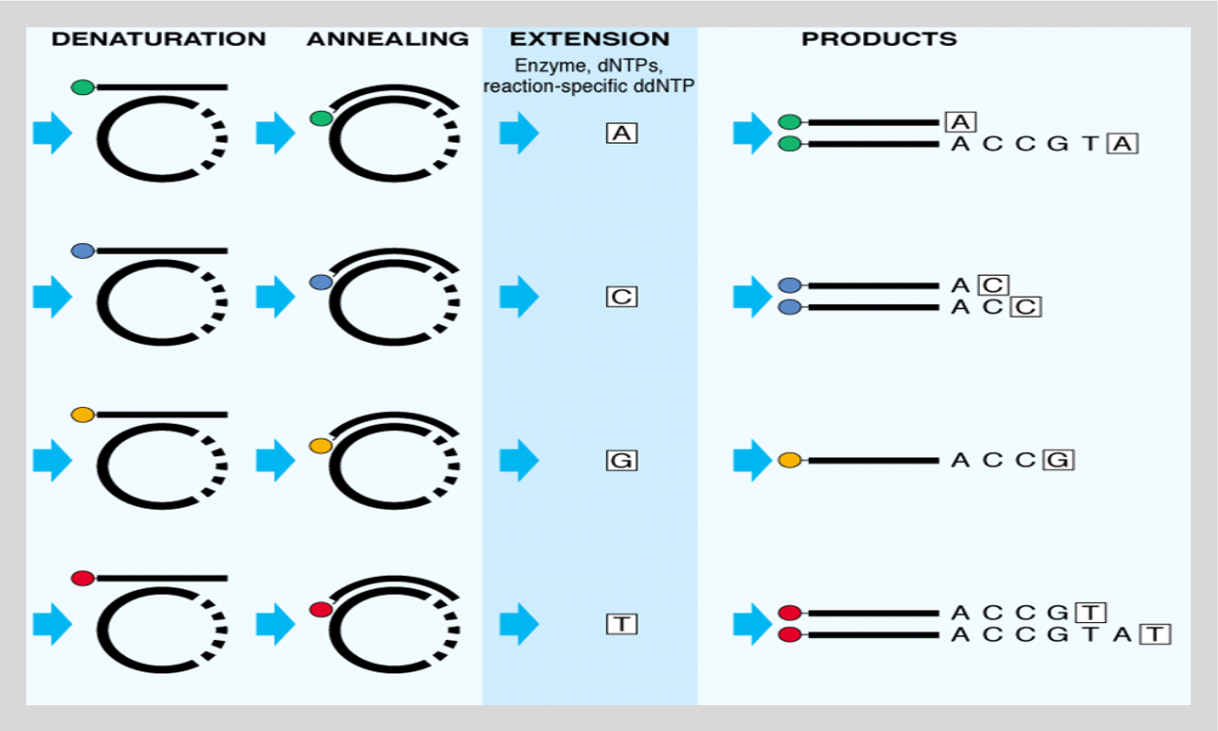

Figure 2.2.3.1. One cycle of dye primer cycle sequencing. Features of Dye-labeled Primer Reactions: Dye primer chemistries generally produce more even signal intensities than dye terminator chemistries. Labeled primers are available for common priming sites. Custom primers can also be labeled. Four-color dye-labeled reactions are loaded onto a single lane or capillary injection (www.biosytems, May 2011).

Although either approach could be used, depending on the objectives of the experiment, dye-terminator reactions approach has advantages over dye primer approach. These include: (1) use of unlabeled primer and performing reactions in a single tube; (2) fewer pipetting steps; and (3) false stops, that is, fragments that are not terminated by a dideoxynucleotide go undetected because there is no dye attached.

2.2.4 Sequencing Chemistry

During the development of a new sequencing chemistry, alternative dye/base relationships were investigated to see which produces the most uniform signal in the analyzed data. For this reason, different sequencing chemistries may have different dye/base relationships. Three different sequencing chemistries were developed [237]: (1) Rhodamine dye terminators, which utilizes terminator dye label (A - R6G, C – ROX, G - R110, T – TAMRA); (2) dichloroRhodamine (dRhodamine) terminators, also utilizes terminator dye label (A - dichloro[R6G], C - dichloro[TAMRA], G -dichloro[R110], T - dichloro[ROX]); and (3) BigDye™ terminators, which uses the terminator acceptor dye (A - dichloro[R6G], C - dichloro[ROX], G - dichloro[R110], T - dichloro[TAMRA]).

In this PhD project we used BigDye^{TM} terminator sequencing chemistry because of the following reasons: BigDye^{TM} terminator chemistry is: (1) 2–3 times brighter than the rhodamine dye terminators; (2) have a narrower emission spectra than the rhodamine dye terminators, giving less spectral overlap and therefore less noise; and (3) has the brighter signal and decreased noise which provide an overall 4–5X gain in signal-to-noise ratio [238]. In addition, the dNTP mix used in BigDye^{TM} terminator sequencing chemistry includes dITP in place of dGTP to minimise band compressions. The dNTP mix also uses dUTP in place of dTTP and the dUTP is known to improve the incorporation of the T terminator resulting in a better T pattern. Applied Biosystems BigDye^{TM} terminator cycle sequencing chemistry provides a comprehensive solution for today's wide range of sequencing applications. For applications at every throughput level, these robust chemistries give longer reads and the

highest data quality ever. They successfully read through challenging motifs and challenging templates.

2.2.5 Genetic Analyzers

Sequencing machines referred to as genetic analyzers are used to analyze fluorescently labeled DNA fragments generated by sequencing reactions. While there are different series of genetic analyzers, this section briefly discusses 3100 series of the genetic analyzers focusing particularly on the ABI^{TM} 3130XL genetic analyzer since it was used in this PhD project.

ABI 3100 series use virtual filter sets to detect light intensity in four non-overlapping regions on a CCD camera [239]. Each region corresponds to a wavelength range that contains or is close to the emission maximum of an ABI PRISM dye used in the sequencing reaction. This process is similar to using a physical filter to separate light of different wavelengths, however, the instruments do not use physical filtering hardware hence they are called virtual filters. The intensity display from the four light-collection regions is color coded by data collection software to appear as the blue, green, black (yellow on gel images), and red peaks in the raw data. The software always displays analyzed data with A as green, C as blue, G as black, and T as red in the electropherogram view.

The ABI 3130XL Genetic Analyzer is a 16-capillary, fluorescence-based capillary electrophoresis system [240]. In addition to being a DNA sequencer it can run a wide variety of sequencing and fragment analysis applications including microsatellite analysis, amplified

fragment length polymorphisms (AFLP), loss of heterozygosity (LOH), SNP validation, and SNP screening. ABI 3130XL Genetic Analyzer can perform *de novo* sequencing and resequencing (mutational profiling). This system incorporates an automated polymer delivery process. This eliminates the need for manual filling, priming and loading of polymer, reducing time and decreasing run to run variation. This versatile platform delivers performance with higher data quality, improved automation and ease-of-use, faster turnaround times and higher reliability across the complete range of sequencing, resequencing, and fragment analysis applications. The full range of applications can be run on a single polymer and capillary array meaning applications could be mixed on one plate (www.appliedbiosystems, August 2010).

The ABI 3130XL Genetic Analyzer has DNA Sequencing Analysis Software v5.1 integrated into it. This software is designed to analyze, display, edit, save and print sample files generated from ABI 3130XL Genetic Analyzer. The software contains a novel basecaller algorithm that performs base calling for pure and mixed base calls, it generates quality values to provide basecall accuracy information for pure and mixed base calls. This software also gives the analysis report that helps to troubleshoot and provide easy assessment of data quality.

2.3. Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

Single nucleotide polymorphisms (SNPs) usually occur once every several hundred base pairs [241]. The ability to accurately determine the DNA sequence at specific sites throughout the individual's genome is important for a number of applications including disease detection, personalized medication, etc. (see section 2.1). The human genome project reported over 10 million SNPs that occur in the human genome (www.ornl.gov). A number of rapid and reliable methods or techniques have been developed to analyze DNA for the purpose of detecting DNA variation or SNPs.

2.3.1. Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a technique that allows production of large quantities of a specific DNA sequence in a simple enzymatic reaction *in vitro* [242, 243]. PCR has transformed the way that almost all studies requiring the manipulation of DNA fragments may be performed as a result of its simplicity, usefulness and cost effectiveness [244]. PCR generates large quantities of a target sequence which are enough for downstream analysis. Unique oligonucleotide primers complimentary to the ends of the target sequence to be amplified need to be designed. These primers and the DNA template containing the target sequence are mixed with commercially available PCR reagents: PCR buffer; magnesium chloride ($Mgcl_2$); dNTPs; enzyme (DNA polymerase); and appropriate buffer. This reaction mixture is then subjected to repeated cycles of denaturation of the original strands of DNA,

primer annealing to the opposite ends of the denatured target DNA, and extension of primers by DNA polymerase which adds dNTPs to extend primers.

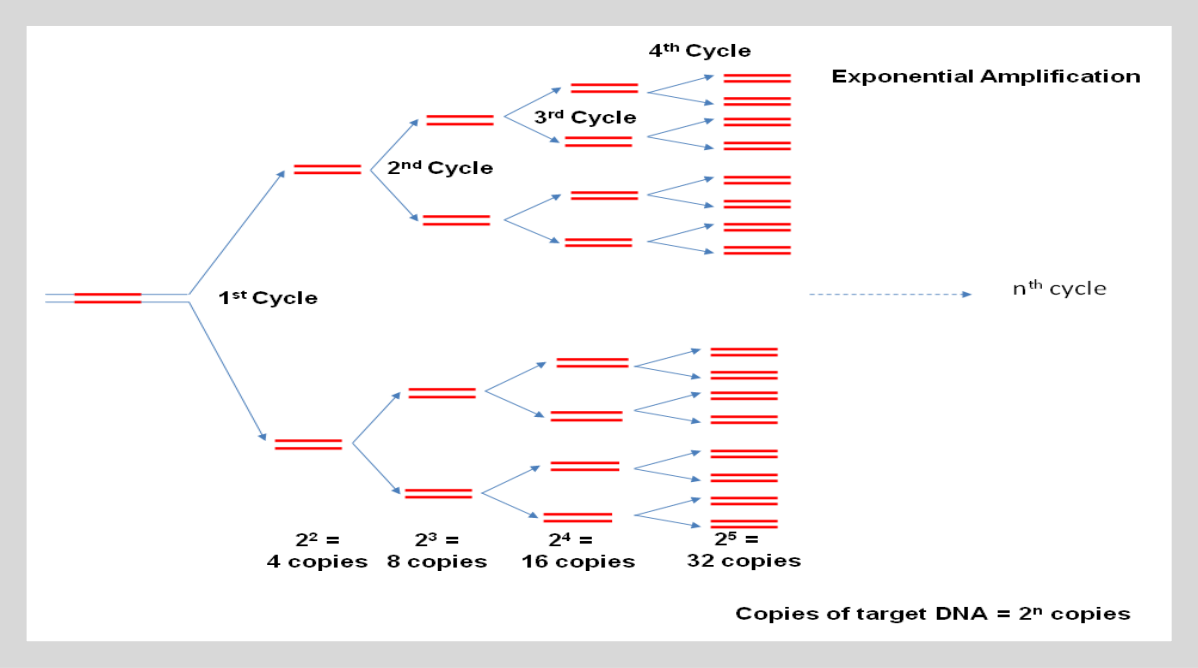

Figure 2.3.1. Target DNA (represented in red) amplification by PCR. The amount of target DNA produced by PCR doubles every cycle.

The newly synthesized strand will overlap the binding site of the opposite oligonucleotide primer. Large quantities of target DNA get produced as repeated cycles of PCR continue (fig. 2.3.1). The end result is an exponential increase in the total number of DNA copies of the targeted sequence, which are finally represented as a theoretical abundance of 2^n where “n” is the number of cycles [244-246].

2.3.2 Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) is a technique used for detection of polymorphisms that result in the loss or creation of a recognition site for a restriction enzymes to cleave the DNA [247]. Restriction enzymes recognize the restriction sites on the DNA molecule and cleave the DNA molecule at these sites to generate DNA molecules of varying length, which are called RFLP. DNA molecules are negatively charged at neutral pH due to their phosphate backbone and therefore migrate towards the cathode when subjected to an electric potential (fig. 2.3.2 A) [248]. In most cases somatic mutations in disease-related genes do not give rise to a functional change of the mutated cell which would allow its isolation or expansion *in vitro*. Therefore, selection of mutated cells on the basis of an altered phenotype has to be replaced by biochemical separation and detection of the altered sequence of the gene of interest.

Evidently such 'genotypic' mutation analysis requires large numbers of cells at the outset since the expansion of mutated cells is avoided. To overcome this hurdle RFLP technique was combined with polymerase chain reaction (PCR) to develop a new technique called polymerase chain reaction-restriction fragment length polymorphisms (PCR-RFLP) [222]. PCR amplifies the target DNA sequence to large quantities and the PCR product is subsequently subjected to corresponding restriction enzyme to generate RFLP.

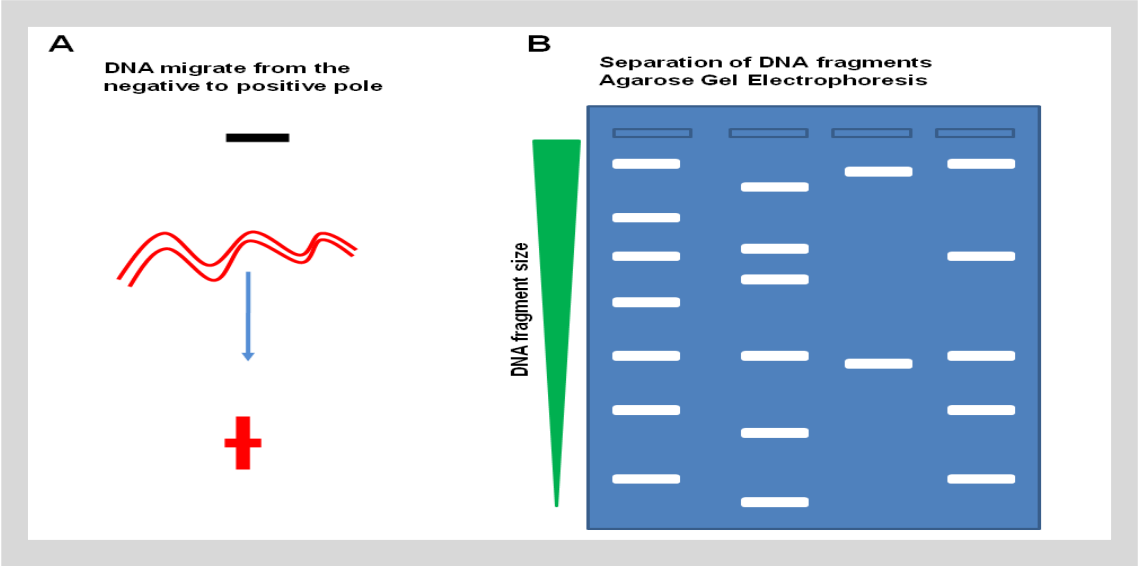

Figure 2.3.2. A, schematic diagram showing DNA migration when subjected to an electric field; **B,** separation of DNA fragments by agarose gel electrophoresis following restriction digestion.

The generated fragments are separated to their respective sizes accordingly by gel electrophoresis. The restriction enzymes recognize specific short palindromic sequence of four to eight nucleotides and hydrolyze the phosphodiester backbone of both DNA strands at these specific recognition sites [247].

Although a SNP in the restriction site or genetic code does not necessarily affect the function of the gene, these SNPs may alter the recognition site of enzymes used for PCR-RFLP analysis. Polymorphism occurring at the restriction site gives rise to banding patterns of DNA restriction fragments. Fragments of the digested DNA are separated by agarose gel

electrophoresis to resolve the banding patterns and measure the size of restriction fragments (fig. 2.3.2B). DNA molecules (or restriction fragments) sieve through the pores of agarose gel matrix at different rates, determined largely by their mass because the charge to mass ratio of all molecules is uniform [249]. The smaller DNA molecules migrate the furthest distance down the gel with the larger molecules remaining closest to the well after electrophoresis.

If several samples have been loaded into adjacent wells of the gel, they run parallel in individual lanes. Depending on the number of different fragment sizes, each lane shows different bands on the gel which correspond to each fragment size. Incomplete separation of the fragments leads to overlapping bands, or to indistinguishable smears representing multiple unresolved DNA fragments [250].

2.4. TaqMan® SNP Genotyping Assays

TaqMan® SNP Genotyping Assay is a single-tube PCR assay that exploits the 5' exonuclease activity of AmpliTaq Gold® DNA Polymerase [251, 252]. The assay includes two locus-specific PCR primers that flank the SNP of interest, and two allele-specific oligonucleotide TaqMan® probes. These probes have a fluorescent reporter dye at the 5' end, and a non-fluorescent quencher (NFQ) with a minor groove binder (MGB) at the 3' end [253].

An intact probe emits minimal fluorescent signal when excited because 5’ fluorophore is in close physical proximity with the 3’ quencher and this causes the fluorescent resonance energy transfer (FRET) effect to quench the fluorescence emitted by the fluorophore [209]. A strong fluorescent signal is generated when the intact probe hybridized to the target allele, is cleaved by the 5’ exonuclease activity of AmpliTaq Gold® DNA Polymerase. The PCR primers amplify a specific locus containing the SNP of interest and each fluorescent dye-labeled hybridization probe reports the presence of its associated allele in the DNA sample (fig. 2.4) [254, 255].

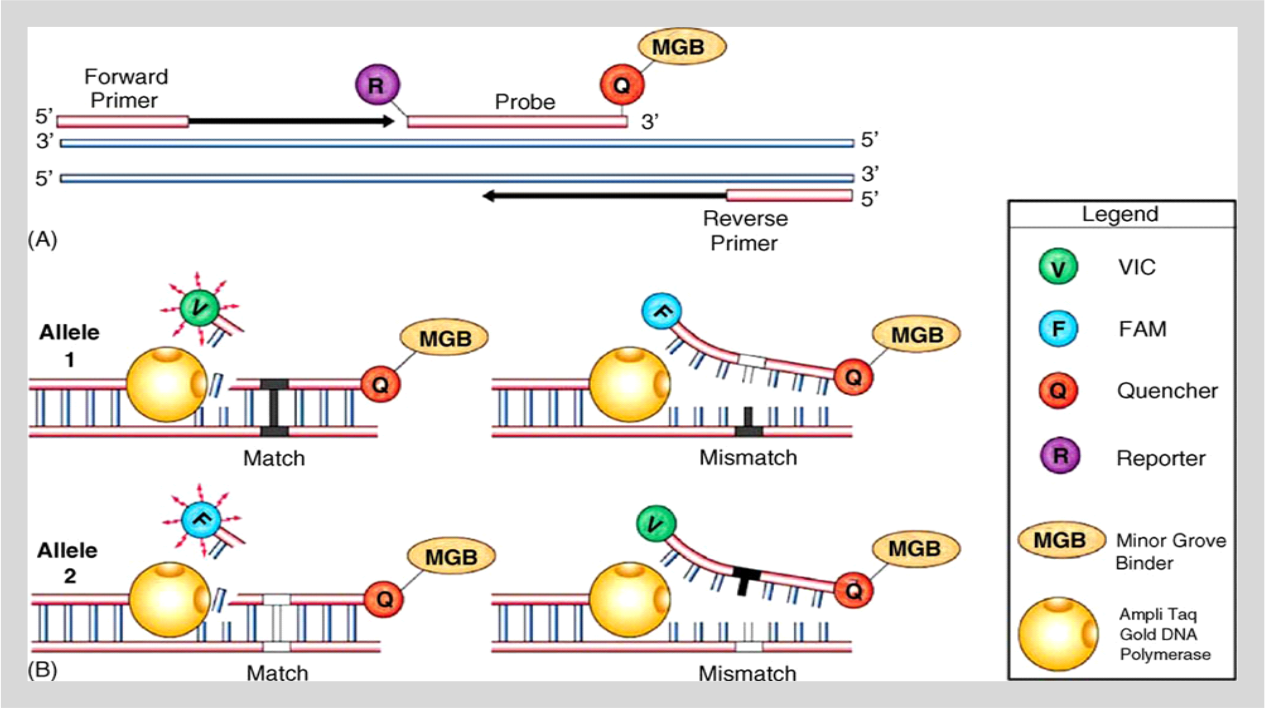

Figure 2.4. TaqMan reaction (www.appliedbiosystems.com, May 2011).

In each PCR cycle, cleavage of one or both allele-specific probes produces an exponentially increasing fluorescent signal by freeing the 5’ fluorophore from the 3’ quencher. The use of

two probes, one specific to each allele of the SNP and labeled with two fluorophores, allows detection of both alleles in a single tube.

Fluorogenic probes with an MGB produce enhanced allelic discrimination, because the MGB stabilizes the double stranded probe-template complex, thereby increasing the probe melting temperature without increasing probe length [256]. This provides enhanced mismatch discrimination between these shorter probes, resulting in improved allele specificity. These probes also increase the signal-to-noise ratio of an assay, because the reduced distance between the 5' fluorophore and the 3' quencher provides more efficient quenching of an intact probe [255].

2.4.1 TaqMan Assay Design

ABI has predesigned TaqMan SNP assays for almost all the SNPs that were discovered from the human genome project; these assays are commercially available and are ready to use (www.appliedbiosystems.com). However, newly described SNPs and few other SNPs existing in the SNP database (www.ncbi.nlm.nih.gov/SNP) would not have predesigned TaqMan SNP assays. For these SNPs, ABI developed a proprietary algorithm that allows researchers to design TaqMan primers and probes for TaqMan SNP Genotyping Assays. This primer and probe design algorithm implements the thermodynamic and heuristic rules described in their TaqMan assay design guidelines [253], as well as additional empirically derived factors that facilitate manufacturing and improve assay performance [251].

To increase the probability of generating a successful assay design, it is important first to mask all known SNPs and other polymorphisms within the region of the target polymorphism (reviewed in [255]). The masking of known SNPs ensures that: (1) the design pipeline does not place an oligonucleotide primer or probe over a base that is masked; (2) primers and probes are designed to the most conserved regions between individuals; and (3) oligonucleotide primers hybridize efficiently to the target regions that flank the SNP of interest.

2.4.2 Allele calling and visualization of results

The TaqMan SNP Genotyping Assay is read at the endpoint of PCR rather than in real time. Genotyping is performed in 96- or 384-well plates that allow many DNA samples to be genotyped simultaneously.

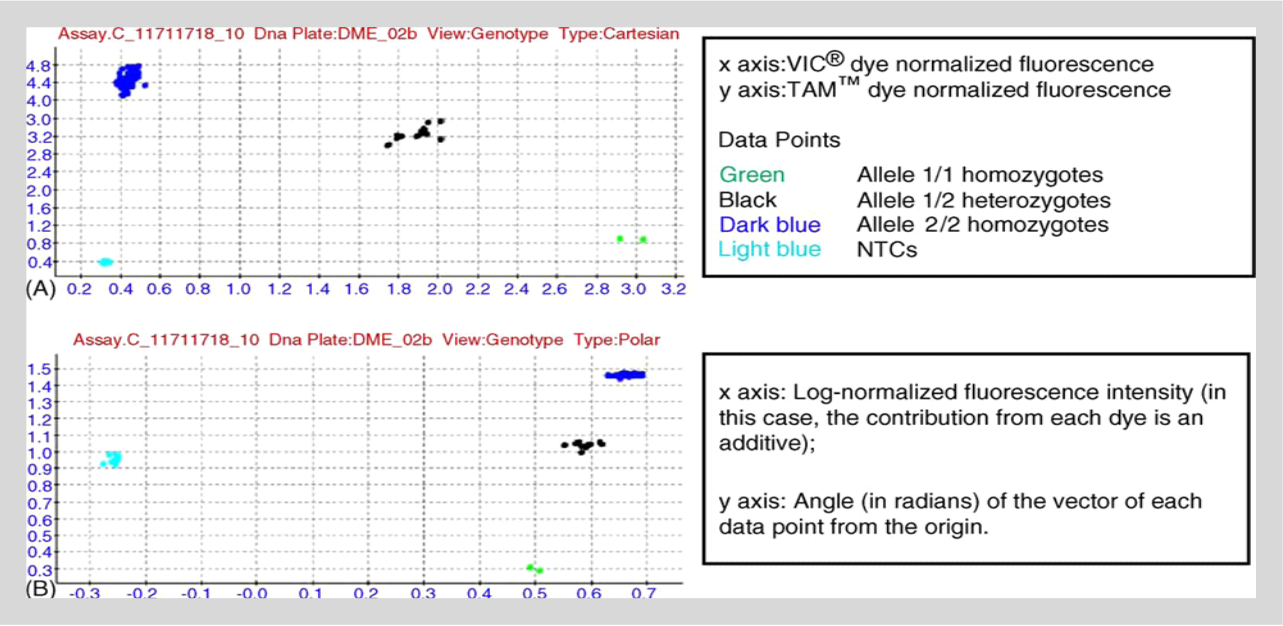

Figure 2.4.2.1. A, Cluster plot of 88 Coriell DNA samples and eight no-template controls (NTCs) from a typical TaqMan assay. **B,** Polar plot of same data. TaqMan assay data are not usually displayed in a polar plot (not available in the current commercially available software), but they are shown in this view to facilitate interpreting and comparing the SNPlex Genotyping System data [255] shown in the next figure.

Individual alleles of a SNP are presented as a dot of the normalized intensity of the reporter dyes in each sample well on a cartesian plot, also known as a scatter or cluster plot. A clustering algorithm in the data analysis software assigns individual sample data to a particular genotype cluster, represented as colored dots in the cluster plot (fig. 2.4.2.1).

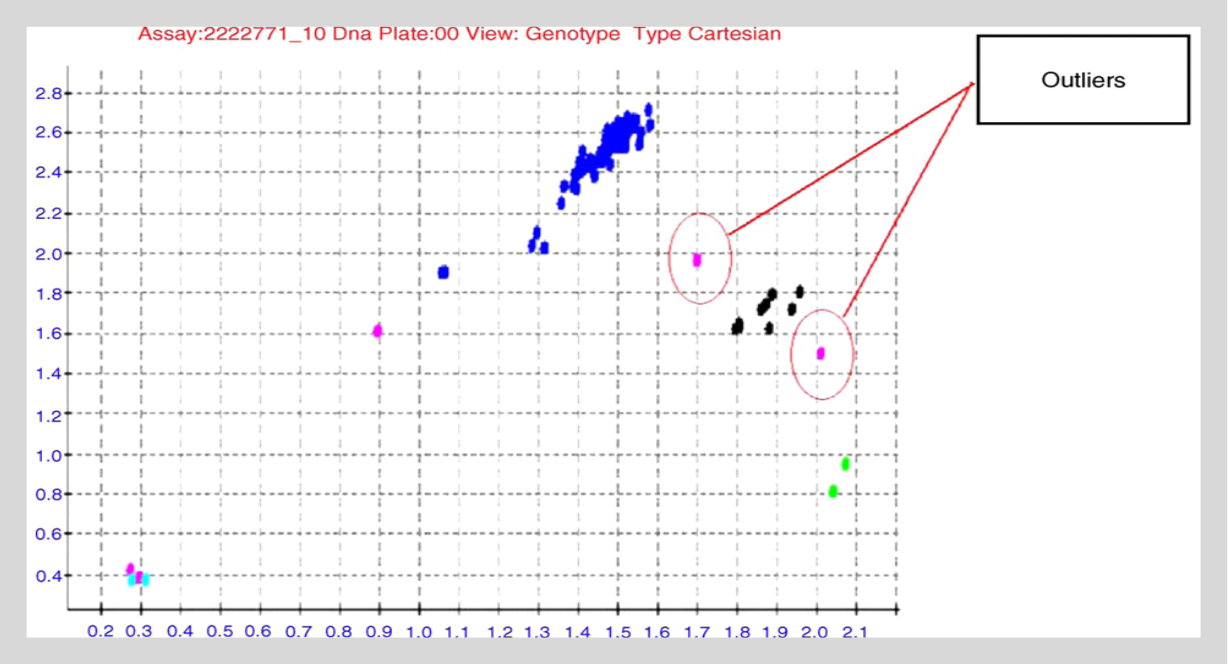

Figure 2.4.2.2. An example of multi-cluster data produced by a target SNP, CYP2D6, a gene that has additional copies in some individuals [255].

SNP Genotyping Assays that produce more than the three clusters expected for bi allelic SNPs provide clues about biological differences in samples that behave anomalously. Extra clusters could be caused by the differences in DNA samples which include (1) unidentified or rare SNP that underlies a primer or probe, (2) the targeted SNP region having an additional copy (exon, gene, or genomic region) in the genome, and (3) the fact that a SNP of interest may contain more than two alleles [255]. These clues, exhibited as anomalous results in genotype cluster plots, are often masked in other technologies such as PCR-RFLP (fig 2.4.2.2).

TaqMan SNP Genotyping Assays provide six significant technological advantages, which include the fact that these assays require only a single enzymatic step, use universal reactions and thermal cycling conditions. TaqMan SNP Genotyping Assays contain two specific primers targeting the region flanking the SNP site and two TaqMan fluorescent probes and can also genotype insertions/deletion polymorphisms in addition to SNPs [257]. The workflow is simple involving the adding of Universal Master Mix and the assay to the sample, followed by PCR and endpoint read [257]. The availability of a large number of validated, off the-shelf TaqMan SNP Genotyping Assays [257] makes it possible to implement a streamlined laboratory workflow. In addition, these assays require only simple, pre-PCR liquid-handling steps, which can be easily automated with a robotic liquid-handling process.

2.5 Amplified Luminescent Proximity Homogeneous Assay Screen (AlphaScreen)

The observation that, besides intrinsic enzymatic activities, proteins exert virtually all of their functions through interacting with other molecules such as nucleic acids, lipids, carbohydrates or small molecules and other proteins, has driven the development of technologies to examine these interactions [258]. The amplified luminescent proximity homogeneous assay screen (AlphaScreen) is an example of such technology developed to measure these interactions [259].

This technology employs oxygen channeling chemistry developed from a diagnostic assay technology known as luminescent oxygen channeling immunoassay (LOCI) [260, 261]. LOCI cause the excitation of a cascading series of chemical reactions generating

chemiluminescent signal, a process referred to as chemical generation. This approach takes the advantage of the short diffusional distance of singlet oxygen to initiate a chemiluminescent reaction in the proximity of chemical generation. AlphaScreen is insensitive to interference by particles or other substances present in biological sample matrices [259].

The assay format comprises two discrete ligand coated polystyrene beads, designated as “Donor” and “Acceptor” beads, which form pairs in the presence of analyte. The pairs must be within about 200 nm (approximately the diameter of a bead particle) in order for a chemiluminescent signal to be generated [259]. Donor beads contain a photosensitizing agent (phthalocyanine) that, when irradiated at 680 nm, excites ambient oxygen to a singlet state [260, 261]. Excitation of each Donor bead generates approximately 60,000 oxygen singlets per second [262], resulting in an amplified response when they come into contact with Acceptor beads.

In an AlphaScreen assay, the Acceptor beads contain three chemical dyes namely, thioxene, anthracene and rubrene (hence the designation, ‘TAR’ beads) [259]. Thioxene reacts initially with singlet oxygen to produce light energy, which is subsequently transferred to anthracene and thence to rubrene. The final compound in the cascade, rubrene, emits light at wavelengths of 520-620 nm [260]. This process occurs against a very dark background and the bead particles are often present in low concentrations in the assay, making the assay highly sensitive.

Consequently, non-specific close proximity interactions of Donor and Acceptor particles is rare and again the background is low (fig. 2.5), which increase the sensitivity and specificity of this assay [259].

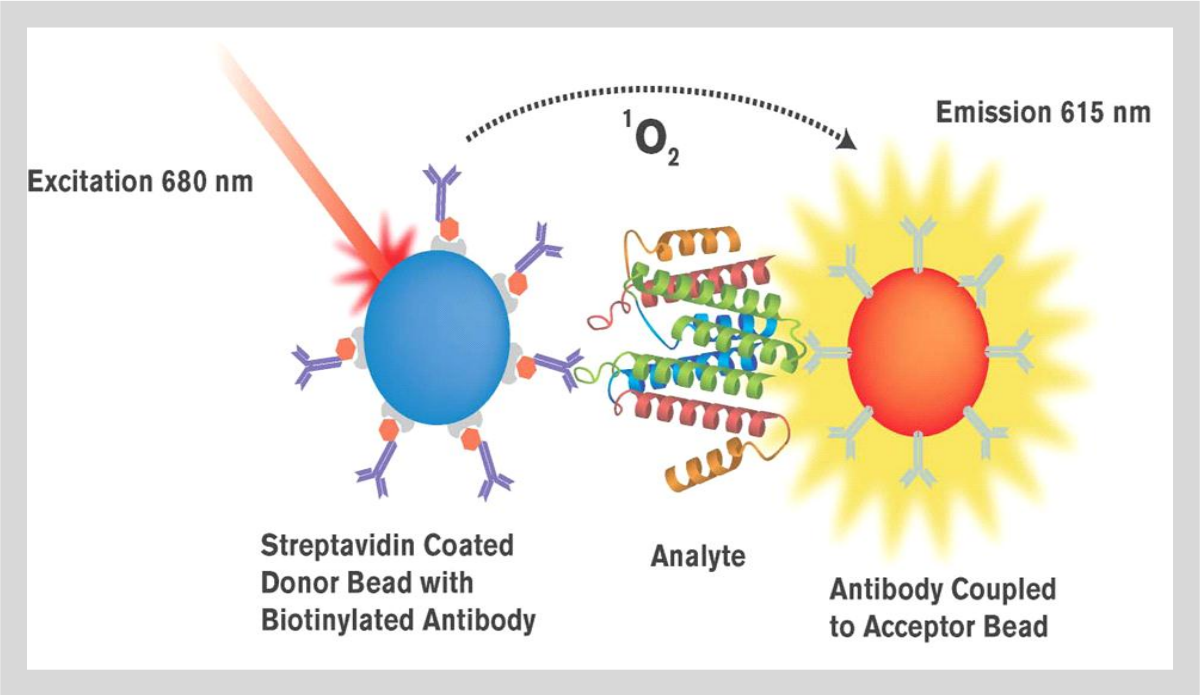

Figure 2.5. The AlphaScreen assay is based on an oxygen channeling technology. When the Donor (blue bead) containing phthalocyanine is laser excited (at 680 nm) ambient oxygen is converted to singlet oxygen. Singlet oxygen molecules travel at least 200 nm in aqueous solution before decay. If the Donor and Acceptor (gold beads) beads are within that proximity, energy transfer occurs. Singlet oxygen molecules react with chemicals in the Acceptor beads to produce a luminescent response. If the Acceptor bead contains Europium an intense luminescence is emitted at a wavelength of 615 nm [259].

The lifetime of the singlet oxygen reactive species in aqueous solutions is very short (approx. 4 ms) therefore donor and acceptor beads rely on an immunological complex to be bound to one another in order to generate a signal. Unbound beads experience insignificantly low concentrations of singlet oxygen, contributing minimally to the background signal [259]. The AlphaScreen technique has been shown to be remarkably acquiescent for the detection of protein–DNA, protein–RNA, or protein–small molecule interactions, as well as protein–protein interactions [258]. In the current study, AlphaScreen was used to measure protein-protein interaction.

2.5.1 Protein-Protein Interaction Assay Using AlphaScreen

AlphaScreen was one of the technologies developed to measure protein-protein interactions that can be used to identify either small molecule inhibitors or novel interacting partners. The interaction of cells occurs through complex protein-protein interactions, ranging from ligand binding, G protein coupling reaction, interaction of kinases with cognate substrates as well as the interaction of transcription factors with nuclear co-activators and co-repressors [258].

Growth factor proteins such as tumor necrosis factor alpha (TNFα) receptors are implicated in the etiology of many inflammatory and immunological disorders. Therefore an AlphaScreen assay was developed to measure ligand binding to a member of the TNF receptor super-family, OX40 (CD 134) [263]. This group developed a fusion protein where OX40 was coupled to a domain of human IgG, allowing OX40 to bind to Acceptor beads coated with protein A and OX40 ligand (OX40L-CD8) tagged with biotin which was coordinated by streptavidin coated Donor beads [263]. This led to the identification of several

peptides and small molecules that inhibited OX40L binding. Human papilloma virus (HPV) protein E6 interacts with ubiquitin ligase, E6AP to inhibit the activity of a tumor suppressor protein (p53) and block transformed cells from undergoing apoptosis [264]. Since HPV causes cervical cancer by inhibiting the interaction between E6 and E6AP [265] then AlphaScreen assay was developed to measure E6 binding to E6AP. This assay was subsequently used to detect inhibitors of this interaction [266].

The interaction between HIV-1 IN–LEDGF/p75 represents an attractive target for antiviral therapy. AlphaScreen assay has also been used to identify small-molecule inhibitors that target HIV-1 IN–LEDGF/p75 interaction [267]. Subsequently, Christ *et al*. rationally designed a series of 2-(quinolin-3-yl) acetic acid derivatives–LEDGINs–that act as potent inhibitors of the HIV-1 IN–LEDGF/p75 interaction [208]. This study demonstrated that LEDGINs inhibit HIV-1 replication by blocking the integration. The 2-(quinolin-3-yl) acetic acid derivatives could be defined as the first genuine allosteric HIV-1 integrase inhibitors because of their potent inhibition of HIV-1 replication and lack of cross resistance [208]. The data presented above indicate that the AlphaScreen technology provides a simplistic assay platform to measure protein-protein interaction.

CHAPTER THREE

3. Materials and Methods

3.1 Study Participants

The Center for the AIDS Programme of Research in South Africa Acute Infection 002 (CAPRISA AI 002) [268, 269] and the Sinikithemba [84, 270, 271] cohorts were used for this study. The CAPRISA AI 002 cohort is an ongoing observational natural history study of HIV-1 subtype C infection established in Durban, KwaZulu-Natal, South Africa in 2004. HIV negative females (n=245) at high risk for HIV infection were enrolled into Phase I of the study. Participants in this cohort were screened monthly for recent HIV-1 infection by two rapid HIV-1 antibody tests (Abbott Laboratories, Tokyo, Japan) and Capillus (Trinity Biotech, Jamestown, NY, USA). HIV-1 antibody negative samples were tested for HIV-1 RNA in batches of 10 plasma samples per pool using the Ampliscreen v1.5 assay (Roche Diagnostics, Rotkreuz, Switzerland), which has a detection limit of 10 copies/ml. Samples that tested positive in pooled plasma were individually tested by quantitative RNA (Amplicor v2.0, Roche Diagnostics) and HIV enzyme immunoassay (BEP 2000; Dade Behring, Marburg, Germany) to identify HIV-1 infection. $CD4^+$ T cell counts were determined by a 4-parameter FACSCalibur flow cytometer (Becton Dickinson).

Participants with acute HIV-1 infection were enrolled into Phase II of the study on the basis of a reactive HIV antibody test within 3 months of previously negative results or positive HIV RNA PCR in the absence of antibodies. Date of infection was estimated by taking the midpoint between the last HIV antibody-negative result and the first HIV antibody-positive

result or 14 days before the first positive HIV RNA PCR assay result for those identified as antibody negative but HIV RNA positive. An additional of 34 acutely infected participants (who met the criteria for acute infection, as aforementioned) were recruited from other ongoing CAPRISA cohorts. Participants in Phase II were monitored weekly for 3 weeks, fortnightly for 2 months then monthly for 9 months and quarterly thereafter. 195 out of 217 DNA samples obtained from participants who remained HIV-1 negative and 52 of the 62 DNA samples obtained from participants with acute HIV-1 infection were available for genotyping by TaqMan (fig. 3.1.1).

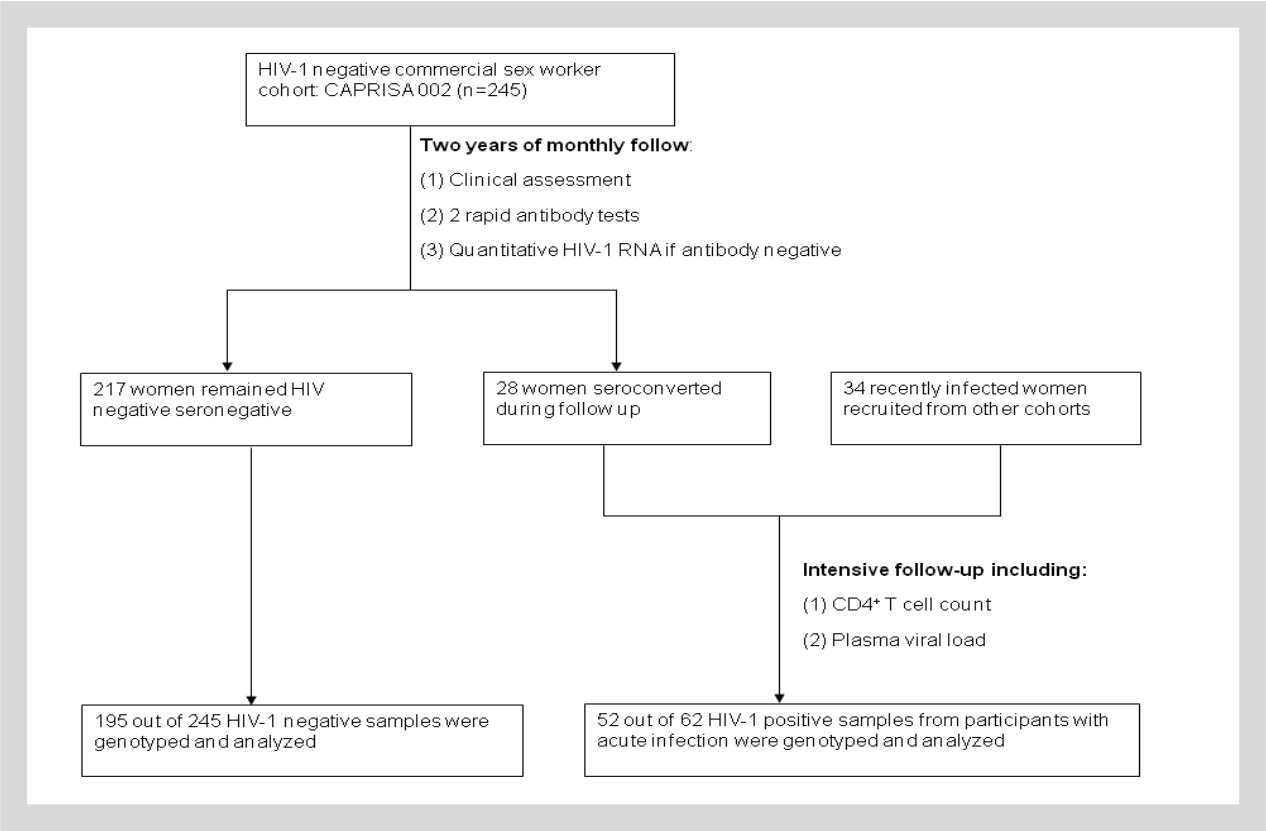

Figure 3.1.1. Description of CAPRISA AI 002 study cohort indicating the number of participants genotyped in the HIV-1 positive and negative groups.

The Sinikithemba cohort comprises 450 antiretroviral naïve, HIV-1 subtype C chronically infected adults enrolled from McCord Hospital (Durban, South Africa) from August 2003 to 2008 and followed longitudinally [84, 270, 271]. Sociodemographic characteristics, plasma viral load and CD4 cell count measurements were obtained at baseline. CD4 cell counts and viral loads were measured every 3 and 6 months respectively from enrollment. Viral loads were determined using the automated CobasAmplicor HIV-1 Monitor test (version 1.5; Roche Diagnostics). $CD4^+$ T cells were enumerated using the Multitest kit (CD4/CD3/CD8/CD45) on a FACSCalibur flow cytometer (Becton Dickinson). 403 out of 450 DNA samples obtained from participants with chronic HIV-1 infection (Sinikithemba cohort) were available for genotyping by TaqMan assays (fig.3.1.2).

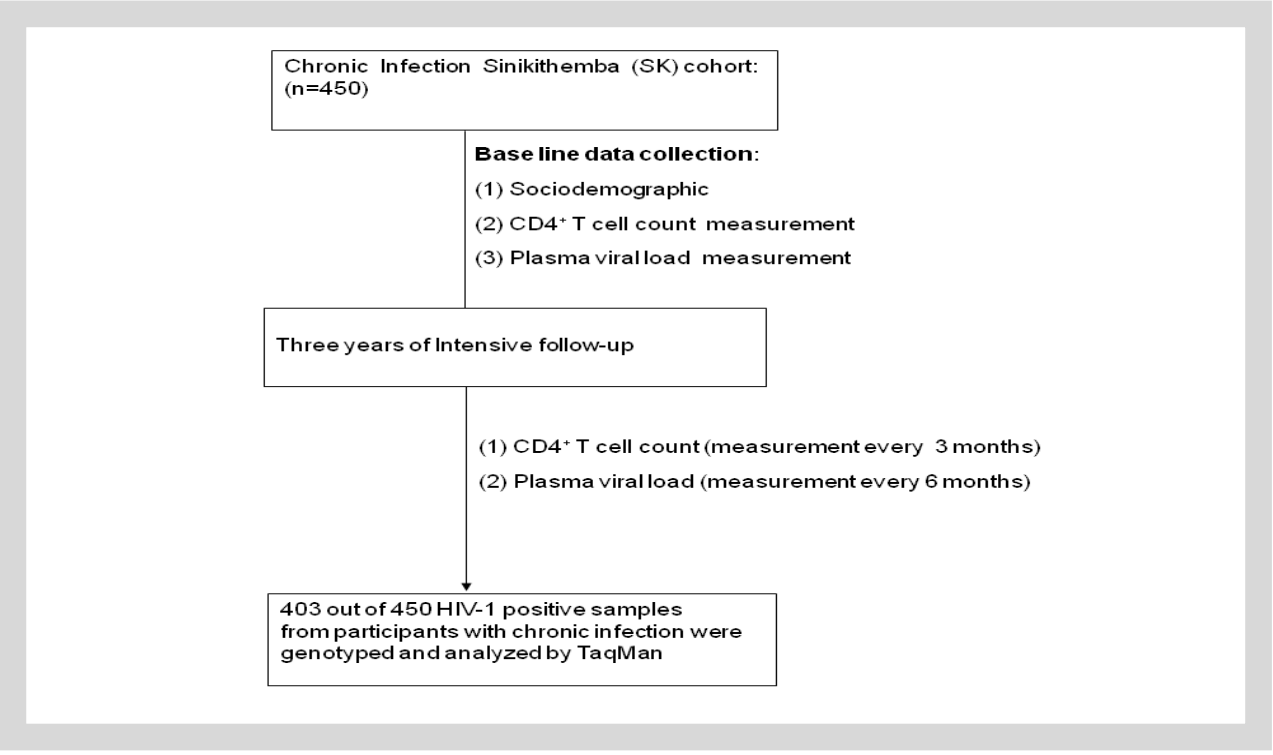

Figure 3.1.2. Description of the Sinikithemba study cohort of chronically infected participants.

Ethical approval for this study was obtained from Biomedical Research Ethics Committee of the University of KwaZulu-Natal. All participants provided written informed consent.

3.2 Molecular Analysis of DNA variation

The DNA was extracted from the buffy coats obtained from the participants of this study using Qiagen DNA extraction kit according to the manufacturer's instruction (www.qiagen.com). The purity and quantity of DNA was measured by Nano Drop 2000/2000c spectrophotometer (Thermo Scientific). The DNA yield obtained using this kit was in the range of 100 ng/µl and the sample was considered pure if the A_{260}/A_{280} reading was 1.8.

3.2.1 Sequencing of the Integrase Binding Domain (IBD) of LEDGF/p75

The IBD that interacts with HIV-1 IN is in the C-terminus region of LEDGF/p75. This study first screened part of the C-terminus region of *PSIP1*, which includes the IBD, for new polymorphisms by re-sequencing a DNA panel obtained from 83 HIV-1 seronegative participants (SN) and 43 HIV-1 seropositive participants (SP) from the CAPRISA AI 002 cohort. A set of three sequencing primers (fig. 3.2.1) and ABI PRISM^{TM} Dye terminator cycle sequencing core kit (Perkin Elmer, Brussels Belgium) was used in re-sequencing this panel of 126 DNA samples in a 3130 XL Genetic Analyzer (Applied Biosystems).

Sequencing primers covered 1679bp long DNA fragment starting from 1190 bp up stream of exon 10 through to the end of exon 13 of the *PSIP1* gene. Firstly, a 2664bp long fragment was amplified using two specific PCR primers: forward primer LEDGFDNA1 5'-TGGGCTCAA

AGC ATTAATCC -3' and a reverse primer Seq5 5'- CTCTGAAGGATTCTACACTAGATAAC -3'.

A 50µl PCR reaction containing 1X PCR buffer, 3.5 mmol/µl $MgCl_2$, 0.25 mmol/µl dNTPs, 0.25 U/µl Expand Taq (Roche), 0.4ρmol/µl of each primer (forward and reverse primer) and 100ng DNA sample was prepared. This was amplified at 95 ºC for 10 minutes followed by 35 cycles of 94 ºC for 30 seconds, 60 ºC for 30 seconds, 72 ºC for 45 seconds, and a final 10 minute extension step at 72 ºC. The PCR products were purified using Qiagen PCR Purification Kit (Qiagen) according to manufacturer's instruction (www.qiagen.com).

3.2.2 Genotyping for regulatory SNPs in the *PPIA* gene using PCR-RFLP

DNA samples from 47 SPs and 168 SNs from the CAPRISA AI 002 cohort were available for genotyping by PCR-RFLP. For analysis of these polymorphisms in the promoter region of the *PPIA*, DNA samples were amplified by PCR using Taq DNA polymerase (Invitrogen) in the presence of 1.25 mmol/µl $MgCl_2$, 0.4 mmol/µl dNTPs, and a primer pair CYPAex-1s (5'-AAGTCGCAGACCCGATTG-3') and CYPAex-1a (5'-ACTTTCTGGGCCCCATTC-3') at 10 ρmol/µl each. The following amplification cycles were used: 10 min 95 ºC; 35 cycles of 30 s 94 ºC, 30 s 60 ºC, 45 s 72 ºC; 5 min 72 ºC.

Subsequently, PCR products (250 bp in length) were subjected to a restriction digest with either *Rsa*I or *HAE*III (4 hours to overnight at 37^{o}C; New England Biolabs) to detect

polymorphisms and analyzed by 4% agarose gel electrophoresis. *Rsa*I restriction digestion was used to detect the SNP A1650G: a PCR product containing the wild type genotype (AA, referred to as 1650AA in this report) of the SNP A1650G resulted in 2 restriction fragments of 130 bp and 120 bp. A PCR product containing the homozygous mutant genotype (GG, referred to as 1650GG in this report) of the SNP A1650G was undigested following restriction digest reaction resulting in a 250 bp undigested fragment. A PCR product containing the heterozygous mutant genotype (AG, referred to as 1650AG in this report) of the SNP A1650G resulted in 3 restriction fragments of 250 bp, 130 bp and 120 bp. *HAE*III restriction digestion was used to detect the SNP C1604G: a PCR product containing the wild type genotype (CC) of the SNP C1604G would result in 5 restriction fragments of 110 bp, 40 bp, 36 bp, 28 bp and 19 bp. A PCR product containing the mutant homozygous genotype (GG) of the SNP C1604G would result resulting in 4 restriction fragments of 138 bp; 40 bp, 36 bp, and 19 bp. A PCR product containing the heterozygous genotype (CG) of the SNP C1604G would result in 3 restriction fragments of 138 bp, 110 bp and 40 bp.

3.2.3 Genotyping of SNPs in the *TNPO3* and *PSIP1* genes using TaqMan Assays

A final volume of 15 µl of 20X homemade master mix which was 5 µl of 1X PCR Buffer containing 1.5 mmol/µl $MgCl_2$ (Roche), 4 µl of 2.5 mmol/µl $MgCl_2$ (ABI's), 0.31 µl of 80% Glycerol, 0.05 µl of 50% Tween 20, 3.4 µl of distilled sterile water, 0.27 µl 10 mmol/µl dNTPs, 0.11 µl of 1µM SFHD, 0.05 U/µl DNA polymerase, and 0.068 U/µl of Taq Gold (Roche) and AOD/ABD/TaqMan probes was prepared. TaqMan Assays were performed in 384 well plates and each well contained 13 µl of the homemade master mix and 2 µl of 10 ng/µl DNA

sample. This was amplified at 95 ºC for 3 minutes, 40 cycles of 92 ºC for 15 seconds and 60 ºC for 60 seconds.

TaqMan assays were used to genotype intronic and exonic haplotype tagging (ht) SNPs in *TNPO3* gene and only intronic ht SNPs in *PSIP1* gene. SNP rs61744944 which is in the exon region of *PSIP1* was also included in this analysis. The 6 intronic and 2 exonic ht SNPs in *TNPO3* were rs13242262, rs2305325, rs11768572, rs1154330, rs35060568, rs8043, rs6957529 and rs10229001 all available from NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP) and HapMap databases (http://www.hapmap.org) selected by considering location, spacing, and allele frequency of at least 10% (Table 3.2.3).

Table 3.2.3. Showing the rs numbers of *TNPO3* and *PSIP1* SNPs and TaqMan Assays

Gene	SNP number	exon or intron region	rs number (amino acid change)	TaqMan Assay ID Number or Assay sequences
TNPO3	SNP1	regulatory	rs13242262	C_2691237_10
	SNP2	intron	rs2305325	C_7877026_1
	SNP3	intron	rs11768572	C_31283360_10
	SNP4	intron	rs1154330	C_8338470_10
	SNP5	exon	rs35060568	C_22273560_10
	SNP6	exon	rs8043	C_8336193_1
	SNP7	intron	rs6975529	C_31283343_10
	SNP8	intron	rs10229001	C_2691246_10
PSIP1	SNP1	intron	rs2277191	C_15883595_10
	SNP2	intron	rs10283923	C_29529242_10
	SNP3	intron	rs12339417	C_31936110_10
	SNP4	intron	rs1033056	C_2757693_20
	SNP5	exon	rs61744944 (Q472L)	PSIP1_Q472L_s AAAACCAAA GATCAAGGGAAGAAA Psip1_q472l_a TGTGAAATTGT TGGCTTTTTACCA

The 4 intronic ht SNPs in *PSIP1* were rs2277191, rs1033056, rs12339417 and rs10283923, all available in the databases mentioned above and were selected as per aforementioned criteria (Table 3.2.3). The ht SNPs were chosen because they are all ht SNPs for *TNPO3* and *PSIP1* genes in the Yoruba population from Nigeria (http://www.snp.cshl.org). The exonic SNP rs61744944 was selected based on its association with HIV-1 clinical outcomes in the preliminary analysis of the sequencing data (section 4.3.2.2). The SNPs in both *TNPO3* and *PSIP1* genes were genotyped in 247 patient samples (195 SNs and 52 SPs) from the CAPRISA AI 002 cohort and 403 patient samples from the Sinikithemba cohort.

Genotyping was performed by TaqMan SNP assay as per manufacturer's protocol (Applied Biosystems). TaqMan assays were obtained from the Assay-by-Demand service of Applied Biosystems (http://www.appliedbiosystems.com). Eight negative controls which contained water instead of DNA were included in each plate to monitor the potential of PCR contamination. Samples were genotyped in duplicate and genotypes obtained were free of water contamination or of inconsistencies between duplicates.

3.2.4 Peripheral blood mononuclear cell (PBMC) sample processing and RNA isolation

A modified Trizol method [272] was used to extract RNA from peripheral blood mononuclear cells (PBMCs). In this modified method, PBMCs were homogenized in 1ml Trizol reagent, followed by incubation for 10 minutes at ambient temperature. The homogenate was centrifuged @ 12,000 revolutions per minute (rpm) for 10 minutes at 4 ºC and a 200µl volume of chloroform was added to the same tube. The combination of chloroform and

centrifuged homogenate was thoroughly mixed by shaking the tube vigorously for 30 seconds followed by incubation at room temperature for 5 minutes and centrifugation @ 12,000 rpm for 15 minutes at 4^{o}C.

The supernatant which contained the total RNA was transferred into a new 1.5ml eppendorf tube. 500µl isopropanol and 1.5µl glycogen was added into this new tube to precipitate out the total RNA and the tube was incubated at room temperature for 10 minutes. This was followed by centrifugation for 10 minutes at 4 ºC @ 12,000 rpm. The supernatant was discarded leaving the total RNA pelleted at the bottom of the tube. The RNA pellet was washed by resuspending it in 1ml of 75% ethanol and this RNA suspension was centrifuged @ 12,000 rpm for 5 minutes at 4 ºC. The supernatant was discarded and the RNA pellet was air dried then dissolved in 30µl nuclear free (DEPC) water.

The RNA concentration was quantified using a NanoDrop spectrophotometer 2000 (Thermal Scientific) and the purity of RNA was determined by the ratio of absorbance at the wavelength of 260 (A_{260}) and 280 (A_{280}). The samples were used only if the A_{260}/A_{280} ratio was 1.90 or greater. All RNA samples were DNAse treated using the fermentas DNASE kit (Fermentas cat. no. EN0521).

3.2.4.1 Reverse Transcription

An amount of 1 µg of the total RNA from each sample was reverse transcribed using the iScript cDNA synthesis kit (Biorad) as per the manufacturer's instructions. Briefly, the

reaction mixture was constituted of 4µl of 5x iScript reaction mix, 1µl of iScript reverse transcriptase, 1µg RNA and DEPC water to a final volume of 15µl. The above reaction mixture was prepared on ice in a sterile PCR tube. The reaction mixture was incubated for 5 minutes at 25 ºC, 42 ºC for 30 minutes, and 85 ºC for 5 minutes and held at 4 ºC for 45 minutes in the Gene Amp 9700 PCR System (Applied Biosystems, California, USA).

3.2.4.2 Quantitation of RNA using real-time PCR.

The PCR primers and cycling conditions used for glyceraldehyde 3-diphosphate dehydrogenase (GAPDH) and CypA real-time quantitative PCR are provided in section 4.1.1.5 below. GAPDH is a house keeping gene and was determined to be the most suitable reference gene based on PCR efficiency. Each PCR reaction consisted of 3 mmol/µl $MgCl_2$, 0.5pmol/µl of the respective primers for GAPDH and CypA (Table 4.1.1.5), 1µl Fast Start SYBR Green I (Roche), 1 µg cDNA and water (10 µl total volume) was prepared. Reactions were run on a Roche LightCycler v1.5 (1 cycle at 95 ˚C for 10 min), then 45 cycles of denaturation, annealing and extension (see section 4.1.1.1 for details). Melting curve analysis was used to confirm the amplification specificity. Serial dilutions of cDNA from total RNA were performed for each target gene. These served as standard curves for quantitative analysis.

CHAPTER FOUR

4. RESULTS

4.1 Genetic Variant A1650G of Cyclophilin A Gene (*PPIA*) Accelerate Progression to AIDS in Black South Africans

HIV-1 is an obligate intracellular parasite and therefore it relies, on cellular cofactors for efficient and productive infection. CypA, also known as peptidyl prolyl isomerase A (PPIA), is a cytoplasmic factor that is essential for efficient infection of HIV-1 [134, 141, 146]. CypA is a member of the cyclophilin family, members of which all possess peptidyl-prolyl cis/trans isomerase activity. Peptidyl prolyl cis/trans isomerases catalyze the cis/trans isomerization of prolyl peptide bonds and are believed to be involved in protein folding [72]. This suggests that CypA has a possible role in uncoating of the viral core following entry into the cytoplasm. CypA is incorporated into the HIV-1 virion capsid through the direct binding between prolyl peptide bond located in a proline-rich loop of the fourth and fifth helices of the HIV-1 capsid and the active sites of CypA [129, 273].

CypA has long been known to enhance HIV-1 replication in host cells [134, 135] (by convention, we refer to the protein as CypA and the gene as *PPIA*). Disruption of CypA incorporation, either by HIV-1 Gag mutations or by cyclosporine A, an immunosuppressive drug that prevents HIV-1 Gag binding to CypA, leads to an attenuation of HIV-1 infectivity [134, 274]. HIV-1 replication was significantly inhibited in CypA-null human $CD4^+$ T cells, in which the *PPIA* was deleted through homologous recombination [145]. CypA is therefore an important host factor required for efficient and productive HIV-1 infection. It has also been

postulated that binding of CypA to capsid protects HIV-1 from an unknown restriction factor in humans [275].

Genetic polymorphisms in the regulatory region of *PPIA* gene (fig. 4.1) have been implicated in HIV-1 infection and pathogenesis [61, 73, 74]. Bleiber *et al*. reported an association between the minor allele (G) of SNP A1650G (1650G) and rapid disease progression in the Swiss Caucasian population [61]. Subsequently, An *et al*. reported that the same allele, 1650G was more frequent among SPs suggesting that the 1650G might be associated with increased susceptibility to HIV-1 infection in the American [73]. However, the 1650G was neither significantly associated with susceptibility to HIV-1 infection nor disease progression in this study [73]. Instead, this group found a significant association of the minor allele (C) of SNP C1604G (1604G) with accelerated progression to AIDS in the American population.

In another study, Rits *et al*. found the frequency of the 1650G to be significantly increased in high-risk seronegative (HRSN) group compared to SP group of the Amsterdam Cohort studies (ACS) and they also reported reduced levels of CypA mRNA expression in PBMCs from SN group [74]. According to this report, these findings suggested that the 1650G might be associated with reduced susceptibility to HIV-1 infection in participants of the ACS. It is important to note that the data reported by this this group [74] regarding the susceptibility to HIV-1 infection is in sharp contrast with the previous report [73]. Nevertheless, this group found an association between the 1604G and accelerated disease progression, measured by increasing viral load, confirming the effect of the 1604G on disease progression [73, 74].

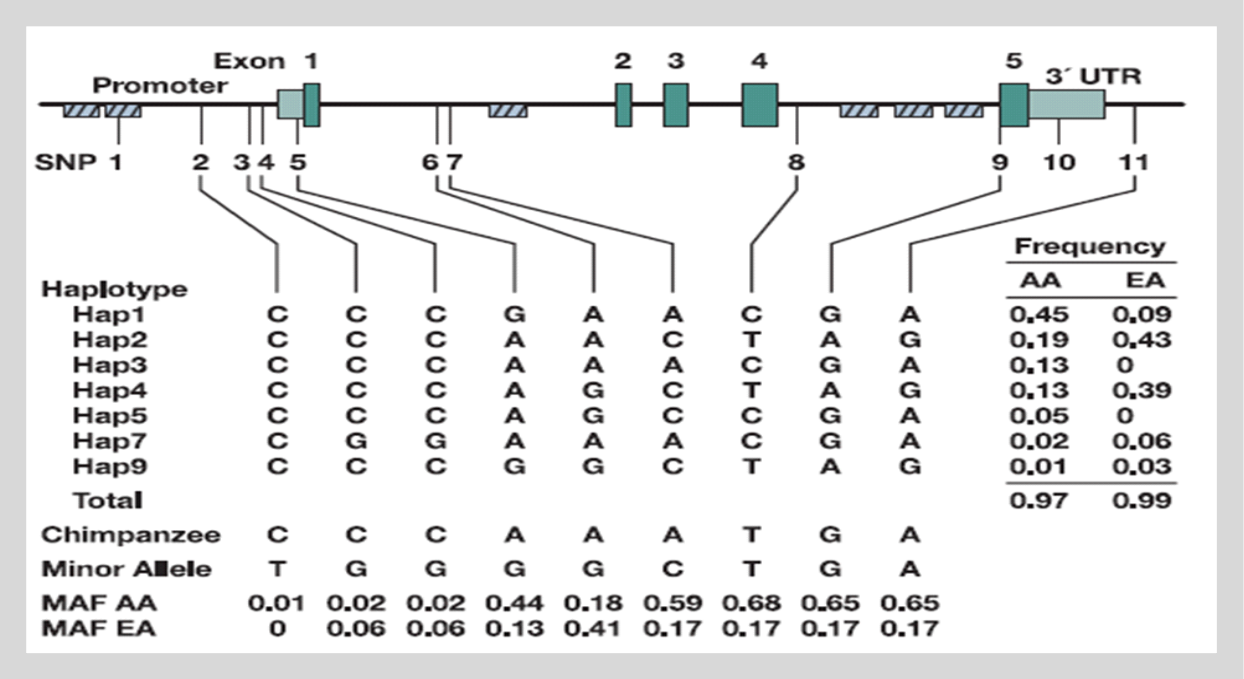

Figure 4.1. Gene Map, SNPs, and Haplotypes in the Human *PPIA* on Chromosome 7p13. Coding exons are marked by green blocks, and 59 and 39 UTRs by light green blocks. Hatched black boxes represent Alu repeats. Haplotype structure and frequencies are shown in the middle panel. The corresponding nucleotides in the chimpanzee and allele frequencies of the SNPs are presented in the bottom panel. MAF: minor allele frequency of SNPs. AA: African Americans. EA: Europeans Americans [73].

The data obtained for the role of *PPIA* genetic variation in susceptibility to HIV-1 infection and disease progression were inconsistent [61, 73, 74]. The plausible reason for these discrepancies could be attributed to different cohort designs. Bleiber *et al*. studied the effect of *PPIA* genetic variation in a cohort of Swiss origin [61], An *et al*. studied the effect of *PPIA*

genetic variation in cohorts of American origin [73] while *Rits et al*. investigated the effect of *PPIA* genetic variation in cohorts of Amsterdam origin [74]. In the current study, we hypothesized that the differences observed in the effect of *PPIA* genetic variation on HIV-1 disease outcome could be due to different population groups (or study cohorts). Therefore, in the present study we wanted to test the effect of *PPIA* genetic variation, SNPs A1650G and C1604G in particular, on susceptibility to HIV-1 infection and disease progression in Black South African women from the CAPRISA AI 002 study cohort. Using some patient samples from the same cohort, we also studied the impact of *PPIA* genetic variation on CypA mRNA expression levels.

4.1.1 Materials and Methods

Study participants. A cohort of black South African women (CAPRISA AI 002) [269] was used for this study (see section 3.1 for detailed description). Ethical approval for this study was obtained from Biomedical Research Ethics Committee of the University of KwaZulu-Natal. All participants provided written informed consent.

4.1.1.1 Genotyping for regulatory SNPs (A1650G and C1604G) in the *PPIA* gene in the CAPRISA AI 002 cohort

DNA samples from 47 seropositive (SP) and 168 seronegative (SN) participants from the CAPRISA AI 002 cohort were available for genotyping. For the analysis of the C1604G and A1650G polymorphisms in the promoter region of the *PPIA* gene, DNA samples were amplified by PCR using *Taq* DNA polymerase (Invitrogen) in the presence of 1X AmpliTaq

Gold PCR buffer, 1.25 mmol/µl $MgCl_2$, 0.4mmol/µl dNTPs, 10 ρmol/µl of each primer pair CYPAex-1s (5'-AAGTCGCAGACCCGATTG-3') and CYPAex-1a (5'-ACTTTCTGGGCCCCATTC-3'). The following amplification cycles were used: 10 min 95 ºC; 35 cycles of 30 s 94 ºC, 30 s 60 ºC, 45 s 72 ºC; 5 min 72 ºC. Subsequently, PCR products were subjected to a restriction digest with either *Rsa*I or *HAE*III (overnight at 37^{o}C; New England Biolabs) to detect polymorphisms and analyzed on a 4% agarose gel (see section 3.2.2).

4.1.1.2 CypA expression analysis and in vitro replication assay.

Peripheral blood mononuclear cells (PBMCs) were isolated from 30 SNs and 28 SPs from the CAPRISA AI 002 cohort as previously described [276] and these participants were genotyped for SNP A1650G. There were at least two study time points available for each of the participants in the primary HIV-1 infection phase and a total of 75 separate samples were analyzed for this group. Samples were available before and after HIV-1 infection for 13 of SP participants.

4.1.1.3 Sample processing, viral load quantification and CD4 cell enumeration.

PBMCs were isolated by Ficoll-Histopaque (Sigma) density gradient centrifugation from blood within 6 hours of phlebotomy and frozen in liquid nitrogen until use. Viral load was determined using the automated COBAS AMPLICOR HIV-1 Monitor Test v1.5 (Roche). $CD4^{+}$ T cells were enumerated by using the Multitest kit (CD4/CD3/CD8/CD45) on a four-parameter FACSCalibur flow cytometer (Becton Dickinson).

4.1.1.4 RNA isolation and analysis.

For all samples, RNA was extracted immediately after thawing and counting of PBMCs without *in vitro* stimulation. RNA was extracted from 2 x 10^6 PBMCs using the TRIzol LS Reagent (Invitrogen). The total RNA concentration was quantified and samples were used only if the A_{260}/A_{280} ratio was 1.90 or greater. All RNA samples were DNAse treated. 1 µg of total RNA from each sample was reverse transcribed using the iScript cDNA synthesis kit (Biorad).

4.1.1.5 Real-time PCR RNA quantitation.

Following RNA isolation and cDNA synthesis, real time quantitative (q) PCR was performed to measure the CypA mRNA expression levels. For this experiment GAPDH was used as the reference gene based on its PCR efficiency. PCR primers and cycling conditions used for GAPDH and CypA real-time qPCR are shown in Table 4.1.1.5. Each PCR reaction consisted of 3 mmol/µl $MgCl_2$, the respective primers 0.5 pmol/µl for GAPDH and CypA, 1µl Fast Start SYBR Green I (Roche), 1 µg cDNA and water (10 µl total volume). Reactions were run on a Roche LightCycler v1.5 (1 cycle at 95 ˚C (10 min), then 45 cycles of denaturation, annealing and extension.

Table 4.1.1.5. Primers and cycling conditions used in real time qPCR for CypA expression assay

Gene	Accession number	Primer Sequence 5'-3'	Cycling conditions: denaturation, annealing and extension
GAPDH	**NM_002046**	F: 5'-AAGGTCGGAGTCAACGGATT-3' R: 5'-CTCCTGGAAGATGGTGATGG-3'	(95 ºC, 6 s), (60 ºC, 6 s) and (72 ºC, 10s)
CypA	**NM_021130**	F: 5'-GTCAACCCCACCGTGTTCTTC-3' R: 5'-TTTCTGCTGTCTTTGGGACCTTG -3'	(95 ºC, 6 s), (60 ºC, 6 s) and (72 ºC, 10s)

GAPDH: Housekeeping gene, glyceraldehyde 3-diphosphate dehydrogenase

CypA: Cyclophilin A

F: forward primer

R: reverse primer

To confirm amplification specificity, the PCR products were subjected to a melting curve analysis. Serial dilutions of cDNA synthesized from total RNA were made for each target gene. These serial dilutions were then used to create standard curves for quantitative analysis.

4.1.1.6 Viral infection.

Replication capacity of HIV-1 in PBMCs ($2x10^6$) isolated from healthy 26 SNs that were genotyped for A1650G was investigated. 8 1650AA wild type genotypes and 18 AG/GG mutant genotypes were placed in a 12-well plate in R20 and incubated for 3 days at 37^oC, 5% CO_2. Following stimulation with IL2, cells were washed with R10 and then infected with HIV IIIB (NIH AIDS Reagent Repository) by spinoculation (2h, 2500 RPMs, 37^oC) at $2x10^5$ cells/well in a 96-well plate. Virus was subsequently removed; cells were washed once, and then allowed to incubate for an additional 2-7 days before analysis of CypA expression real-time PCR. Cell culture supernatants from day 2 and 7 samples were harvested and analyzed by p24 ELISA (Becton Dickinson).

Statistical analysis. The difference in allele frequency distribution between the SN and SP group was determined by Fisher's exact test (FET) for the SNP A1650G to test the null hypothesis that allele frequencies were the same in the two groups.

The effect of the SNP A1650G on HIV-1 viral load and $CD4^+$ T cell count was determined using a Generalized Estimating Equation (GEE) model [277] taking into account longitudinal measures for each participant. Viral loads were log-transformed and the square root of $CD4^+$ T cell count was used to normalize their measurements.

Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effect of each SNP on time to HIV-1 infection after enrollment and on the rate of progression to AIDS defined as $CD4^+$ T cell decline to less than 350 cells/μl (CD4<350). Decline in CD4 levels was determined and compared for the group with one or two copies of the minor allele to a reference group with two copies of the major allele (dominant genetic model). The significance of genotypic associations and relative hazard (RH) was determined by unadjusted Cox regression analysis for the dominant genetic model.

CypA mRNA expression levels were compared between SNs and SPs by performing dot plot graphical representation, nonparametric statistical analysis, and correlation (Pearson). Values were expressed as median values. Differences between the 2 groups were evaluated using Dunn's multiple comparison test, whereas the Mann-Whitney *U* test was used for any 2-group comparisons. The software used for the analysis was SAS version 9.1.3 (SAS Institute Inc., Cary, NC). A result was considered significant if the P-value was < 0.05.

4.1.2 Results

4.1.2.1 Genotyping for regulatory SNPs (A1650G and C1604G) in the *PPIA* gene.

Although there are many SNPs in the *PPIA* gene [41], in this study we focused only on two SNPs, A1650G and C1604G because they have been shown to have an association with disease outcome in the previous studies [61, 73, 74]. DNA was extracted from buffy coats obtained from 215 participants in the CAPRISA AI 002 cohort according to the manufacturer's instructions (fig. 4.1.2.1.1) (see section 3.2 for more information).

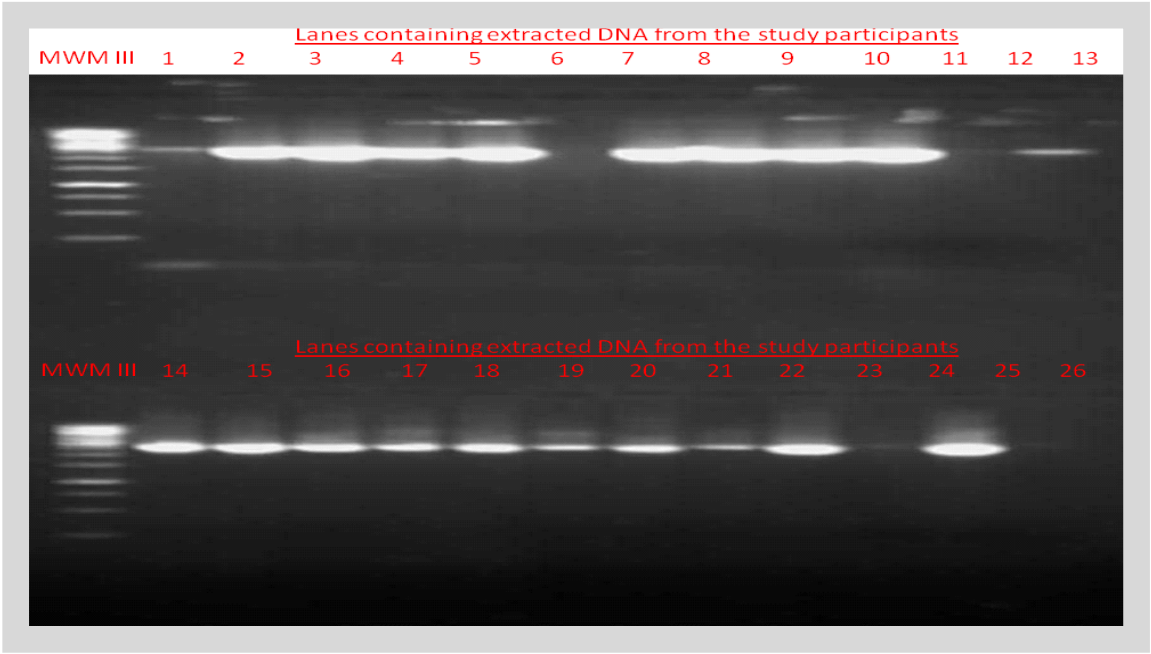

Figure 4.1.2.1.1. Analysis of DNA extraction by 0.8% agarose gel electrophoresis. MWM III, DNA ladder molecular weight marker III (Fermentas). Numbers 1-26, represents the lanes in which extracted DNA was loaded for analysis. Lane 6 and lane 26 contains negative controls that contained water instead of DNA.

215 (47 SP and 168 SN) out of 279 buffy coats samples gave good quality DNA yields suitable for genotyping analysis. SNPs A1650G and C1604G which are in the promoter region of the *PPIA* gene were analyzed by PCR-RFLP method (section 3.2.2). Firstly the DNA samples were amplified by PCR (fig. 4.1.2.1.2) to large quantities required for RFLP analysis.

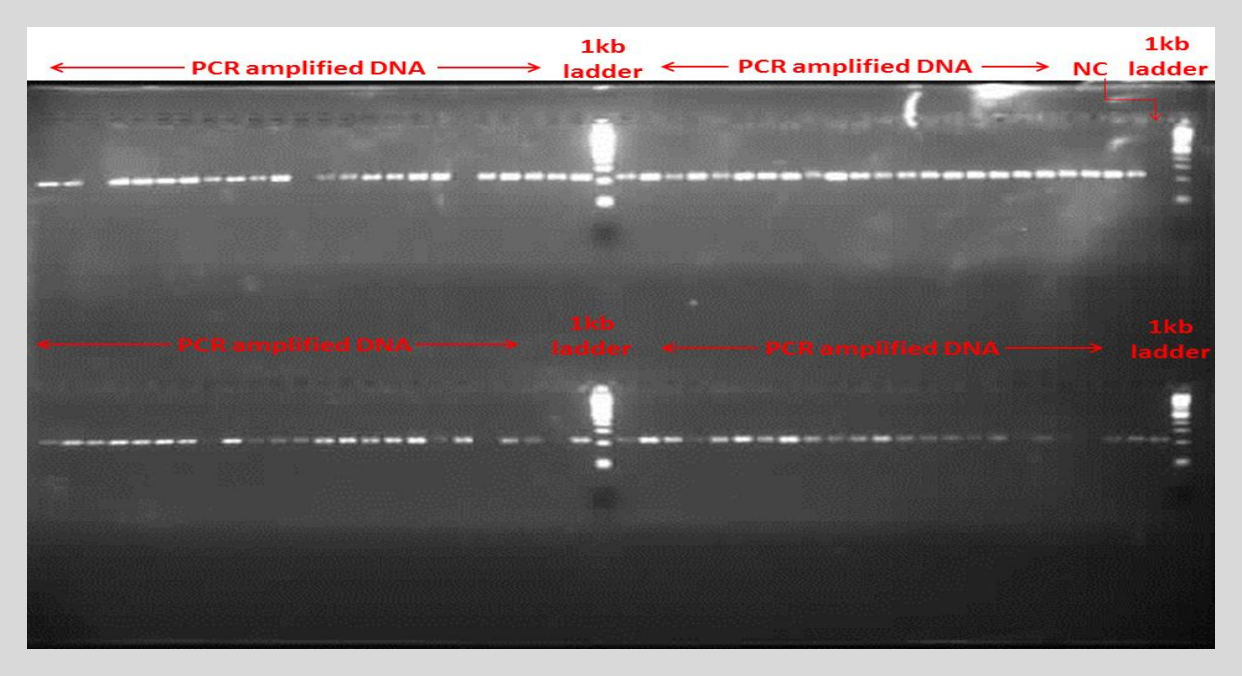

Figure 4.1.2.1.2. Polymerase chain reaction (PCR) amplification of DNA extracted from buffy coats obtained from study participants. 1kb DNA molecular weight marker (Fermentas) was used. <- PCR amplified DNA -> indicate the lanes containing DNA that was amplified by PCR. NC represents negative control.

Following PCR amplification, the PCR products were then subject to restriction digest by restriction enzymes for RFLP analysis. For this analysis two restriction enzymes *Rsa*I and *HAE*III were used for analyzing SNPs A1650G and C1604G, respectively. Restriction of the PCR products using *Rsa*I yielded distinct restriction fragments which are seen as DNA bands after 4% agarose gel electrophoresis analysis (fig. 4.1.2.1.3A). After restriction digestion with *Rsa*I, a PCR product containing homozygous wild type 1650AA genotype yielded two restriction fragments of 130 bp and 120 bp in size. Homozygous mutant 1650GG genotype did not get digested yielding undigested PCR product of 250 bp in size. Consequently, a

PCR product containing heterozygous 1650AG genotype yielded three restriction fragments of 250 bp; 130 bp and 120 bp in size. These results demonstrate that the homozygous mutant 1650GG completely removes *Rsa*I restriction site at this position.

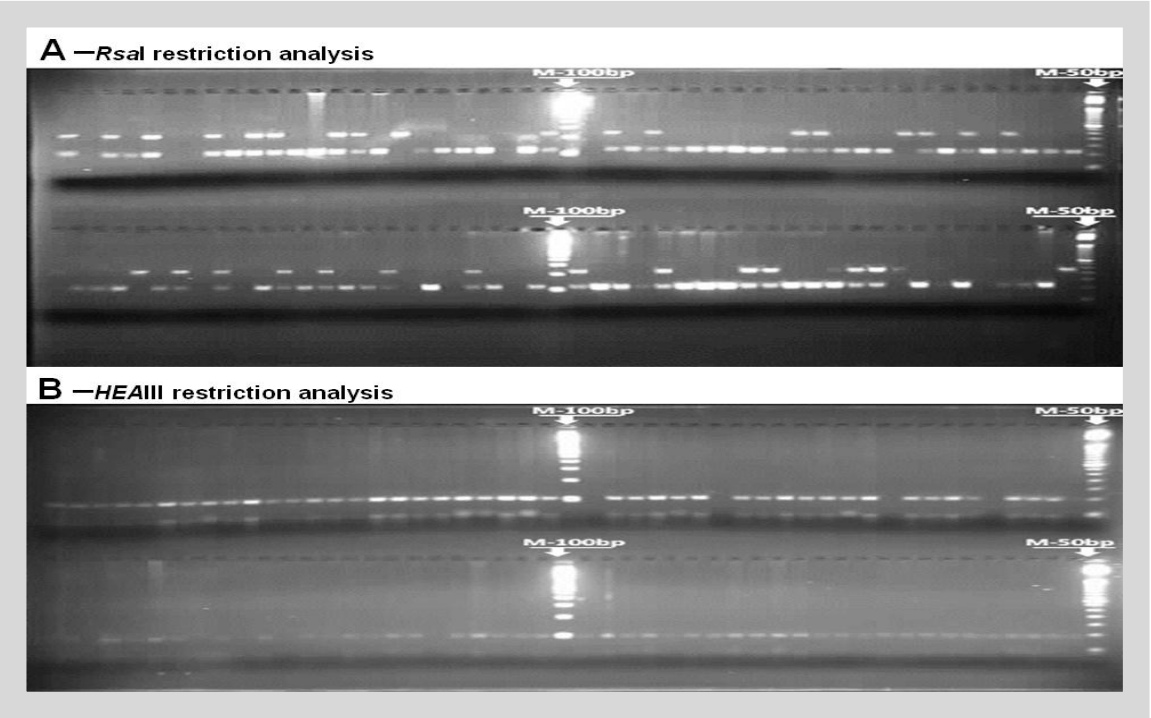

Figure 4.1.2.1.3. Restriction fragment length polymorphism (RFLP) analysis of PCR amplified DNA. RFLP analysis was performed using two restriction enzymes: **A,** RFLP banding patterns following restriction with *Rsa*I enzymes. **B,** RFLP banding patterns following restriction with *HEA*III enzymes.

In contrast, RFLP analysis after restriction with *HEA*III shows uniform banding patterns (fig. 4.1.2.1.3B) suggesting that participants in the CAPRISA AI 002 study cohort were not polymorphic at C1604G locus of the *PPIA* gene. Based on these findings, the SNP C1604G was excluded in the association analysis.

4.1.2.2 Influence of SNP A1650G on $CD4^+$ T cell counts and viral load.

Rits *et al*. did not find an association between the SNP A1650G and $CD4^+$ T cell counts or plasma viral RNA loads in the Amsterdam Cohort of studies (ACS) [74]. In this study we therefore tested the influence of the SNP A1650G on $CD4^+$ T cell counts and viral loads in the CAPRISA AI 002 cohort. The approximate time of HIV-1 infection was known for SPs in the CAPRISA AI 002 cohort [269]. Since viral loads and $CD4^+$ T cell counts fluctuate significantly during acute HIV-1 infection, we analyzed the data in two intervals post infection, 0-3 months (acute phase) and 3-12 months (early chronic phase) to detect possible differences between genotypes during these phases of infection.

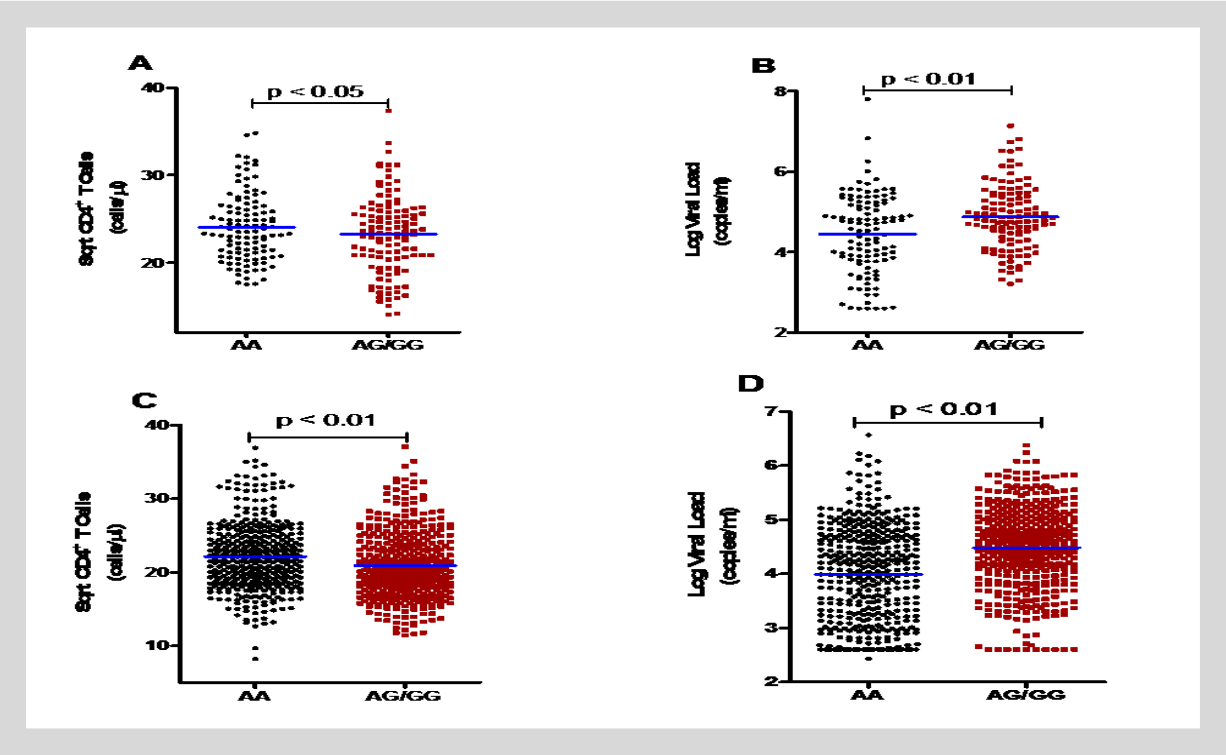

Figure 4.1.2.2. The association between different genotypes of the SNP A1650G and $CD4^+$ T cell counts and viral loads during the primary HIV-1 infection. **A** and **B** show the association of different genotypes with square root $CD4^+$ T cell counts and log viral loads during acute phase (0-3 months) of infection, respectively. **C** and **D** show the association of different genotypes with square root $CD4^+$ T cell counts snd log viral loads during early chronic phase (3-12 months) of infection, respectively. AA→ wild type genotype and AG/GG→mutant genotypes.

A dominant model analysis using GEE showed that the 1650G was significantly associated with lower $CD4^+$ T cell counts ($p=0.02$) (fig. 4.1.2.2A) and higher viral loads ($p<0.01$) (fig. 4.1.2.2B) during the acute phase of infection. Interestingly, similar results were observed

during the early chronic phase of infection where the 1650G was significantly associated with lower $CD4^+$ T cell counts ($p<0.01$) (fig. 4.1.2.2C) and higher viral loads ($p<0.01$) (fig. 4.12.2D).

Although An *et al*. did not find significant association of the1650G with clinical outcomes in either European American or African American [73], Bleiber *et al*. reported an association of the 1650G with rapid CD4 cell depletion in a Swiss Caucasian HIV-1 cohort and a trend toward higher *in vitro* HIV-1 replication [61]. Interestingly, the results obtained from the current study are consistent with the results obtained from a Swiss Caucasian HIV-1 cohort [61]. The current study reports the association of the 1650G with enhanced HIV-1 replication as reflected by significantly higher viral loads and lower $CD4^+$ T cell counts during primary HIV-1 infection (fig. 4.1.2.2).

4.1.2.3 Effects of SNP A1650G on disease progression

Next, we tested the influence of the SNP A1650G on disease progression in the SP group. Kaplan Meier and Cox proportional hazard analysis were used to test for potential differences in $CD4^+$ T cell decline between individuals who carried the 1650AA or 1650AG/GG genotype (dominant model). Our results show borderline significant association between the 1650AG/GG and rapid $CD4^+$ T cell decline (RH=3.18; p=0.05) (fig. 4.1.2.3).

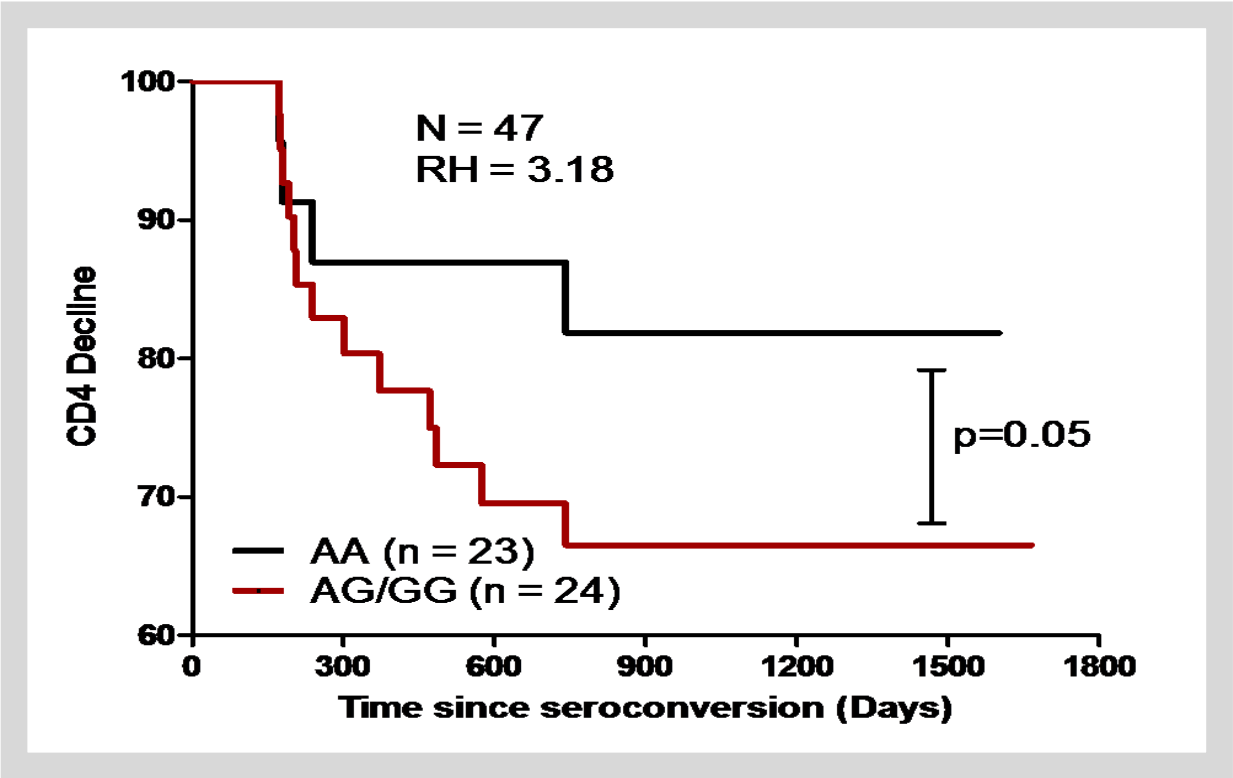

Figure 4.1.2.3. Kaplan Meier survival curves showing differences in rates of progression to $CD4^+$ T cells below 350 cells per ml between individuals who carried the 1650AA or 1650AG/GG genotype (dominant model).

This association supports the earlier observations (fig. 4.1.2.2) that the 1650G is associated with enhanced HIV-1 replication in the black South African HIV-1 (CAPRISA AI 002) cohort.

Table 4.1.2.3. Effect of the *PPIA* SNP A1650G on HIV-1 Acquisition Risk

SNP	Risk group	n	Genotype Number of Carriers (Frequency)			SP versus SN RH (95% CI)	*p*
A1650G			AA	AG	GG	AG + GG versus AA	
	SN	168	79 (0.42)	76 (0.41)	13 (0.07)		
	SP	47	23 (0.49)	18 (0.38)	6 (0.13)	1.57 (0.64 – 3.62)	0.31

RH - represents the relative hazard of being HIV-1-SP compared to being HIV-1-SN.
n - number of individuals.
P-value < 0.05 was considered significant.
Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effect of SNP A1659G on time to HIV-1 acquisition risk.

Although the 1650G frequency was higher in the SP group as compared to the SN group, this difference was not significant (p=0.31) (Table 4.1.2.3). Unlike the previous studies [73, 74], our results suggest that the 1650G may not be associated with susceptibility to HIV-1 infection in this setting

4.1.2.4 Effect of 1650G on CypA expression levels in vivo.

Next we analyzed whether the SNP A1650G of the *PPIA* gene influenced CypA mRNA expression levels *in vivo*. We performed quantitative RT-PCR to measure CypA mRNA levels in 30 SNs and 28 SPs. The data obtained from this experiment indicates that PBMCs obtained from SPs had elevated levels of CypA mRNA expression compared to PBMCs obtained from SNs, however, the difference was not significant (fig. 4.1.2.4A). For 13 of SPs, samples pre- (baseline) and post-HIV-1 infection were available for analysis. Baseline PBMCs from seroconverters had higher CypA expression compared to PBMC from non-

seroconverters but this was not significantly different (fig. 4.1.2.4B). These results provide no convincing evidence that CypA might be associated with susceptibility to HIV-1 infection.

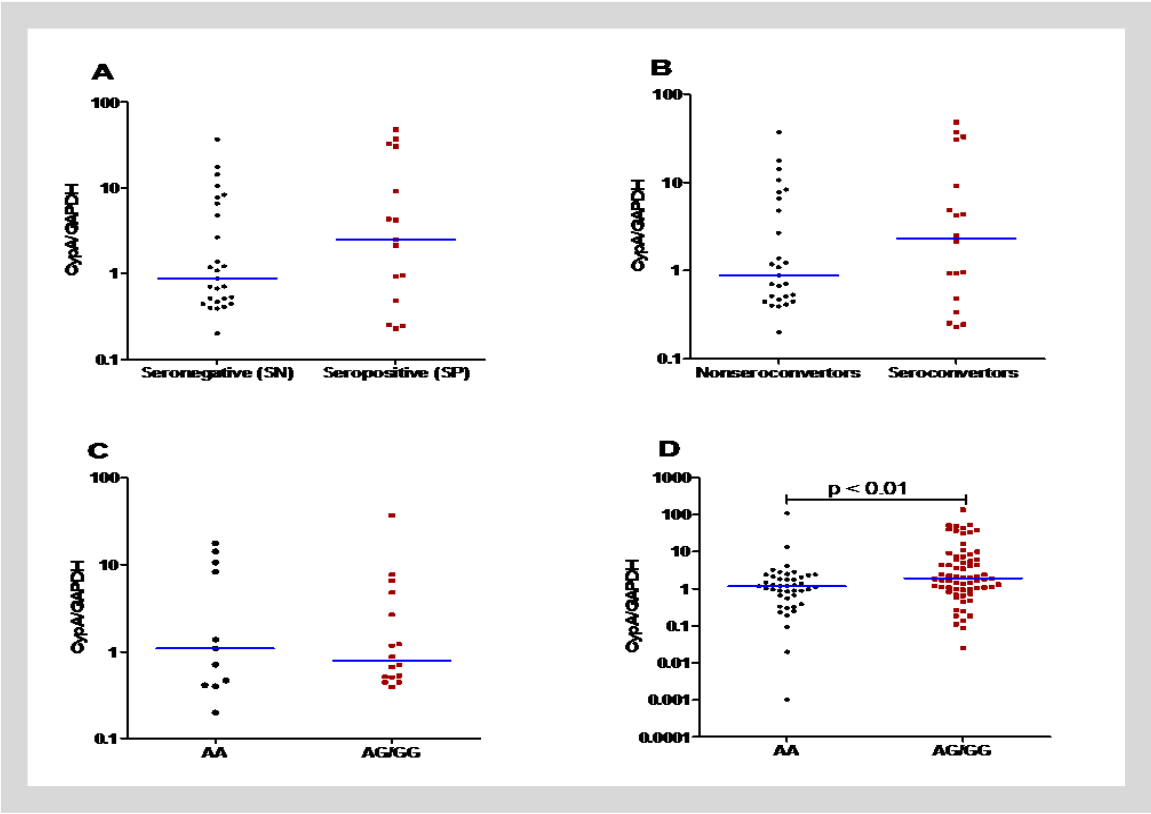

Figure 4.1.2.4. CypA mRNA Expression in peripheral blood mononuclear cells (PBMCs) represented as the normalized ratio of CypA to GAPDH. **A,** Expression of CypA mRNA in PBMCs obtained from SNs versus SPs. **B,** Expression of CypA mRNA in PBMCs obtained seroconverters versus nonseroconverters. **C,** Expression of CypA mRNA as modulated by different genotypes among SNs. **D,** Expression of CypA mRNA as modulated by different genotypes among SPs. ***AA***→ wild type and ***AG/GG***→ mutant genotypes for SNP A1650G.

CypA expression between the 1650AA and 1650AG/GG genotypes was not significantly different amongst the SN group (fig. 4.1.2.4C). Interestingly, the 1650G was significantly associated with elevated levels of CypA mRNA expression post HIV-1 infection ($p<0.01$) (fig. 4.1.2.4D) suggesting that the 1650G drives higher expression of CypA mRNA levels which in turn promotes HIV-1 replication *in vivo* in the black South African cohort.

4.1.2.5 Effect of 1650G on CypA expression levels and HIV-1 replication *in vitro*

Next we analyzed whether the SNP A1650G was associated with altered CypA expression levels *in vitro*. Again, we performed quantitative RT-PCR on CypA mRNA levels in PBMCs of 26 SN (or highly exposed but persistently seronegative (HEPS)) group, genotyped for A1650G SNP. Pre-HIV-1 infection expression levels of CypA were significantly higher in PBMC genotyped for the1650G compared to CypA expression levels in PBMC harboring the 1650AA genotype ($p<0.01$) (fig. 4.1.2.5A). At day 2 post infection, we observed a significant drop in CypA mRNA levels in PBMC genotyped for the 1650G.

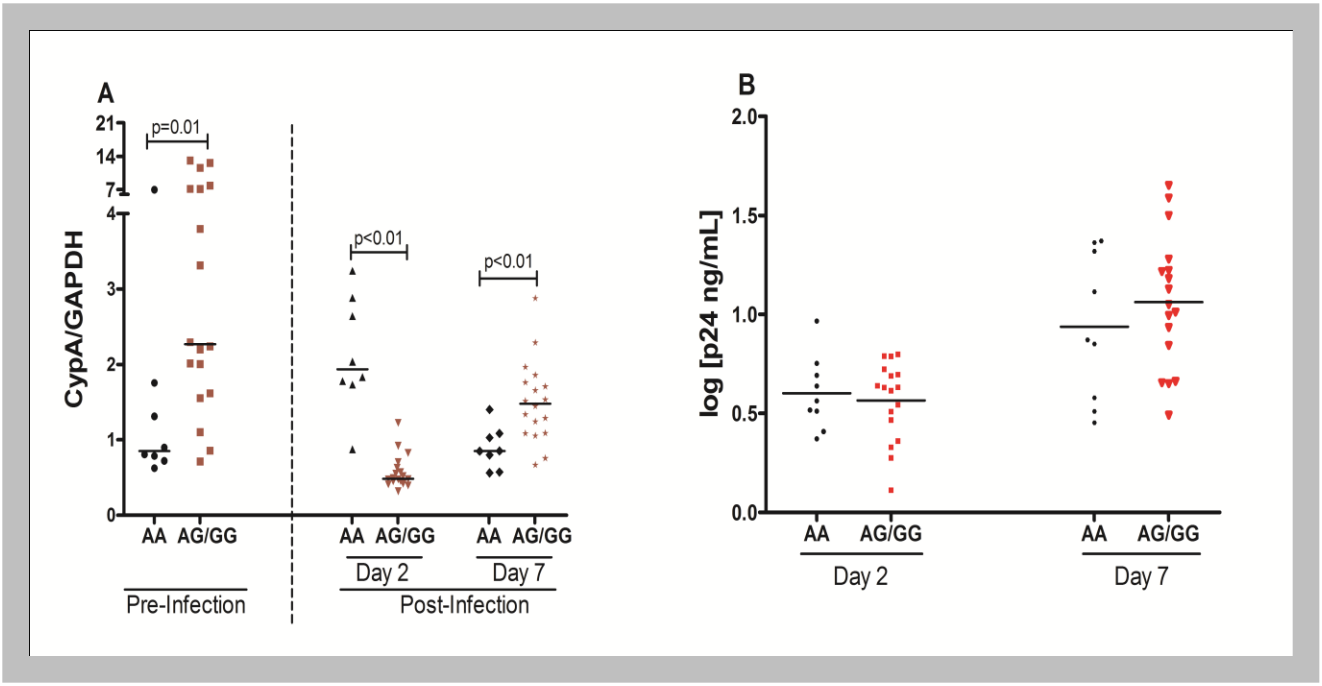

Figure 4.1.2.5. Expression of CypA mRNA in PBMCs represented as the normalized ratio of CypA to GAPDH. PBMCs were obtained from highly exposed but persistently seronegative (HEPS). **A,** expression levels of CypA mRNA pre-and post-HIV-1 infection of PBMCs. **B,** replication assays of HIV-1 in PBMCs harboring different genotypes of the SNP A1650G.

Interestingly, at day 7 post infection, our results demonstrated that the expression levels of CypA in PBMCs genotyped for the 1650G were restored to significantly higher levels ($p<0.01$) (fig. 4.1.2.5A). These results demonstrated that CypA expression is altered by HIV-1 infection. Elevated levels of CypA expression from day 7 onwards support the findings presented above that 1650G promotes HIV-1 replication and rapid disease progression.

In order to determine when does the 1650G start to have an impact on HIV-1 replication, we analyzed the replicative capacity of HIV-1 in IL2 stimulated PBMC from 26 HEPS individuals

with known genotypes for the SNP A1650G. IL2-stimulated PBMC were inoculated with 5ng of NL4.3 per $2X10^6$ cells. Subsequently, virus replication was assessed by measuring p24 antigen levels in the culture supernatant every other day for 7 days. Similar to the previous study [74], our results did not show significant differences in p24 production at any time point during the culture period, irrespective of the genotype of the SNP A1650G (figure 4.1.2.5B). These results suggest that the effect of the SNP A1650G is not shown in short-term tissue culture and therefore its effect may be subtle rather than dramatic.

4.1.3 Discussion and Conclusion

The role of CypA in promoting HIV-1 replication was initially established through extensive *in vitro* experiments [134, 135, 145, 274]. Subsequent studies validated CypA as an important cellular cofactor of HIV-1 by identifying SNPs within the *PPIA* gene that are associated with HIV-1 disease outcome [61, 73, 74]. Although numerous SNPs within the *PPIA* gene have been implicated to HIV-1 clinical outcome [73], this study focused only on one SNP, A1650G, because it was polymorphic in a cohort of black South African women (CAPRISA AI 002). The minor allele (1650G) was associated with enhanced *in vivo* HIV-1 replication as evident from significantly higher viral loads and lower $CD4^+$ T cell counts following HIV-1 infection in the CAPRISA AI 002 cohort. These results corroborate the data obtained from a Swiss Caucasian HIV-1 cohort and *in vitro* data which showed a trend towards higher HIV-1 replication [61]. However, the 1650G was not associated with disease progression in either the European Americans or African Americans [73].

While there was no evidence of association between the 1650G and susceptibility to HIV-1 infection in both the American and Swiss cohorts [61, 73], results obtained from the Amsterdam cohort suggest that 1650G may be associated with protection against HIV-1 infection [74]. In this study, we report that the 1650G was similarly distributed between SN and SP groups of the CAPRISA AI 002 cohort. Thus, the role of the 1650G in susceptibility to HIV-1 infection remains inconclusive and warrants further investigation. The discrepancy in the observed phenotypes could be attributed to the fact that these studies used populations from different geographical areas. There are differences in allele frequencies among potential disease influencing gene variants between ethnic groups and geographically separated populations [107].

The entire *PPIA* gene was re-sequenced and no genetic variation was found in the coding sequence of the *PPIA* gene and the dbSNP database has no documented records of variation in the coding sequence of the *PPIA* gene [73]. This suggests that differential genetic impact of *PPIA* SNPs on HIV-1 replication and disease progression are likely due to the 1650G affecting protein levels since the SNP A1650G is in the regulatory region of *PPIA* gene. Although CypA mRNA expression levels are altered by HIV-1 infection, our results demonstrated that PBMCs genotyped for the 1650G expresses higher levels of CypA mRNA than PBMCs genotyped for wild type genotype, 1650AA. Rapid depletion of CD4 cells to less 350 cell/ml could be partially attributed to higher expression levels of CypA since participants genotyped for 1650G lost their CD4 cells more rapidly than wild type individuals.

The replicative capacity of HIV-1 (NL4.3) in IL2 stimulated PBMCs was higher in PBMCs genotyped for 1650G as was measured by p24 production. However, there was no significant difference in the amount of p24 produced between PBMCs genotyped for 1650AA and 1650AG/GG. Lack of significant difference could be attributed to a shorter period (7 days) of incubation of PMBC with HIV-1.

In summary, we demonstrate that SNP A1650G is associated with higher HIV-1 replication and rapid $CD4^+$ T depletion in black South Africans. These findings corroborate the understanding that the genetic variation of the *PPIA* gene influences HIV-1 disease progression [61, 73] and further confirm that CypA is a critical host protein crucial for efficient HIV-1 replication. Small molecule inhibitors may be designed to block the interaction between HIV-1 and CypA thereby inhibiting HIV-1 replication.

4.2 The Influence of Genetic Variation in the *TNPO3* Gene on Susceptibility to HIV-1 Infection and Disease outcomes

HIV-1 has the capacity to infect nondividing cells, including cells of monocyte/macrophage lineage. The early events of HIV-1 replication include reverse transcription to synthesize viral copy DNA (cDNA) from the RNA genome and subsequent integration of viral cDNA into the cellular chromosome. Reverse transcription and integration of HIV-1 cDNA occur within the context of higher-molecular-weight nuclear protein complexes derived from viral cDNA, proteins and host proteins. Cellular chromosome resides within the nucleus and therefore, HIV-1 must enter the nucleus to replicate successfully [32].

However, the interior of the nucleus is separated from the cytoplasm by a double-layer nuclear membrane possessing nuclear pore complexes that allow the trafficking of molecules between the cytoplasm and the nucleus in interphase (nondividing) cells [278]. Nuclear pores allow the diffusion of ions and molecules smaller than 9 nm in diameter. Other studies have reported nuclear pores to facilitate passage of larger molecules up 39 nm in diameter under certain conditions (reviewed in [31]. At 56 nm [30], viral cDNA, packed in the preintegration complex (PIC) grossly exceeds the diffusion limit and yet HIV-1 PIC is actively transported during interphase [32, 33], reflecting its ability to efficiently infect nondividing cells.

A number of HIV-1 proteins–matrix (MA), viral protein R (Vpr), integrase (IN), and the central DNA flap–constituting the PIC have been implicated in nuclear import. However, this aspect of HIV-1 biology remains highly controversial with reports both supporting and refuting the

role of these viral proteins in the nuclear import (reviewed in [75, 206, 279]). In addition to HIV-1 PIC proteins, host proteins, including the importin α/importin β heterodimer [35, 280, 281], importin 7 [280, 282, 283], NUP153 [284], and transportin-SR2 (TRNSR2) [178, 285] have been implicated in HIV-1 nuclear transport.

TRN-SR2 is a member of the karyopherin protein family that imports serine/arginine-rich splicing factors (SR proteins) into the nucleus [157]. TRN-SR2 has recently been identified as a cellular protein that interacts with HIV-1 integrase (IN) to import HIV-1 PIC into the nucleus [160, 285]. Recently, Engelman's group published results that discounted the important role for IN in determining the requirement for transportin 3 (TNPO3) during infection and highlighted a dominant role for CA in this aspect of HIV-1 biology [32]. The gene and its protein are referred to as *TNPO3* and TRN-SR2, respectively (section 1.6.2).

Previous research has shown that the disruption of this interaction inhibits HIV-1 replication *in vitro*. TRN-SR2 depletion significantly inhibited HIV-1 replication [67, 68, 160]. Christ *et al*. demonstrated that the inhibition was due to the blocking of PIC at nuclear import [160], indicating a potential role of TRN-SR2 in nuclear import of HIV-1 PIC. The role of TRN-SR2 in HIV-1 infection, so far, has been studied *in vitro*.

Identification of genetic variation and its association with disease outcome is a powerful way of validating the *in vitro* data. Previous studies have used this approach to identify genetic

variants that affect susceptibility to HIV-1 infection and disease progression. A 32 base pair deletion in the portion of the human *CCR5* open reading frame encodes a truncated protein, designate Δ32. Individuals who are homozygous for *CCR5Δ32* gene are protected against HIV-1 infection [286] and AIDS progression is delayed in heterozygous individuals [47, 49, 286-288].

Association of genetic variation and disease outcome is a powerful tool for validating the importance of a gene or gene product in HIV-1 infection and pathogenesis. Since the association of genetic variation of the TRN-SR2 gene (*TNPO3*) and HIV-1 disease outcome had never been studied before, the current study investigated the association of genetic variation in the *TNPO3* with susceptibility to HIV-1 infection and disease progression.

4.2.1 Materials and Methods

Study participants. Two South African cohorts CAPRISA AI 002 [269] and the Sinikithemba [271] were used for this study (see section 3.1).

4.2.1.1 Genotyping of SNPs in *TNP03* gene in the CAPRISA AI 002 Participants

6 intronic and 2 exonic haplotype tagging (ht) SNPs rs13242262, rs2305325, rs11768572, rs1154330, rs35060568, rs8043, rs6957529 and rs10229001 in the *TNPO3* gene selected as described in section 3.2.3 were genotyped in 247 patient samples (52 SPs and 195 SNs) from the CAPRISA AI 002 cohort. The ht SNPs were chosen because they are ht SNPs for

the *TNPO3* gene in the Yoruba population from Nigeria (http://www.snp.cshl.org). Genotyping was performed using the TaqMan SNP assay according to the manufacturer's protocol (Applied Biosystems). TaqMan assays were obtained from the Assay-by-Demand service of Applied Biosystems (http://www.appliedbiosystems.com). Eight negative controls which contained water instead of DNA were included in each plate to rule out PCR contamination. Samples were genotyped in duplicate and genotypes were accepted after confirmation of no contamination or inconsistencies between duplicates.

4.2.1.2 Genotyping of the SNP rs1154330 of the *TNPO3* gene in the Sinikithemba Participants

Analysis of the *TNPO3* gene genotyping results from the CAPRISA AI 002 cohort showed an association between the SNPs in the *TNPO3* gene and HIV-1 outcomes. However, the interpretation of the data from the CAPRISA AI 002 cohort was complicated by a small sample size, only 52 SPs. Since the minor allele (G) of the SNP rs1154330 (rs1154330G) gave consistent results in the CAPRISA AI 002 cohort, we decided to extend the analysis of this SNP to a larger cohort of 450 HIV-1 chronically infected individuals from Sinikithemba cohort, to either confirm or refute the results obtained from the CAPRISA AI 002 cohort. Genotyping assays were performed as described above for CAPRISA AI 002 cohort.

4.2.1.3 Statistical Analysis

The differences in allele frequency distributions between the SP and SN groups were determined by Fisher's exact test (FET) for each SNP to test the null hypothesis that allele

frequencies were equally distributed between the SN and SP groups. The effect of each SNP on HIV-1 viral loads and $CD4^+$ T cell counts was determined using a GEE model [246] which takes into account the longitudinal measures for each participant. Viral loads were log transformed and the square root of $CD4^+$ T cell counts was used to normalize both $CD4^+$ T cell count and viral load measurements.

Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effect of each SNP on time to HIV-1 infection after enrollment in the CAPRISA AI 002 cohort and on the rate of $CD4^+$ T cell decline to less than 350 cells/μl (CD4<350). The rate of $CD4^+$ T cells decline was compared between the reference group (SPs containing the wildtype genotype) and a group with containing the mutant genotype (dominant genetic model) for each SNP. $P<0.05$ and $P<0.0015$ were considered to denote statistical significance before and after correction for multiple comparisons, respectively. Statistical significance after a Bonferroni correction for multiple comparisons is indicated by an asterisk.

4.2.2 Results

4.2.2.1 Description of *TNPO3* variations

The *TNPO3* gene is approximately 100.25kb in length consisting of twenty two exons (fig. 4.2.2.1). 6 intronic and 2 exonic ht SNPs available from dbSNP database (www.ncbi.nlm.nih.gov/SNP) were selected to genotype participants from the CAPRISA AI 002 cohort (fig. 4.2.2.1 and Table 4.2.2.1).

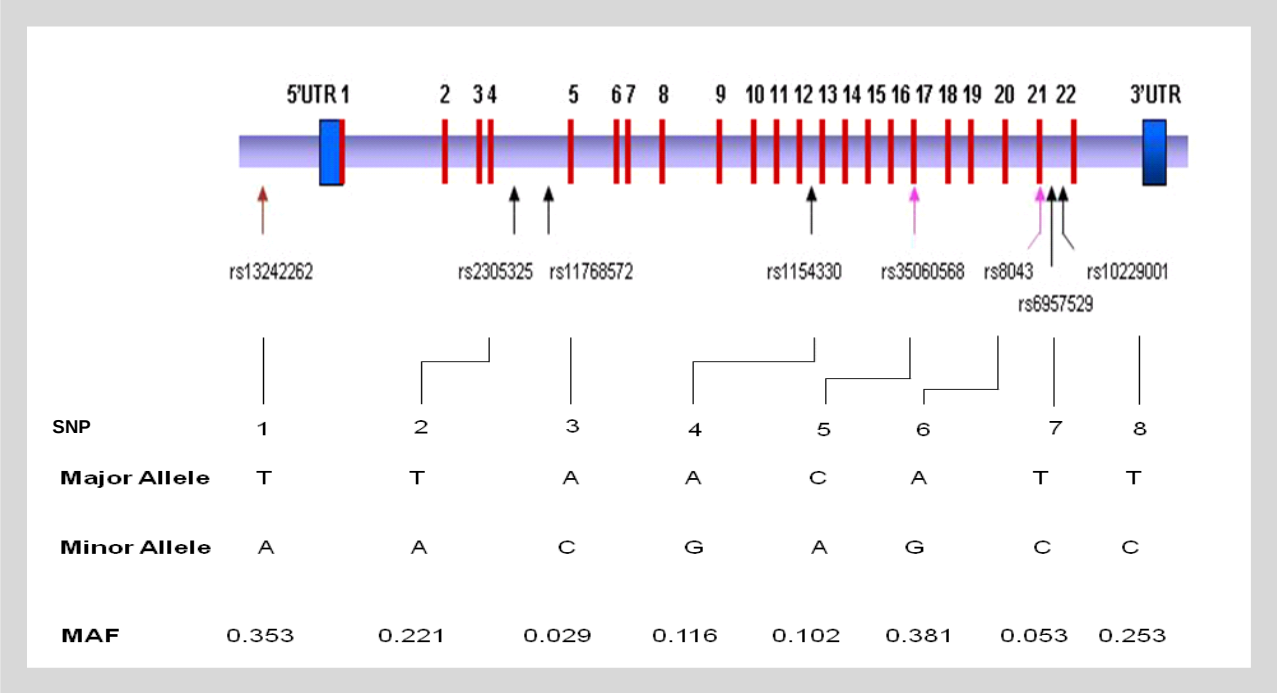

Figure 4.2.2.1. Locations of the *TNPO3* single nucleotide polymorphisms (SNPs) on chromosome 7. Coding exons of a gene are marked with red lines. Minor allele frequencies (MAF) of the SNPs are shown at the bottom panel.

The SNP rs1154330 was further genotyped in the Sinikithemba cohort. The genotypic frequencies of all SNPs conformed to the frequencies expected under Hardy Weinberg (HW) equilibrium (Table 4.2.2.1).

Table 4.2.2.1. Description of *TNPO*3 ht SNPs

dbSNP rs number	SNP number	Location	Nucleotide Change	Hardy Weinberg	MAF
rs13242262	SNP1	regulatory	T→A	0.55	0.35
rs2305325	SNP2	intron 4	T→A	0.86	0.22
rs11768572	SNP3	intron 4	A→C	5.84	0.03
rs1154330	SNP4	intron 12	A→G	0.11	0.12
rs35060568	SNP5	exon 17	C→A	21.52	0.10
rs8043	SNP6	exon 21	A→G	0.26	0.38
rs6957529	SNP7	Intron 21	T→C	12.00	0.05
rs10229001	SNP8	Intron 21	T→ C	0.92	0.25

Hardy Weinberg was calculated using the CAPRISA AI 002 cohort MAF- minor allele frequency

4.2.2.2 Effect of the *TNPO3* SNPs on Susceptibility to HIV-1 Infection in the CAPRISA AI cohort

The study design of the CAPRISA AI 002 cohort allowed us to study the influence of genetic variation of the *TNPO3* gene on susceptibility to HIV-1 infection. Firstly, we compared the

TNPO3 SNP allele frequency distribution between the SP and SN groups and did not find distortions in allele distribution between the groups as shown for SNP rs1154330G (fig. 4.2.2.2A).

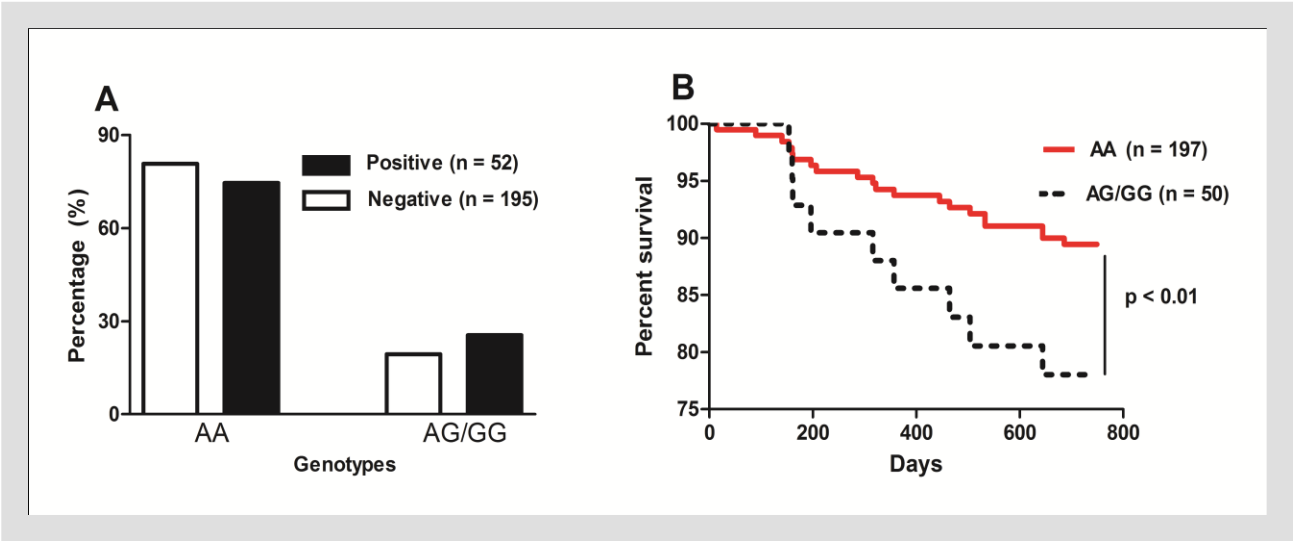

Figure 4.2.2.2. Association of the minor allele (G) of SNP rs1154330 (referred to as rs1154330G) with susceptibility to HIV-1 infection in the CAPRISA AI 002 cohort. **A,** frequency distribution of the rs1154330G between the SP and SN groups. **B,** The rs1154330G represented by a red curve on the graph showed significant association with fast acquisition of HIV-1 infection compared to the wild type allele (rs1154330A) at this locus (HR = 3.24, P < 0.01).

Kaplan-Meier survival analysis of time to HIV-1 acquisition and showed that the minor allele (G) of SNP rs1154330 (referred to as rs1154330G) was associated with faster acquisition of HIV-1 infection (HR 7.13, 95% confidence interval [CI]: 3.06 - 16.50; p<0.01, Cox model) (fig. 4.2.2.2B). Although the rs1154330G was more frequent amongst SPs, this was not

statistically significant. No association was observed between HIV infection and the rest of SNPs (Table 4.2.2.2).

Table 4.2.2.2. Effect of *TNPO3* SNPs on HIV-1 Infection Risk in CAPRISA cohort

rs number Risk Group	n	Genotype Number (Frequency)			22 and 12 versus 11 (reference group)* HR (95% CI)	P - value
rs13242262 (SNP1)		TT	AT	AA		
SN	195	78 (0.4)	97 (0.5)	20 (0.10)		
SP	52	24 (0.46)	20 (0.38)	8 (0.15)	2.55 (1.12 – 5.81)	0.73
rs2305325 (SNP2)		TT	AT	AA		
SN	195	120 (0.59)	68 (0.36)	7 (0.038)		
SP	52	32 (0.62)	15 (0.29)	5 (0.09)	2.71 (1.15 – 6.31)	0.64
rs11768572 (SNP3)		AA	AC	CC		
SN	195	180 (0.92)	15 (0.08)	0 (0.00)		
SP	52	50 (0.96)	2 (0.04)	0 (0.00)	0.00 (0.00 – 0.00)	0.99
rs1154330 (SNP4)		AA	AG	GG		
SN	195	158 (0.81)	31 (0.16)	6 (0.03)		
SP	52	39 (0.75)	13 (0.25)	0 (0.00)	7.13 (3.06 – 16.50)	<0.01
rs35060568 (SNP5)		CC	AC	AA		
SN	195	155 (0.79)	40 (0.20)	0 (0.00)		
SP	52	42 (0.81)	10 (0.19)	0 (0.00)	0.76 (0.18 – 3.21)	0.15
rs8043 (SNP6)		AA	AG	GG		
SN	195	102 (0.52)	69 (0.35)	24 (0.12)		
SP	52	23 (0.44)	20 (0.29)	9 (0.17)	2.09 (0.86 – 4.99)	0.91
rs6957529 (SNP7)		TT	CT	CC		
SN	195	172 (0.88)	23 (0.12)	0 (0.00)		
SP	52	47 (0.90)	5 (0.10)	0 (0.00)	3.41 (0.75 – 8.51)	0.81
rs10229001 (SNP8)		TT	CT	CC		
SN	195	108 (0.55)	74 (0.38)	13 (0.07)		
SP	52	30 (0.58)	19 (0.36)	3 (0.06)	3.28 (1.47 – 4.93)	0.33

4.2.2.3 Effect of the *TNPO3* SNPs on $CD4^+$ T cell counts and viral loads in the CAPRISA AI 002 Cohort

Since the approximate date of infection was known for CAPRISA AI 002 cohort, we analyzed the data in two specific intervals, 0-3 months post infection (acute phase of infection) and 3-12 months post infection (early chronic phase of infection). The influence of genetic variation on $CD4^+$ T cell counts and viral loads was performed in these specific intervals because viral loads and $CD4^+$ T cell counts fluctuate significantly during primary infection. The rs numbers instead of SNP numbers will be used below, please refer to the table above (Table 4.2.2.1) to find the corresponding SNP numbers for rs numbers. Dominant model analysis, using GEE which takes into account repeated measurements, showed that the minor alleles (A) and (C) of rs2305325 and rs11768572, respectively were significantly associated with higher $CD4^+$ T cell counts during the acute phase of infection (fig. 4.2.2.3A). The minor allele (A) of rs2305325 was associated with lower viral loads whereas the minor alleles (C) of SNPs rs11768572, (G) of rs1154330 (referred to as rs1154330G), (A) of rs35060568 and (C) of rs6957529 were significantly associated with higher viral loads during acute phase of infection (fig. 4.2.2.3B).

The rs2305325A and rs10229001C were associated with higher $CD4^+$ T cell counts while the rs1154330G, rs35060568A and rs6957529C were associated with lower $CD4^+$ T cell counts during the early chronic phase of infection (fig. 4.2.2.3C).

The rs10229001C was associated with lower viral loads while the rs1154330G, rs35060568A and rs6957529C were associated with higher viral loads during the early chronic phase of infection (fig. 4.2.2.3D). Interestingly, the rs1154330G was significantly associated with lower $CD4^+$ T cell counts and higher viral loads during both the acute phase and the early chronic phase of infection in the CAPRISA AI 002 cohort (fig. 4.2.2.3). This observation was consistent with the observation that the rs1154330G was significantly associated with increased susceptibility. However, rs8043 was neither association with $CD4^+$ T cell counts nor with viral loads during primary HIV-1 infection (fig. 4.2.2.3).

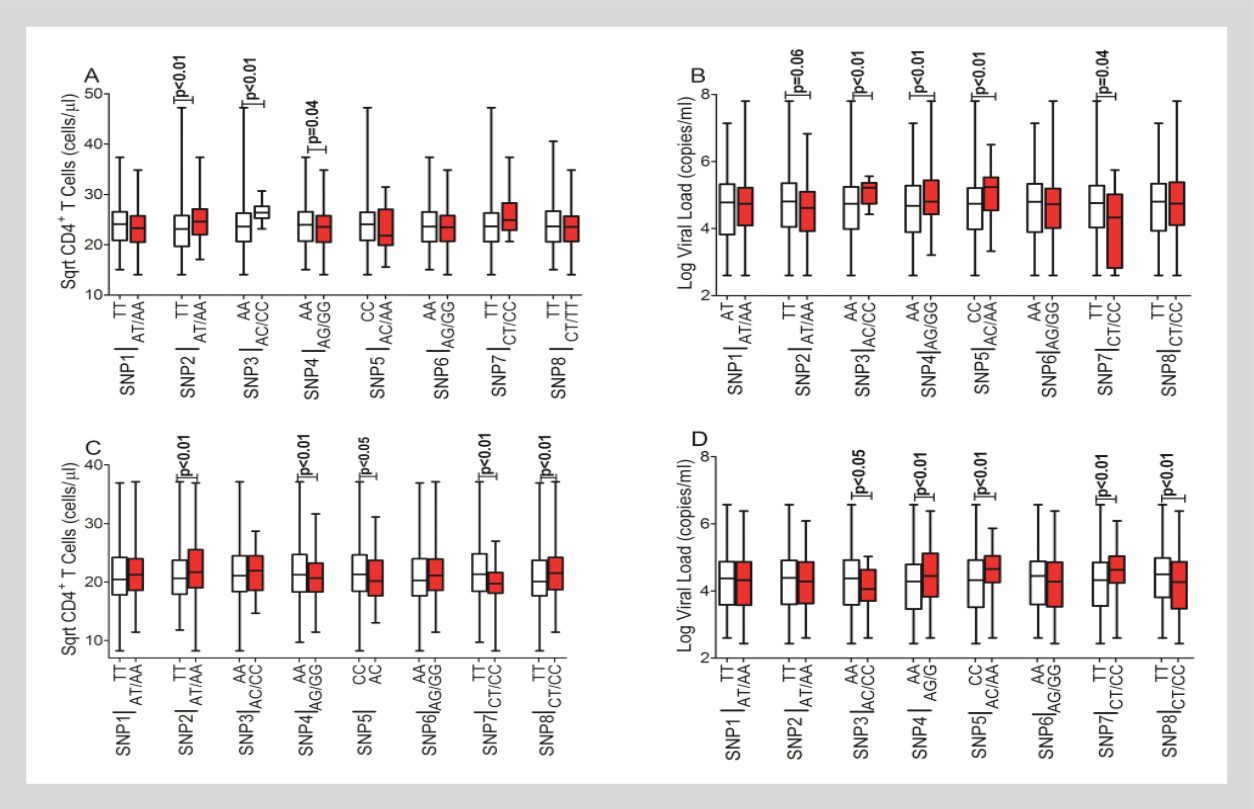

Figure 4.2.2.3. Effect of *TNP03* polymorphism on CD4$^+$ T cells and viral load in primary HIV-1 infected black South Africans (n =52). **A and B,** association of individual *TNPO3* SNP with differential CD4$^+$ T cell counts and viral loads, respectively, during the acute phase of infection. **C and D,** association of individual *TNPO3* SNP with differential CD4$^+$ T cell counts and viral loads, respectively, during the early chronic phase of infection (refer to Table 4.2.2.1 for corresponding rs number for the SNP number).

Next, we tested the effect of each SNP on the rate of CD4$^+$ T cell decline in the CAPRISA cohort using Cox model. However, no association was observed between the SNPs and the rate of CD4$^+$ T cell decline in the CAPRISA AI 002 cohort. Lack of association between *TNPO3* genetic variation and the rate of CD4$^+$ T cell decline could be attributed to a small

number of participants in the CAPRISA AI 002 cohort. For example, there were only 13 SPs who carried the rs1154330G.

4.2.2.4 Effect of the SNP rs1154330 of *TNPO3* on $CD4^+$ T cell decline in the Sinikithemba Cohort

The results obtained from the CAPRISA AI 002 cohort suggested that the rs1154330G may be associated with higher susceptibility to HIV-1 infection, lower $CD4^+$ T cell counts post infection and higher viral loads. However, the effect of the rs1154330G on the CD4 decline was not pronounced probably due to fewer (n=13) SPs containing the rs1154330G in the CAPRISA AI 002 cohort. We therefore extended the rs1154330G analysis to the larger Sinikithemba cohort comprising 450 HIV-1 chronically infected participants–within the same geographical area–in order to elucidate the effect of the rs1154330G on the rate of $CD4^+$ T cell decline. Since the Sinikithemba cohort is a seroprevalent cohort with unknown dates of infection, the effect of the rs1154330G was examined at baseline for $CD4^+$ T cell levels and on the trajectory of $CD4^+$ T cell decline over 5 years of follow-up. There was no association observed between the rs1154330G and $CD4^+$ T cell counts at baseline (fig. 4.2.2.4 A), probably due to the fact that participants enrolled into this study cohort were at different stages of the disease or infection. Another possibility is that there truly was no effect in this population**.**

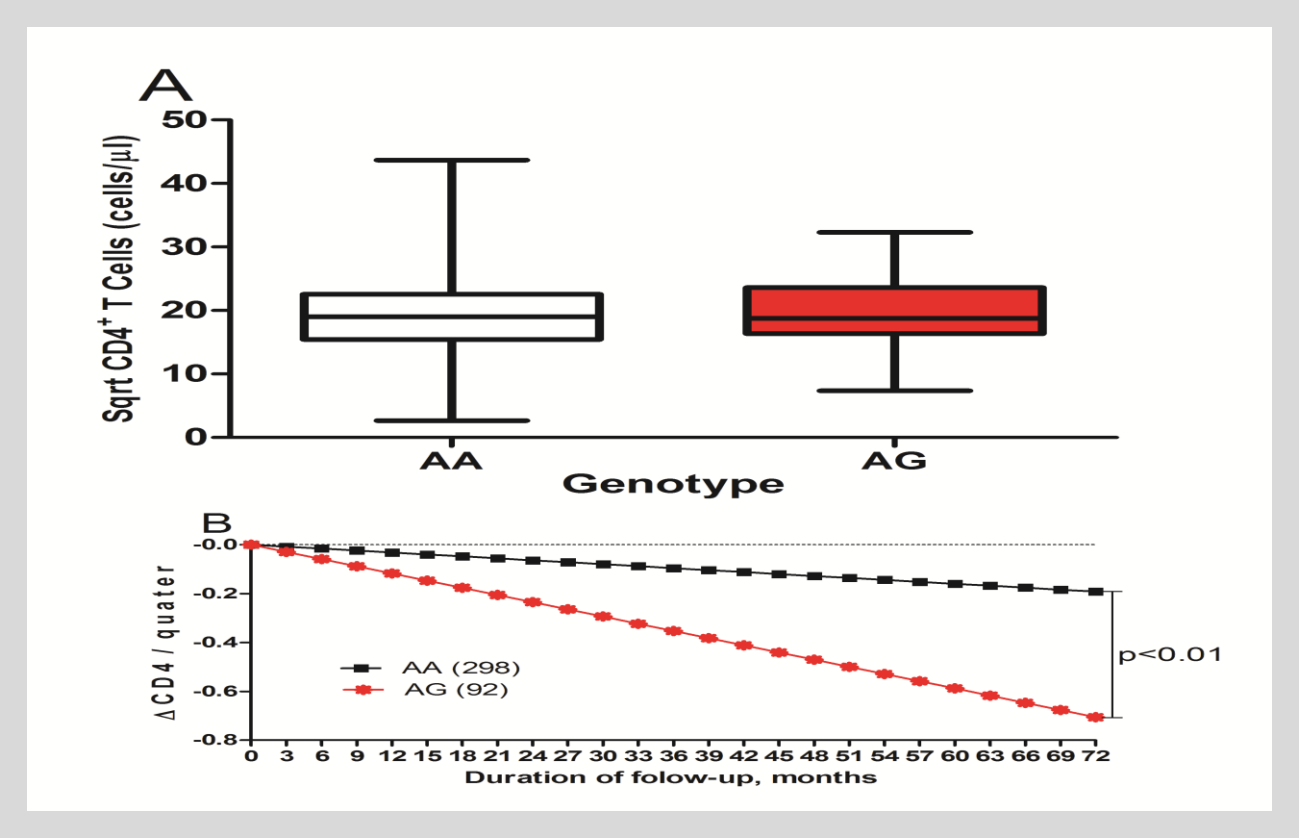

Figure 4.2.2.4. Rates of $CD4^+$ T cell decline stratified by *TNPO3* SNP rs1154330. **A,** influence of SNP rs1154330 on $CD4^+$ T cell counts at baseline; **B,** $CD4^+$ T decline in individuals carrying the wild type genotype (AA) for SNP rs1154330 are represented by a black curve whereas $CD4^+$ T decline in participants carrying the mutant genotype (AG) are represented by a red curve.

Interestingly, the rs1154330G was significantly associated with a faster rate of $CD4^+$ T cell decline in the Sinikithemba cohort ($p<0.01$) (fig. 4.2.2.4B). These results confirmed that the rs1154330G may be associated with enhanced HIV-1 replication *in vivo* as evident from increased susceptibility to HIV-1 infection, higher levels of HIV-1 loads and lower levels of $CD4^+$T cell count post infection in the CAPRISA AI 002 cohort (sections 4.2.2.2 and 4.2.2.3).

4.2.3 Discussion and Conclusions

Interestingly, the SNP rs1154330G was associated with increased susceptibility to HIV-1 infection as was reflected by faster acquisition of HIV-1 infection by individuals who harbored the rs1154330G. The rs1154330G was further associated with lower $CD4^+$ T cell counts and higher viral loads during primary infection in the CAPRISA AI 002 cohort. Lastly, the rs1154330G correlated significantly with rapid disease progression, as was reflected by the faster rate of $CD4^+$ T cell decline in in individuals carrying the rs1154330G in the Sinikithemba cohort. These results corroborate the role of TRN-SR2 as an important cellular cofactor for HIV-1 replication as previously established through *in vitro* studies [32, 67, 160].

In this study, we have investigated the association of genetic variation in the *TNPO3* gene with susceptibility to HIV-1 infection and disease progression in HIV-1 longitudinal South African cohorts. The rs1154330G, which is in intron 12 of *TNPO3*, was associated with increased susceptibility to HIV-1 infection and disease progression in the South African cohorts. These findings suggest that the rs1154330G gives HIV-1 selective advantage for infection and enhanced replication *in vivo*. In addition to the SNP rs1154330, 5 other SNPs also in the intron regions and 2 SNPs in the exon region were also studied in the CAPRISA AI cohort.

The rs2305325A was associated with higher $CD4^+$ T cell count and lower viral load during acute phase of infection. The rs2305325A was not associated with viral load during early chronic phase of infection. The rs35060568A was associated with higher viral load during

both the acute phase and the early chronic phase infection. However, the association between the rs35060568A and $CD4^+$ T cell counts was only observed during the early chronic phase of infection.

The rs6957529C showed inconsistent effect on $CD4^+$ T cell count and viral load between the acute and early chronic phase of infection. The rs6957529C was associated with higher $CD4^+$ T cell counts and lower viral loads during the acute phase of infection whereas the opposite effect was observed during early chronic phase of infection. The plausible explanation for this could be that the virus adapts differently to be able to replicate more effectively during the later stages of HIV-1 infection [202]. The rs10229001C was associated with higher $CD4^+$ T cell counts and lower viral loads during the early chronic phase of infection only.

The analysis of the SNP rs1154330 was extended to a larger Sinikithemba cohort of 450 chronically infected individuals on the basis that this was the only SNP that showed consistent results in the CAPRISA AI 002 cohort. In the Sinikithemba cohort, the rs1154330G could only be examined at baseline for its effect on $CD4^+$ T cell counts and viral loads since the date of infection was not known for participants in this cohort. The effect of the rs1154330G could also be assessed on the rate of $CD4^+$ T cell decline in this cohort. Interestingly, the rs1154330G was significantly associated with rapid $CD4^+$ T cell decline in the individuals harboring this allele. Since the rs1154330G is in the intron 12 region, therefore it probably exerted its effect by affecting the mRNA expression profile of the TRN-

SR2. However, the rs1154330G was neither associated with $CD4^+$ T cell count nor viral load at baseline. Lack of association could be explained by the fact that individuals in the Sinikithemba cohort were enrolled into the study while they were at different stages of HIV-1 disease.

This study had two major limitations: (1) the CAPRISA AI cohort, the only cohort with known dates of infection had few HIV-1 infected individuals and; (2) the date of infection was not known for HIV-1 chronically infected individuals, and therefore the clinical data of these individuals could not to be categorized and analyzed according to the different phases of HIV-1 infection.

Results obtained in this study demonstrated that the rs1154330G was significantly associated with increased susceptibility to HIV-1 infection. This allele was also associated with higher levels of viral loads and lower levels of $CD4^+$ T cell counts during primary infection and was further associated with rapid disease progression as was reflected by rapid $CD4^+$ T cell decline. This study has to be replicated in a larger cohort of HIV-1 infected individuals whose date of infection is known in order to elucidate the role of the genetic variation in*TNPO3* gene during the course of HIV-1 infection.

4.3 Association of Polymorphisms in the LEDGF/p75 Gene (*PSIP1*) with Susceptibility to HIV-1 Infection and Disease Progression

HIV-1 requires host cell factors to successfully complete its replication cycle [289]. Lens epithelium derived growth factor p75 (LEDGF/p75) also known as proprotein convertase 4 (PC4) or splicing factor, arginine/serine-rich 1 (SFRS1) interacting protein 1 (PSIP1), is a ubiquitous protein expressed in a wide range of tissues and cell types at all stages of development (by convention, the protein is referred to as LEDGF/p75 and the gene named *PSIP1*). LEDGF/p75 is known for its involvement in HIV integration and its important role in promoting HIV-1 replication *in vitro* [171, 179]. LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF) family, members of which are involved in chromosomal replication, transcription and chromatin structure formation [174-176, 178-180, 183].

LEDGF/p75 has been shown to interact with HIV-1 integrase (IN) through specific binding that occurs between the integrase binding domain (IBD) of LEDGF/p75 and the catalytic core domain of HIV-1 IN to mediate the nuclear accumulation of IN and to target HIV-1 integration into active transcription unit sites [187, 188, 201, 202, 290, 291]. Disruption of the interaction between LEDGF/p75 and HIV-1 IN, either by HIV-1 IN mutations or by LEDGF/p75 knockdown, leads to an inhibition of HIV-1 replication [177, 189, 292]. The data presented above from intensive *in vitro* studies have convincingly demonstrated that LEDGF/p75 is an important host factor for HIV-1 replication.

Studies of the influence of human genetic variation on susceptibility to HIV-1 infection and disease progression may reveal the essential *in vivo* host factors that interact with HIV-1 and their epidemiologic importance at the population level [73]. Although this approach has previously been used to identify HIV/AIDS modifying variants in human genes, the majority of these studies have been conducted using populations mainly from developed countries [73, 75, 286, 293]. However, it has been clearly demonstrated that there are considerable differences in allele frequencies among potential disease influencing genes between ethnic groups and geographically separated populations [107]. There is therefore a pressing need to extend host genetic studies of HIV-1 infection to developing world populations heavily burdened with HIV/AIDS and to factors that are possible targets for therapeutic or other plausible biomedical interventions. Although LEDGF/p75 has been demonstrated to be essential for HIV-1 integration *in vitro*, there was no data published on the influence of genetic variation in *PSIP1* on HIV-1 disease outcome. In this study, we therefore investigated the influence of genetic variation in *PSIP1* on HIV-1 infection and disease progression in South Africans cohorts.

4.3.1 Materials and Methods

Study participants. The Center for the AIDS Programme of Research in South Africa Acute Infection 002 (CAPRISA AI 002) [268, 269] and the Sinikithemba [270, 271] cohorts were used for this study (see section 3.1 for more details). Ethical approval for this study was obtained from the University of KwaZulu-Natal's Biomedical Research Ethics Committee. All participants provided written informed consent.

4.3.1.1 Identification of polymorphisms in the C terminal region of *PSIP1*

Single nucleotide polymorphisms (SNPs) in the C terminal region of *PSIP1* were screened in a panel of 83 seronegative (SN) and 43 seropositive (SP) black South African women from the CAPRISA AI 002 cohort by resequencing part of the C-terminal region of the *PSIP1* gene. *PSIP1* encodes two proteins by alternative splicing, LEDGF/p75 and p52 that have a high degree of homology but p52 lacks the C-terminus region of LEDGF/p75. Sequence comparison and BLAST search were performed to select LEDGF/p75-specific PCR primers. A 50 µl PCR reaction mixture containing 1X PCR buffer, 3.5 mmol $MgCl_2$, 0.25 mmol dNTP mix, 0.25 U/µl Expand *Taq* polymerase, 0.4 ρmol/µl of forward (LEDGFDNA1) and reverse (Seq 5) primer and 100 ng DNA was prepared (Table 4.3.1.1).

This was amplified at 95 ºC for 10 minutes, 35 cycles of 94 ºC for 30 seconds, 60 ºC for 30 seconds, 72 ºC for 45 seconds, and a final 10 minute extension step at 72 ºC. The PCR product was purified using Qiagen PCR Purification Kit (Qiagen) and sequenced using overlapping primers to cover a 1,679 bp long DNA fragment of the C-terminal region of *PSIP1* (starting from 1,190 bp upstream of exon 10 to 105 bp downstream of exon 13), this region included the IBD (Table 4.3.1.1). BigDye Terminator Kit (Applied Biosystems, Foster City, California, USA) was used for sequencing.

Table 4.3.1.1. PCR, sequencing and real time RT-PCR primers and cycling conditions used in this study

	Gene Name, Accession, SNP and rs number	**Primer /probe name/nucleotide change**	**PCR Primer Sequence 5'→3'** **TaqMan Assay ID# or Assay sequences**
PCR	*PSIP1* AF_199339	LEDGF1 Seq5	F: TGG GCT CAA AGC ATTA ATC C R: TCT GTG GCG TAT ACA CAG TG
Sequencing		Seq1 Seq6 Seq2 Seq5 S9 R1	F: GCC AGA TATGAT TTA ATC TAG R: GTA GAC TTT TCC ATG ATT CCT GAC F: GCC TGT ATA TAG AAA TAC TGG R: TCT GTG GCG TAT ACA CAG TG F: CTT CAA AGG ATA CAT GC R: GTA GAC TTT TCC ATG ATT CCT GAC
TaqMan Assay	SNP1→rs2277191 SNP2→rs10283923 SNP3→rs12339417 SNP4→rs1033056 SNP5→rs61744944	G → A G → C T → C A → G A → T	C_15883595_10 C_29529242_10 C_31936110_10 C_2757693_20 PSIP1_Q472L_s AAAACCAAAGATCAAGGGAAGAAA Psip1_q472l_a TGTGAAATTGTTGGCTTTTTACCA
Real Time –RT PCR		PTZ1 PTZ2	F: GTC AAC CCC ACC GTG TTC TTC R: TTT CTG CTG TCT TTG GGA CCT TG
	GAPDH NM_002046	GAPDH1 GAPDH2	F: AAG GTC GGA GTC AAC GGA TT R: CTC CTG GAA GAT GGT GAT GG

4.3.1.2 Genotyping of SNPs in *PSIP1* from the CAPRISA AI 002 Participants

The rs numbers instead of the SNP numbers will be used below, please refer to the table above (Table 4.3.1.1) to find the corresponding SNP numbers for rs numbers. The four intronic ht SNPs rs2277191, rs1033056, rs10283923 and rs12339417 available from NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP) and HapMap databases (http://www.hapmap.org) were selected by considering location, spacing, and allele frequency of at least 10%. The exonic SNP rs61744944 (Q472L) also available from the aforementioned databases (fig. 4.3.1.2, and Table 4.3.1.1) was selected based on its association with clinical outcome in the preliminary analysis (see section 4.3.21). Therefore all 5 SNPs were genotyped in 247 patient samples (195 SNs and 52 SPs) from the CAPRISA AI 002 cohort.

The intronic ht SNPs were chosen because they are tag SNPs for the *PSIP1* in the Yoruba from Nigeria (http://www.snp.cshl.org) and SNP rs61744944 based on the preliminary analysis of the sequencing data that suggested an association between the minor allele (T) of SNP rs61744944 (rs61744944T) with increased susceptibility to HIV-1 infection, lower $CD4^+$ T cell counts post HIV-1 infection and higher viral loads. Genotyping was performed by TaqMan SNP assay as per manufacturer's protocol (Applied Biosystems). TaqMan assays were obtained from the Assay-by-Demand service of Applied Biosystems (http://www.appliedbiosystems.com). Eight negative water controls were included in each run to rule out PCR contamination. Samples were genotyped in duplicate and genotypes were accepted after confirmation of no contamination or inconsistencies between duplicates.

4.3.1.3 Genotyping of SNPs in *PSIP1* from the Sinikithemba Participants

Analysis of the *PSIP1* polymorphisms in the CAPRISA AI 002 cohort showed an association between some SNPs in *PSIP1* and HIV-1 outcomes. We therefore extended the analyses to a larger cohort of 450 SPs from Sinikithemba cohort, to either confirm or refute the results obtained from the CAPRISA AI 002. Genotyping assays were performed as described above.

4.3.1.4 LEDGF/p75 mRNA expression Analysis

Peripheral blood mononuclear cells (PBMCs) from 57 HIV-1-negative participants and 38 recently infected participants from the CAPRISA AI 002 cohort were isolated and LEDGF/p75 expression levels were measured by quantitative RT-PCR. RNA was extracted from 2 x 10^6 PBMCs immediately after thawing using Trizol LS reagent (Invitrogen, Carlsbad, California, USA) according to the manufacturer's protocol. RNA was reverse transcribed using the iScript complementary DNA (cDNA) synthesis kit (Bio-Rad). LEDGF/p75 mRNA expression was quantified by quantitative TR-PCR using SYBR Green Chemistry on a Roche LightCycler version 1.5 (Roche Diagnostics). LEDGF/p75 cDNA was amplified using primers (Table 4.3.1.1) designed to uniquely amplify the IBD of *PSIP1* (GeneBank accession no. NM_021144). Each PCR reaction contained 3 mmol/µl $MgCl_2$, 0.5 ρmol/ml of each primer, 1 X LightCycler FastStart DNA Master SYBR Green I (Roche), 1 µg of cDNA, and water added to 10 µl final volume. Reactions were run in duplicate. The PCR conditions were: 1 cycle at 95 ºC for 10 min, 45 cycles of 95 ºC for 6 s, at 60^{o}C for 6s, and at 72 ºC for 6s.

Standard curves were generated using 10-fold serial dilutions of cDNA of known concentration. Samples and standards were run in duplicates and average values were used to compute gene copy number. The housekeeping gene, glyceraldehyde 3-diphosphate dehydrogenase (GAPDH) (NM_002046), was used to normalize for variations in cell count or differences in nucleic acid extraction as previously described [250]. Relative mRNA levels of LEDGF/p75 to GAPDH were calculated for each sample by dividing the concentration of LEDGF/p75 by the concentration of GAPDH.

4.3.1.5 Expression and purification of recombinant proteins

His_6-tagged HIV-1 integrase, 3X flag-tagged LEDGF/p75, MBP-JPO2 and MBP-pogZ were purified for AlphaScreen applications as described previously [258, 259].

4.3.1.6 AlphaScreen

The concentration units of the reagents used from this section onwards would be written as they appear in the article that was published from this work.

The AlphaScreen assay is a technique used to measure protein-protein affinity interactions. To measure the influence of the SNP rs61744944 (Q472L) on the binding affinity of LEDGF/p75 we performed AlphaScreen assay according to the manufacturer's protocol (Perkin Elmer, Benelux). Reactions were performed in 25 µl volume in 384-well Optiwell™ microtiter plates (Perkin Elmer). The reaction buffer contained 25 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1 mM $MgCl_2$, 0.01% (v/v) Tween-20 and 0.1% (w/v) bovine serum albumin.

300nM His_6-tagged IN was incubated with 100 nM flag-LEDGF/p75 for an hour at 4°C. Subsequently 5 µl of Ni-chelate–coated acceptor beads and 5 µl anti-flag donor beads were added to a final concentration of 20 µg/ml of both beads. Proteins and beads were incubated for 1 hour at 30 ºC. Exposure of the reaction to direct light was avoided and the emission of light from the acceptor beads was measured in the EnVision plate reader (Perkin Elmer, Benelux) and analyzed using the EnVision manager software. Assays with JPO2 or pogZ, respectively, were essentially performed as described previously [294, 295].

4.3.1.7 Complemented cell lines

Complemented HeLaP4/CCR5 knockdown cells (A3 clone) were generated and grown as described earlier [173]. Briefly, the Q472L mutation was introduced in pLNC_LEDGF BC-Ires-Bsd and MLV-based vectors were generated [204]. Following transduction, cells were selected with 3 µg/ml blasticidin (Invitrogen, Merelbeke, Belgium). Protein expression was verified by Western blot analysis and immunocytochemistry (data not shown).

4.3.1.8 Virus strains

The molecular clone pNL4.3 was obtained through the NIH AIDS Research and Reference Reagent Program. Virus stock ($HIV_{NL4.3}$) was produced as described earlier [204].

4.3.1.9 HIV-1 breakthrough assay

Cells were seeded at 30,000 cells per well in a 6-well dish and infected as described earlier with minor modifications [296]. Briefly, cells were infected with 56,000 pg p24 $HIV_{NL4.3}$ in a

total volume of 2 ml (MOI 0.01). 24 hrs later, cells were washed twice with 1X PBS prior to addition of 4 ml of fresh medium. HIV replication was monitored by sampling the culture medium for p24 ELISA (HIV-1 p24 ELISA kit, Perkin Elmer).

4.3.1.10 Quantitative PCR

Integrated proviral copies were quantified by real-time quantitative PCR (Q-PCR) on genomic DNA using the iQ5 Multicolor RT PCR detection system (BioRad, Nazareth, Belgium). In order to allow quantification of integrated proviral copies in HIV-$1_{NL4.3}$ infected cells, cells were subcultured at day 6 and grown under azidothymidine/ritonavir treatment for 10 days, 0.5µM and 1.5µM, respectively, i.e. 25-fold IC_{50} as determined in MT4/MTT assay [297] to eliminate all non-integrated viral DNA. Genomic DNA was extracted using the GenElute mammalian genomic DNA miniprep kit (Sigma, Bornem, Belgium); for each reaction 100 ng was used for Q-PCR. Integrated copies for HIV-$1_{NL4.3}$ were quantified using a Gag-derived primer-probe set. Each 25 µl reaction contained 12.5 µl 2x iQ Supermix (Biorad), 40 nM primer and 40 nM probe. RNaseP was used as house-keeping gene control (TaqMan RNaseP control reagent, Applied Biosystems, the Netherlands). All samples were run in quadruplet and subjected to 3 min at 95°C, 50 cycles of 95°C for 10s and 55°C for 30 s. Data were analyzed with iQ5 Optical System software (BioRad, Nazareth, Belgium).

4.3.1.11 Statistical Analysis

The difference in allele frequency distribution between the HIV-1-positive and HIV-1-negative group was determined by Fisher's exact test (FET) for each SNP to test the null hypothesis that allele frequencies were the same in the two groups. The effect of each SNP on HIV-1

viral load and $CD4^+$ T cell count was determined using a Generalized Estimating Equation (GEE) model [277] taking into account longitudinal measures for each participant. Viral loads were log-transformed and the square root of $CD4^+$ T cell count was used to normalize their measurements.

Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effect of each SNP on time to HIV-1 acquisition after enrollment and on the rate of progression to AIDS defined as $CD4^+$ T cell decline to less than 350 cells/μl (CD4<350). Decline in CD4 levels was determined and compared for the group with one or two copies of the minor allele to a reference group with two copies of the major allele (dominant genetic model), for each SNP. The significance of genotypic associations and relative hazard (RH) was determined by unadjusted Cox regression analysis for the dominant genetic model.

LEDGF/p75 mRNA expression levels were compared between SNs and SPs by performing dot plot graphical representation, nonparametric statistical analysis, and correlation (Pearson). Values were expressed as median values. Differences between the 2 groups were evaluated using Dunn's multiple comparison test, whereas the Mann-Whitney *U* test was used for any 2-group comparisons. The software used for the analysis was SAS version 9.1.3 (SAS Institute Inc., Cary, NC).

4.3.2 Results

4.3.2.1 Study design and selection of patient samples

Two hundred and forty seven samples (52 SPs and 195 SNs) from the CAPRISA AI 002 cohort and 403 samples from the Sinikithemba cohort were available for genotyping analysis. 32 samples from the CAPRISA AI cohort (10 SPs and 22 SNs) and 37 SPs from the Sinikithemba cohort were excluded from all analysis due to sample unavailability or poor quality genotype data.

4.3.2.2 Identification of Q472L as an exonic polymorphism in the C-terminus of LEDGF/p75

The *PSIP1* gene is approximately 47kb in length; consisting of fifteen exons (fig. 4.3.2.3.1). Therefore LEDGF/p75-specific primers were used to re-sequence a fragment (1.679kb long) of the C-terminus region of *PSIP1* in 43 SPs and 83 SNPs individuals from the CAPRISA AI 002 cohort (fig. 4.3.2.2). Resequencing of this fragment which included the IBD was conducted in order to screen for novel SNPs within this region that might be associated with disease outcome.

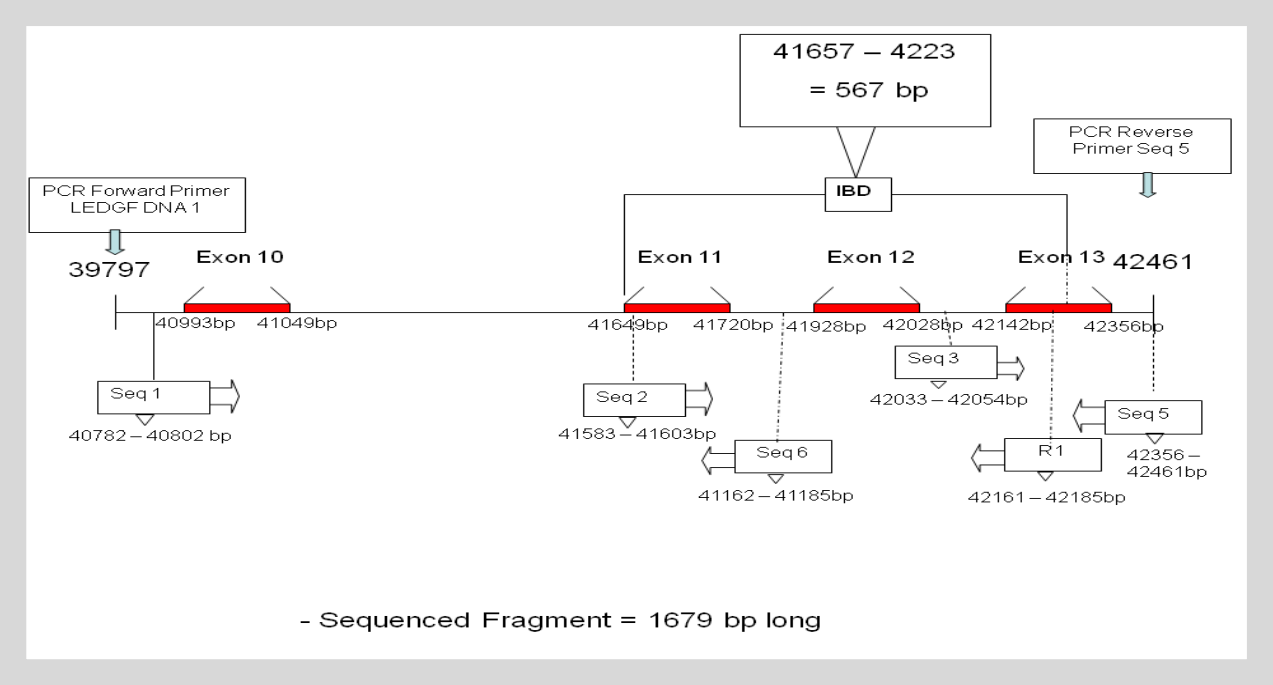

Figure 4.3.2.2. C-terminus fragment of the *PSIP1* gene that was re-sequenced to identify novel polymorphism that might be associated with the attenuated HIV-1 infection.

Fourteen previously described SNPs (dbSNP, www.ncbi.nlm.nih.gov) were discovered in the *PSIP1* gene by resequencing part of the C-terminal region of the *PSIP1* gene. Only one SNP rs61744944; a nonsynonymous SNP which result in amino acid change Glutamine to Leucine at position 472 (Q472L) of the LEDGF/p75 protein, was located in the exon region (exon 13) and the remaining SNPs were in the intron regions. Fisher's exact test was used to determine the effect of each SNP on susceptibility to HIV-1 infection (Table 4.3.2.2).

Table 4.3.2.2. Fourteen SNPs identified by resequencing and their association with susceptibility to HIV-1 infection

Wild type allele	Position (Region)	SNP numbers	Frequency of mutant allele	Susceptibility to HIV-1 acquisition (p-value)
A	41148A/G (Intron)	rs112104292	5/126	0.18
A	41250A/C (Intron)	rs4741508	60/126	0.25
G	41582G/A (Intron)	rs909323	1/126	1.00
C	41841C/A (Intron)	rs73421380	3/126	1.00
C	41857G/A (intron)	rs80047054	2/126	1.00
G	41994C/A (Intron)	rs12684386	7/126	1.00
A	42433A/T (Intron)	rs2795123	7/126	2.21
A	42503A/G (Intron)	rs4741507	8/126	1.00
AAG	42566_42568_Del (Intron)	rs34331255	1/126	1.00
T	42732A/T (Intron)	rs34331255	1/126	0.33
C	42819C/T (Intron)	rs2795124	2/126	1.00
C	44950C/T (Intron)	rs1046388	3/126	0.55
TTTGG	45201_45205_Del (Intron)	rs4741507	1/126	1.00
A	42350A/T (Q472L) (exon)	rs61744944	23/126	0.04

The rs61744944T was significantly associated with increased susceptibility to HIV-1 infection (Table 4.3.2.2). Kaplan-Meier survival analysis presented a clear separation of curves stratified for the major allele (A) of rs61744944 (represented by a blue curve) and the rs61744944T (represented by a black doted curve) on susceptibility to HIV-1 infection in

black South Africans (RH = 2.8, 95% CI = 1.6 – 7.3; uncorrected p = 0.04, Cox model) (fig. 4.3.2.2.1).

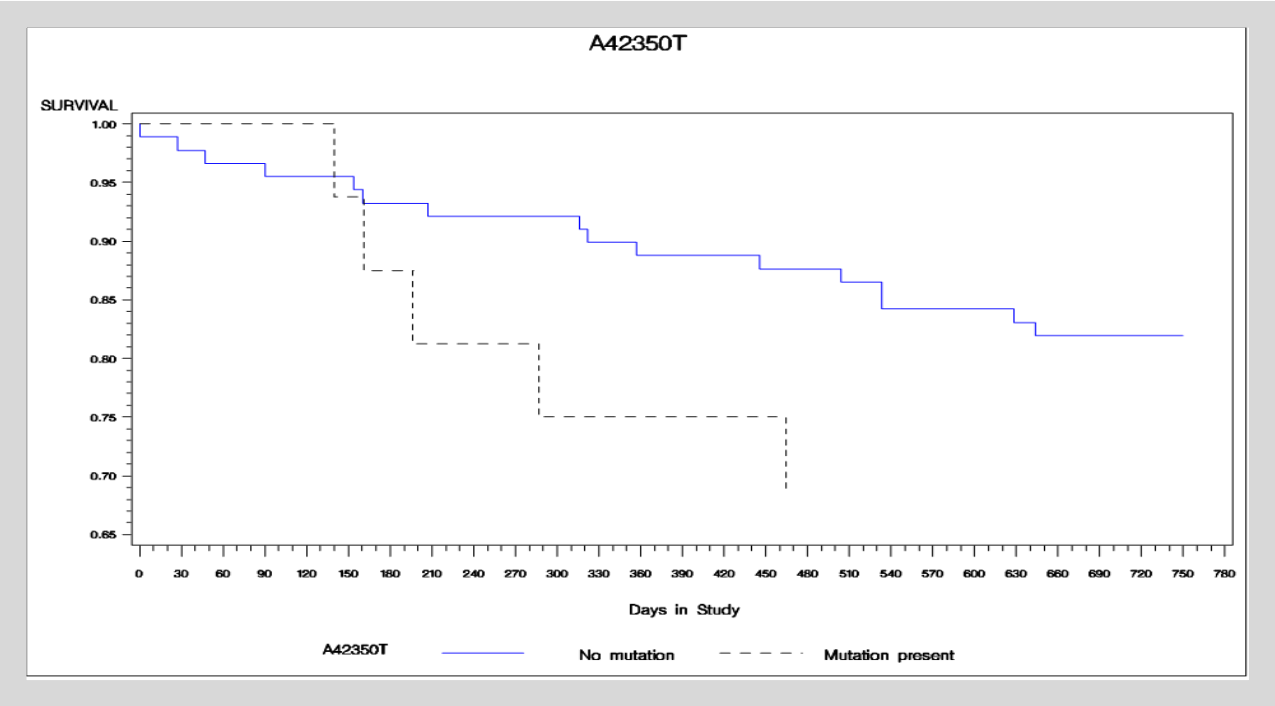

Figure 4.3.2.2.1. Kaplan-Meier survival curve analysis of A42350T (SNP5) on susceptibility to HIV-1 infection in black South Africans.

Subsequently, all 14 SNPs were tested for association with $CD4^+$ T cell counts and viral loads post HIV-1 infection. The rs61744944T was further found to be significantly associated with higher $CD4^+$ T cell count ($p < 0.01$) and lower viral loads ($p < 0.01$) over a period of 24 months (fig. 4.3.2.2.2).

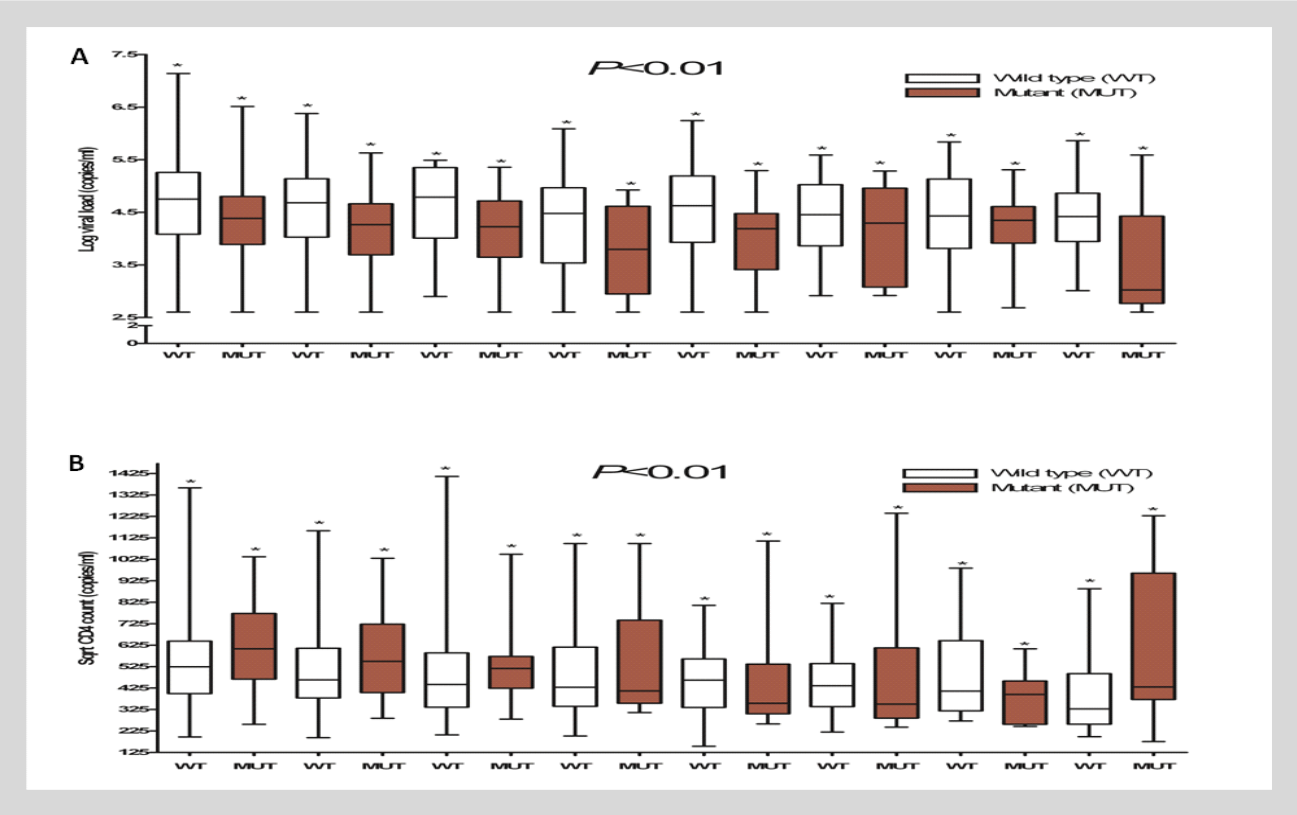

Figure 4.3.2.2.2. The influence of exonic SNP (rs61744944) on HIV-1 clinical outcomes. **A,** the association between viral loads and the SNP rs61744944; **B,** the association between $CD4^+$ T cell counts and the rs61744944 SNP. WT represent the wild type genotype (AA) and MUT represent mutant genotype (AT) of the exonic SNP.

The exonic SNP rs61744944 showed consistent and significant association with disease outcome. However, there was an unusual phenomenon with the results obtained from the preliminary analysis where the rs61744944T was associated with increased susceptibility to HIV-1 infection and yet higher $CD4^+$ T cell counts and lower viral loads post infection. This could also be attributed to the fact that pathophysiologic factors associated with foundational viral infection only partially overlap with those associated with downstream pathogenesis.

4.3.2.3 Effect of *PSIP1* SNPs on Susceptibility to HIV-1 Infection in the CAPRISA AI cohort.

The analysis of the preliminary data (section 4.3.2.2) suggested that genetic variation within the *PSIP1* gene might be associated with clinical outcomes in black South Africans. The analysis of *PSIP1* genetic variation was therefore extended to more participants (n=247) from the CAPRISA AI 002 cohort. 4 haplotype ht SNPs of the *PSIP1* gene and the exonic SNP rs61744944 (Q472L) were selected (see section 4.3.2.2) to genotype 247 participants from CAPRISA AI 002 cohort (fig. 4.3.2.3.1). The intronic ht SNPs were selected because they are in linkage disequilibrium with other SNPs within the *PSIP1* gene and therefore they would give the overall effect of genetic variation of the *PSIP1* gene on HIV-1 disease outcome.

CAPRISA AI 002 cohort comprises high risk individuals who were initially identified as HIV-1 negative and then followed longitudinally. This study design allowed us to test whether genetic variation in *PSIP1* was associated with susceptibility to HIV-1 infection. The minor allele frequencies (MAF) of SNPs and haplotype frequencies are shown at the bottom and right panel, respectively (fig. 4.3.2.3.1).

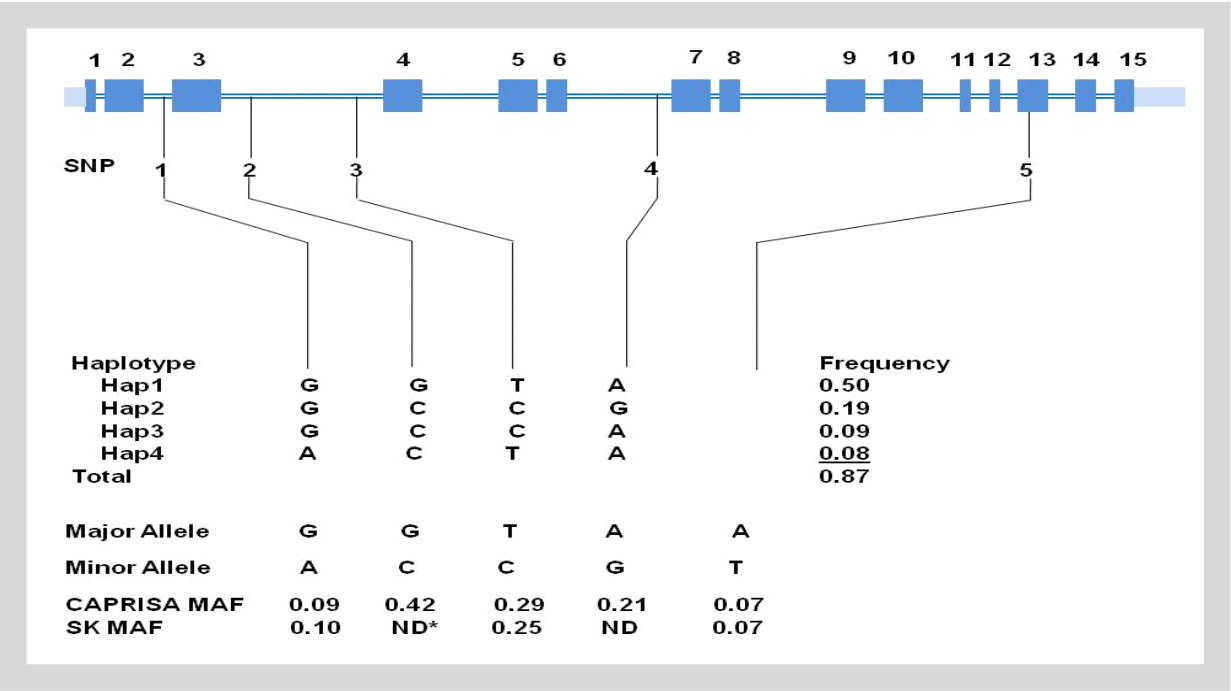

Figure 4.3.2.3.1. *PSIP1* gene: coding exons are marked with solid blue boxes. Minor allele frequencies (MAF) of the SNPs are shown for the CAPRISA and Sinikithemba (SK) cohorts at the bottom of panel. Hap is the acronym used for haplotype.

The minor allele (A) of rs2277191 (rs2277191A) was more frequent (25**%**) among the SPs in comparison to the SNs (p=0.06) (fig. 4.3.2.3.2A), suggesting that there could be an association between the rs2277191A and HIV status.

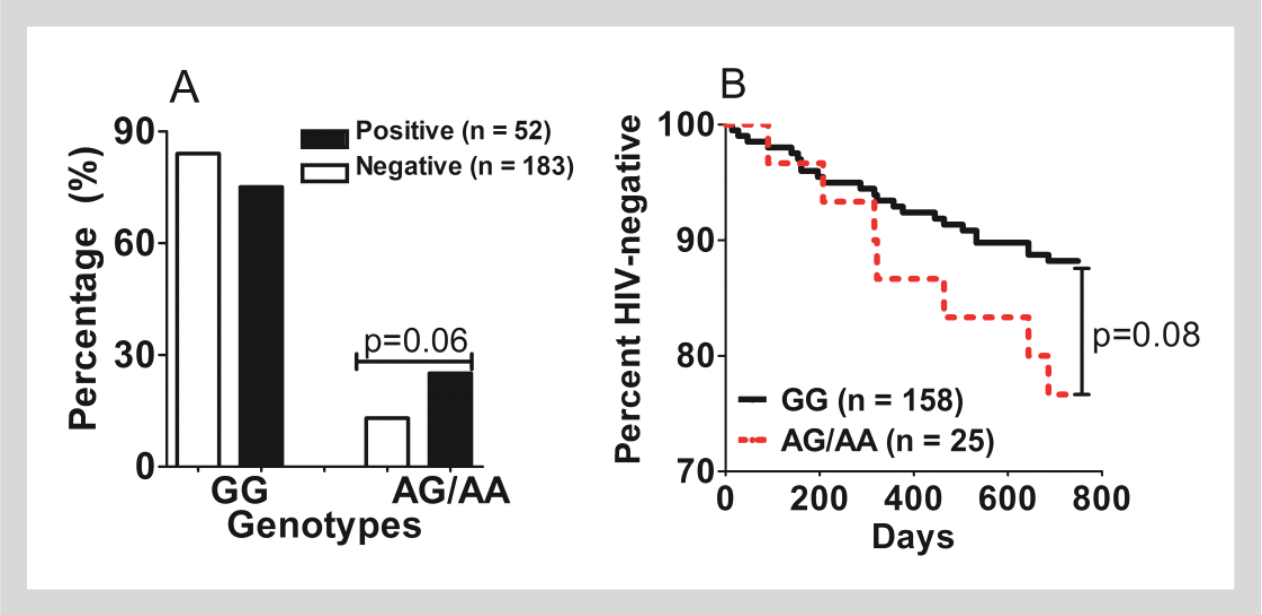

Figure 4.3.2.3.2. Association of the rs2277191A with susceptibility to HIV-1 infection. **A**, frequency distribution of the rs2277191A between HIV-1 negative and positive groups; **B**, Kaplan-Meier survival curve showing a clear separation of curves, GG wild type genotype (represented by black curve) and AG/GG mutant genotypes (represented by red curve) of the SNP rs2277191.

Kaplan-Meier survival analysis of time to HIV-1 acquisition suggested a trend of association between the rs2277191A and HIV-1 acquisition (RH = 2.21, 95% CI = 0.92–5.28; uncorrected p = 0.08, Cox model) (fig. 4.3.2.3.2B).

Table 4.3.2.3. Association of *PSIP1* SNPs with HIV-1 Acquisition in the CAPRISA AI 002 cohort, dominant model

SNP Risk group	N	Genotype Number (Frequency)			22 and 12 versus11 (reference group)*	
					RH (95% CI)	P - value
rs2277191 (SNP1)		GG	AG	AA		
SN	195	164(0.84)	26(0.13)	5 (0.03)		
SP	52	39 (0.75)	12(0.23)	1 (0.01)	2.21 (0.92 – 5.28)	0.08
rs10283923 (SNP2)		GG	CG	CC		
SN	195	80 (0.41)	77(0.39)	38(0.20)		
SP	52	20 (0.38)	23(0.44)	9 (0.17)	0.99 (0.44 – 2.24)	0.98
rs12339417 (SNP3)		TT	CT	CC		
SN	195	99 (0.51)	70(0.36)	26(0.13)		
SP	52	29 (0.56)	19(0.36)	4 (0.08)	0.60 (0.26 – 1.37)	0.23
rs1033056 (SNP4)		AA	AG	GG		
SN	195	121(0.62)	60(0.31)	14(0.07)		
SP	52	35 (0.67)	16(0.31)	1 (0.02)	0.85 (0.36 – 2.01)	0.71
rs61744944 (SNP5)		AA	AT	TT		
SN	195	168(0.86)	24(0.12)	3(0.02)		
SP	52	41 (0.79)	10(0.19)	1 (0.02)	1.90 (0.71 – 5.09)	0.20

Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effect of each SNP on time to HIV-1 acquisition. *P*-values uncorrected for multiple comparisons are shown. *11 represent the wildtype genotype, 12 represent heterozygous genotype and 22 represent homozygous genotype.

Although the rs61744944T was associated with increased susceptibility to HIV-1 infection in the preliminary data analysis of 126 participants, it lost its association when 247 participants

were analyzed. In this larger group of 247 participants only the rs2277191A showed a trend towards association with higher likelihood of acquiring HIV-1 infection (Table 4.3.2.3).

4.3.2.4 Effect of the *PSIP1* SNPs on $CD4^+$ T cell counts and viral loads in the CAPRISA AI 002 Cohort

The approximate time of infection was known for HIV-1 infected participants in the CAPRISA AI 002 cohort. Since $CD4^+$ T cell counts and viral loads can fluctuate significantly during HIV-1 primary infection, we analyzed the data in two specific intervals, 0-3 months post infection (acute phase of infection) and 3-12 months post infection (early chronic phase of infection) in order to detect possible differences between genotypes during these phases.

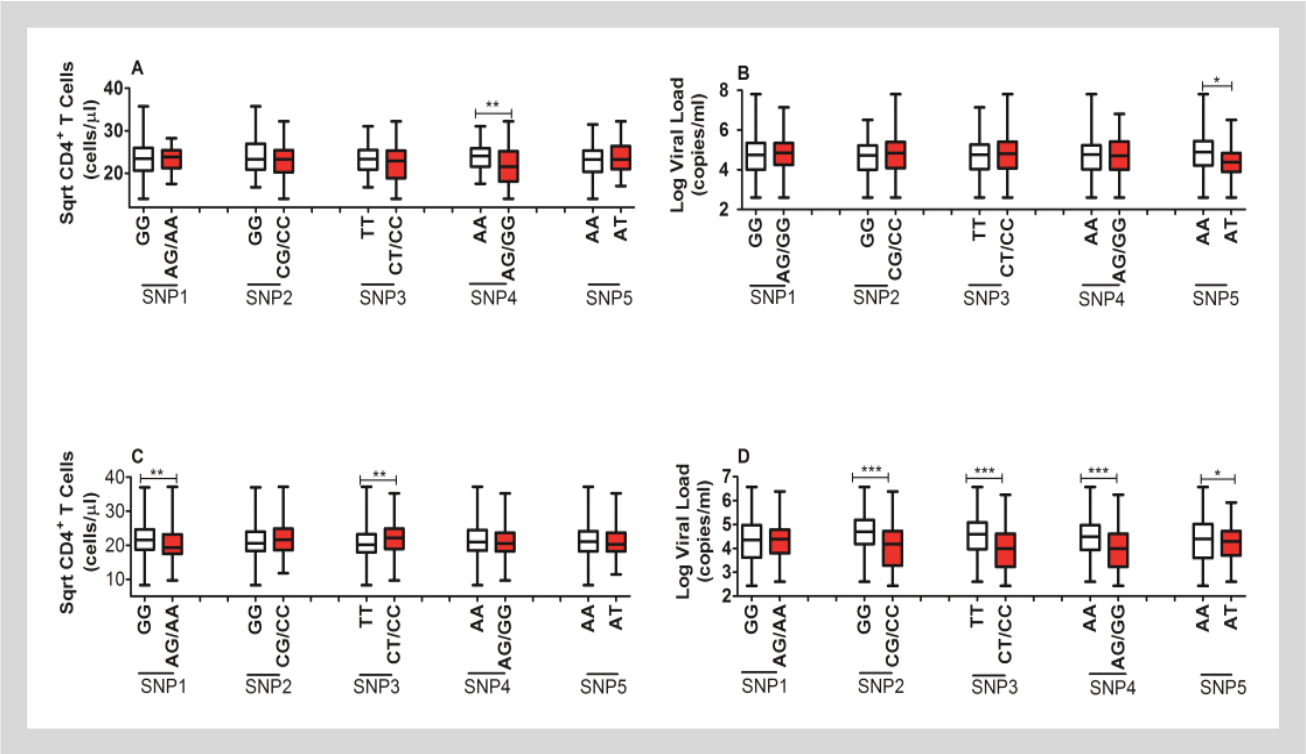

Figure 4.3.2.4. Association between the individual SNPs of the *PSIP1* gene and clinical outcomes reflected by square root $CD4^+$ T cells and log viral loads during acute phase of infection (**A** and **B**) and during early chronic phase of infection (**C** and **D**). SNP1, SNP2, SNP3, SNP4 and SNP5 refer to rs2277191, rs10283923, rs12339417, rs1033056 and rs61744944, respectively.

After correction for multiple comparison, dominant model analysis using GEE showed that the minor allele (A) of rs1033056 and the rs61744944T were significantly associated with lower $CD4^+$ T cells ($p < 0.01$) and viral load ($p = 0.01$), respectively, during the acute phase of infection (fig. 4.3.2.4A and B). The rs2277191A was associated with lower $CD4^+$ T cell count ($p < 0.01$) while the minor allele (C) of rs12339417 (rs12339417C) was associated with

higher $CD4^+$ T cell counts ($p < 0.01$) during early chronic phase of infection (fig. 4.3.2.4C). The minor alleles of all SNPs except rs61744944 were significantly associated with lower viral loads ($p < 0.01$) during the early chronic phase of infection (fig. 4.3.2.4D).

4.3.2.5 Effect of *PSIP1* SNPs and Haplotypes on $CD4^+$ T cell decline in the CAPRISA AI 002 Cohort

Few studies have analyzed the association of genetic factors with the rate of CD4 decline particularly in an African HIV-1 cohort with known time of infection.

Table 4.3.2.5. Effects of *PSIP1* SNPs on disease progression in the CAPRISA AI 002 cohort, dominant model

Endpoint	SNP Number	SNPs on AIDS Progression			Haplotype Number	Haplotypes on AIDS Progression		
		RH	95% CI	P		RH	95% CI	P
CD4 < 350	**SNP1**	5.98	2.27 – 5.82	0.04	**HAP1**	1.84	0.55- 6.12	0.77
CD4 < 350	**SNP2**	2.68	0.99 – 7.24	0.70	**HAP2**	0.84	0.19-3.77	0.21
CD4 < 350	**SNP3**	1.25	0.44 - 3.56	0.29	**HAP3**	0.00	0.00-0.00	0.99
CD4 < 350	**SNP4**	2.05	0.70 - 6.02	0.90	**HAP4**	5.41	1.92 -5.24	0.09
CD5 < 350	**SNP5**	1.74	0.50 – 6.05	0.71	**HAP5**	1.58	0.36 -7.49	0.66

Cox proportional hazards model (Cox model) was used to calculate the rate of CD4 decline to less than 350 cells/ml. *P*-values uncorrected for multiple comparisons are shown. SNP1, SNP2, SNP3, SNP4 and SNP5 refer to rs2277191, rs10283923, rs12339417, rs1033056 and rs61744944, respectively.

We therefore tested the effect of the *PSIP1* SNPs on the rate of $CD4^+$ T cell decline. The rs2277191A showed a weak but significant association with rapid $CD4^+$ T cell decline (p = 0.04, Cox model) (Table 4.3.2.5). Haplotype 4 (HAP4), the only haplotype carrying the rs2277191A, showed a trend of association with rapid $CD4^+$ T cell decline to CD4<350 (RH = 5.41, 95% CI =1.92-15.24, p = 0.09) (Table 4.3.2.5). None of the other SNP genotypes showed differences in the rate of $CD4^+$ T cell decline.

4.3.2.6 Effect of *PSIP1* SNPs in the Sinikithemba Cohort with Chronic HIV Infection

The results from the CAPRISA AI 002 cohort suggested that genetic variation in *PSIP1* gene may be associated with differential disease outcome. However, interpretation of these data was complicated by the small sample size (52 SPs) of the study cohort and the fluctuations in $CD4^+$ T cell counts and viral loads experienced during primary infection. We therefore extended our analysis to a larger cohort (Sinikithemba cohort), from within the same geographical area, in order to elucidate the role of the *PSIP1* genetic variants in HIV-1 pathogenesis.

Two SNPs, rs10283923 and rs1033056, were excluded from this analysis because they showed no association with $CD4^+$ T cell counts during early chronic phase of infection in the CAPRISA AI 002 cohort. The rs61744944 was included in this analysis based on the

preliminary data where it showed consistent association with both $CD4^+$ T cell counts and viral loads over 24 months post infection. The Sinikithemba cohort is a seroprevalent cohort with unknown dates of infection. Therefore, the effects of the SNPs rs2277191, rs12339417 and rs61744944 were examined on the trajectory of $CD4^+$ T cell decline over the 6 years of follow-up.

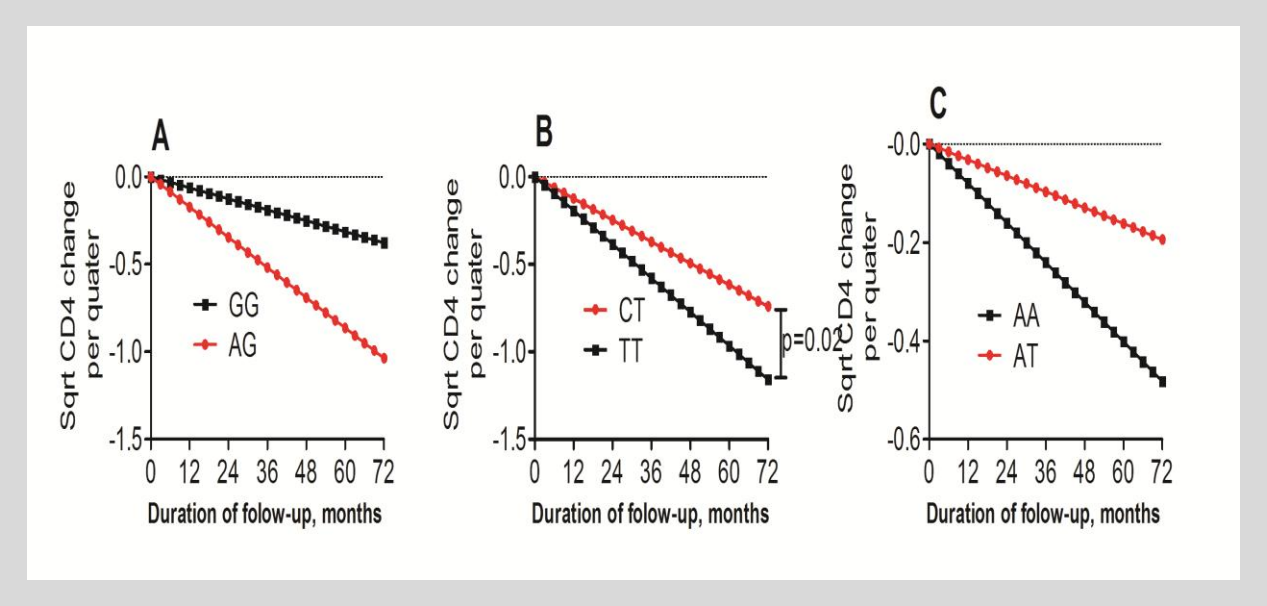

Figure 4.3.2.6. Rates of $CD4^+$ T cells declinestratified by genotypes for **A**- rs2277191, **B**- rs12339417, and **C**- 61744944 for the Sinikithemba seroprevalent cohort.

The influence of each of the three SNPs on disease progression was assessed by comparing the slopes of $CD4^+$ T cell levels for each of the three SNPs using the dominant model. The SNP rs2277191 did not show any association with disease progression in the Sinikithemba cohort. Interestingly, the rs12339417C was associated with delayed disease progression as was reflected by a slower rate of $CD4^+$ T cells decline to CD4<350 (p = 0.02, Cox model)

(Figure 4.3.2.6B). The SNP rs61744944 was also not associated with HIV-1 disease outcome in the Sinikithemba cohort.

4.3.2.7 LEDGF/p75mRNA expression levels

Next we studied the association between the *PSIP1* genetic variation and expression mRNA levels of LEDGF/p75. To this end, we performed quantitative RT-PCR on LEDGF/p75 mRNA levels in 57 HIV-1 negative and 38 recently infected participants from the CAPRISA AI 002 cohort. For 13 of the recently infected participants, samples pre- (at study entry) and post-HIV-1 infection were available for analysis.

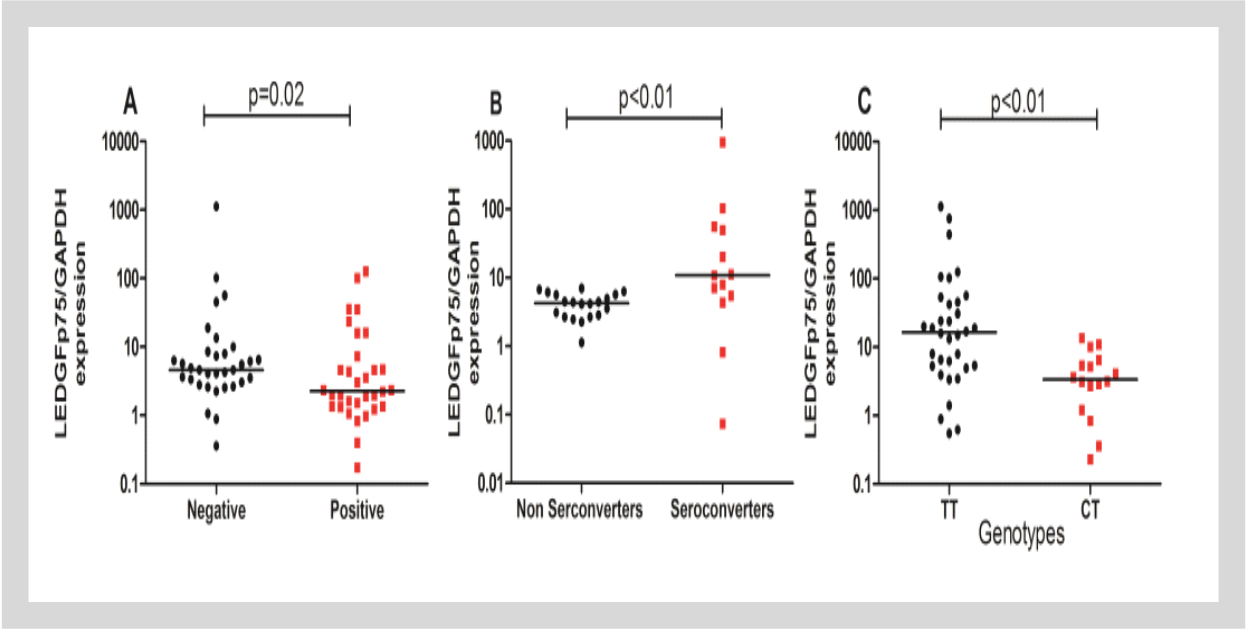

Figure 4.3.2.7. A, Expression of LEDGF/p75 mRNA in PBMCs obtained from HIV-1 positive versus negative participants. **B**, Expression of LEDGF/p75 mRNA in PBMCs obtained seroconverters versus nonseroconverters. **C**, Expression of LEDGF/p75 mRNA as

modulated by different genotypes of the SNP rs12339417. The TT→wild type genotype of SNP rs12339417, CT→heterozygous genotype of SNP rs12339417, the horizontal line in the middle of the points denotes the median value.

LEDGF/p75 mRNA expression data indicate that HIV-1 negative group express higher LEDGF/p75 mRNA levels compared to the HIV-1 positive group ($p = 0.02$) (fig. 4.3.2.7A). In order to ascertain whether LEDGF/p75 expression levels are dysregulated by HIV-1, 13 samples matched (pre- and post-HIV-1) samples were analyzed. There was higher LEDGF/p75 mRNA expression levels in PBMC at baseline (preinfection) from seroconverters as compared to PBMC obtained from non-seroconverters ($p < 0.01$) (fig. 4.3.2.7B). These results suggest that higher levels of LEDGF/p75 mRNA expression may be associated with a higher likelihood of acquiring HIV-1 infection and they also indicate that LEDGF/p75 mRNA expression levels are dysregulated by HIV-1 infection.

Lastly, we studied the association between *PSIP1* genetic variation and differential expression of LEDGF/p75 mRNA in the seronegative group. Results obtained from this analysis indicate that the differential expression of LEDGF/p75 mRNA was influenced by the SNP rs12339417. The rs12339417C was associated with reduced levels of LEDGF/p75 mRNA expression ($p < 0.01$) (fig. 4.3.2.7C). Interestingly, the expression data is consistent with the SNP rs12339417 genotype data where the rs12339417C was associated with higher $CD4^+$ T cell counts, lower viral loads and slower rate of $CD4^+$ T cell decline (section 4.3.2.4 and section 4.3.2.6).

4.3.2.8 Functional analysis of Q472L LEDGF/p75

To determine the impact of the Q472L mutation on the interaction of LEDGF/p75 with IN we measured its affinity in an AlphaScreen assay (Table 4.3.2.8) and evaluated the effect of the Q472L mutation on the interaction with JPO2 and pogZ, two cellular binding partners of LEDGF/p75 [258, 259]. While the affinity for JPO2 and pogZ was identical for the wild-type and the mutant protein, an almost 2-fold decrease in the K_d value for the Q472L mutant was observed (Table 4.3.2.8), suggesting stronger binding of the latter to HIV-1 IN.

Table 4.3.2.8. Interaction of WT and Q472L LEDGF/p75 with HIV-1 integrase

Protein	**Protein-protein interaction (K_d, nM)** [1]	
	HIV-1 IN	**JPO2pogZ**
LEDGF/p75 WT	47.7±4.1	79.3±13.6122.2±16.0
LEDGF/p75 Q472L	28.4±5.2	68.6±13.5 128.7±12.3

[1] K_d for interaction of LEDGF/p75 WT or Q472L with indicated recombinant proteins as measured by AlphaScreen. Average values ± SD for two independent measurements in triplicate.

Next, we evaluated the Q472L mutant in cell culture. Potent RNAi-mediated knockdown of LEDGF/p75 severely hampers HIV replication affecting the integration step of the provirus

[188, 296], a phenotype that is readily rescued upon re-introduction of RNAi resistant LEDGF/p75 (LEDGF BC) [204].

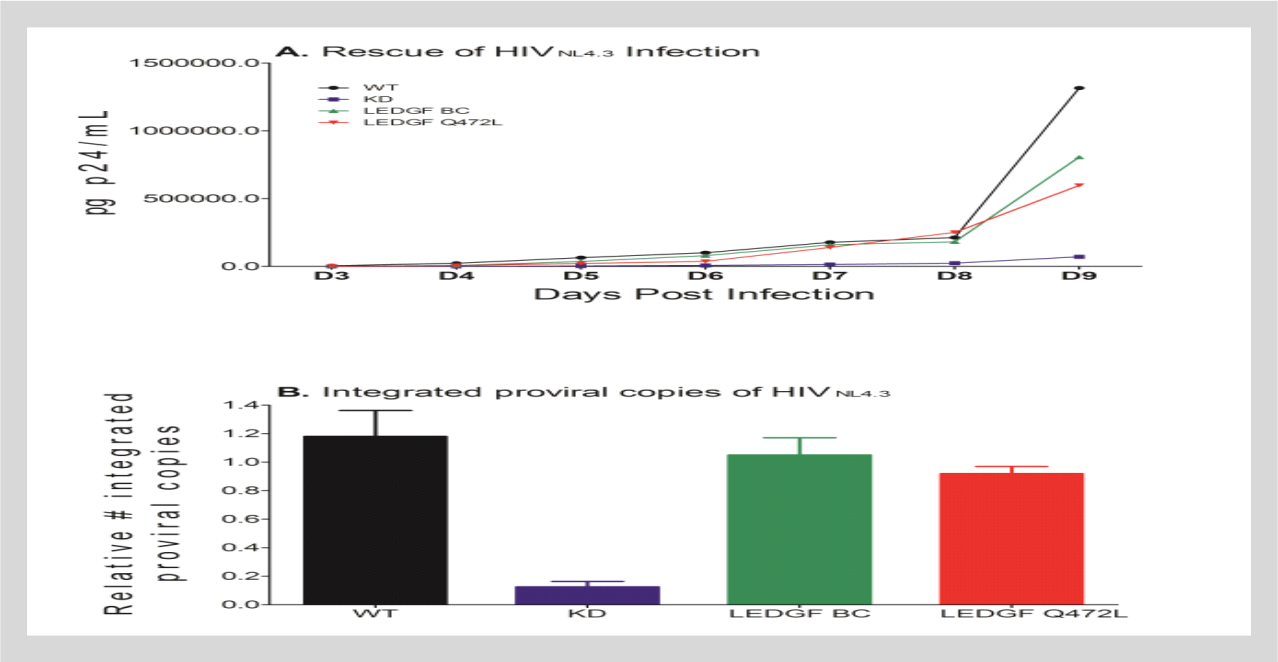

Figure 4.3.2.8. Rescue of spreading HIV NL4.3 infection. The Q472L LEDGF/p75 mutation was tested for its ability to rescue spreading $HIV_{NL4.3}$ infection in LEDGF/p75 knockdown cells [173]. HeLaP4/CCR5 cells (WT) and knockdown cells complemented with WT LEDGF/p75 BC were used as controls. **A**, Cells were infected in quadruplet with $HIV_{NL4.3}$ at an MOI of 0.01. HIV replication was monitored by sampling the supernatant for p24 at regular time-points post infection. A representative HIV-1 breakthrough experiment is shown; **B**, $HIV_{NL4.3}$ integration was measured by real-time quantitative PCR in the same experiment as in **A**, except that cells were treated with azidothymidine /ritonavir at day 7 to block further HIV replication.

In a similar setting, we complemented potent knockdown cells (KD) with the Q472L variant (LEDGF BC Q472L). LEDGF BC and the LEDGF BC Q472L supported HIV-1 replication to the levels observed for wild-type cells (WT) (fig. 4.3.2.8A). In parallel, we quantified integrated proviral copies. The nine-fold decrease in integrated proviral copies observed in KD cells was rescued to near wild-type levels in cells complemented with either LEDGF BC or the Q472L mutant (fig. 4.3.2.8B).

4.3.3 Discussion and Conclusions

LEDGF/p75 has been shown to promote HIV-1 replication *in vitro* [160, 171, 186-188, 201, 202, 290, 291]. In this study, we have investigated the association between genetic variation in the *PSIP1* gene and susceptibility to HIV-1 infection and disease progression in two South African HIV-1 cohorts. We found that, the rs2277191A, was associated with higher likelihood of HIV-1 acquisition. The rs2277191A was also associated with lower levels of $CD4^+$ T cell counts during the early chronic phase of infection and rapid $CD4^+$ T cell depletion, reflecting an early stage of disease progression in CAPRISA AI 002 cohort. However, rs10283923 deviated from the frequencies expected under the Hardy Weinberg equilibrium. Genotyping was repeated to eliminate genotyping error as a reason for rs10283923 not to conform to the Hardy Weinberg expectations and interestingly, genotypes obtained were consistent between duplicates and free of contamination as the negative controls did not amplify.

The rs12339417C was associated with higher $CD4^+$ T cell count and lower viral load during the early chronic phase of infection in the CAPRISA AI 002 cohort. We further found that the

SNP rs12339417 had an influence on LEDGF/p75 mRNA expression levels. The rs12339417C was associated with lower levels of LEDGF/p75 mRNA expression. These findings suggest that the variation at the SNP rs12339417 locus inhibits HIV-1 replication.

The interpretation of the CAPRISA AI 002 data was complicated by the small sample size of the study cohort. We therefore extended our analysis to a larger cohort, the Sinikithemba cohort, within the same geographical area, in order to get a clearer sense of the role of *PSIP1* genetic variants in HIV-1 pathogenesis. We excluded rs10283923 and rs1033056 from the analysis in the Sinikithemba cohort because these two SNPs showed no association with $CD4^+$ T cell count during primary infection and rs10283923 did not conform to the frequencies expected under HWE in the CAPRISA AI 002 cohort. Three SNPs (rs2277191, rs12339417 and rs61744944) were therefore selected for analysis in the Sinikithemba cohort. Sinikithemba cohort is comprised of chronically HIV-1 infected individuals whose date of infection is unknown and this made it impossible to study the effect of *PSIP1* genetic variation on susceptibility to HIV-1 infection.

Polymorphisms in the *PSIP1* gene were not associated with either $CD4^+$ T cell count or viral loads at baseline, which could be explained by the fact that individuals in this cohort were at different stages of HIV-1 infection upon enrolment into the study and immune activation [298, 299]. The rs12339417C was significantly associated with delayed disease progression as was reflected by a slower rate of $CD4^+$ T cell decline in the Sinikithemba cohort. The protective effect of the rs12339417C was consistent between the CAPRISA AI 002 and the

Sinikithemba cohort suggesting that a nucleotide change at the SNP rs12339417 locus maybe associated with inhibition of HIV-1 replication.

Lastly, we investigated the association between the *PSIP1* genetic variation and mRNA expression levels of LEDGF/p75 and found that rs12339417 influenced the expression of LEDGF/p75 with the rs12339417C significantly associated with lower levels of LEDGF/p75 mRNA. Seroconverters had significantly higher LEDGF/p75 mRNA levels compared to nonseroconverters at baseline. However, HIV-1 infected individuals expressed significantly lower levels of LEDGF/p75 mRNA compared to uninfected individuals. These findings suggested that high levels of LEDGF/p75 mRNA might be associated with a higher likelihood of HIV-1 acquisition and this corroborates the role of LEDGF/p75 as a cellular cofactor of HIV-1 infection in patients.

Previous studies on human genes APOBEC3G [268] and TRIM5α (TRIM5αhu) [276] showed that the mRNA of these two genes is down-regulated in HIV-1 infected PBMCs, compared with uninfected PBMCs. Both APOBEC3G and TRIM5αhu are intrinsic antiviral factors. Here we extend these studies to a cellular cofactor of HIV-1 replication. We investigated whether HIV-1 infection was associated with dysregulation of LEDGF/p75 expression. However, no significant difference was observed in the median level of LEDGF/p75 expression between matched preinfection and post-infection samples.

Although median levels of LEDGF/p75 mRNA did not reach statistical significance, they varied relatively throughout primary HIV-1 infection showing reduced levels of LEDGF/p75 mRNA post infection. We therefore examined the kinetics of LEDGF/p75 mRNA expression in matched HIV-1 - negative and – positive samples obtained from 13 individuals who acquired HIV-1 infection during follow-up.

PBMCs of these participants were available at baseline (before infection), within 3 months after infection (during the acute phase), and between 3 and 12 months after infection (during early chronic phase). The LEDGF/p75 mRNA expression levels were not significantly different among the different time points although there was an overall trend of decreased levels of LEDGF/p75 after infection. However, LEDGF/p75 mRNA expression levels fluctuated at the individual participant level, with 10 individuals showing a decrease and three showing an increase in LEDGF/p75 expression level after infection.

Our findings demonstrate, for the first time, that genetic variation of *PSIP1* may influence susceptibility to HIV-1 infection and the disease progression, which provide *in vivo* evidence that LEDGF/p75 is an important host cofactor for HIV-1 replication. Participants who expressed high levels of LEDGF/p75 mRNA at baseline were most likely to become HIV-1-infected.

Differential expression of LEDGF/p75 was associated with the SNP rs12339417 where the rs12339417C was showing significantly reduced mRNA levels of LEDGF/p75. This rs12339417C showed a slower rate of $CD4^+$ T decline in the Sinikithemba cohort suggesting that HIV-1 replication is inhibited by lower levels of LEDGF/p75 in a cell, which confirms the previous *in vitro* study [201]. These results suggest that repressing LEDGF/p75 mRNA levels could be used as a novel strategy to control HIV-1 replication.

CHAPTER FIVE

5.1 Discussion

The role of CypA, TRN-SR2 and LEDGF/p75 in promoting HIV-1 replication has been well established through extensive *in vitro* experiments [32, 134, 135, 160, 161, 171, 187, 201, 202, 273]. In addition to *in vitro* studies, the role of CypA has been studied *in vivo* using populations from the developed world [61, 73, 74]. In this study, we investigated the association of genetic variation in the *PPIA*, *TNPO3* and *PSIP1* genes with susceptibility to HIV-1 infection and disease progression in South African cohorts.

The minor allele (G) of SNP A1650G (1650G) in the promoter region of *PPIA* was significantly associated with lower $CD4^+$ T cell count and higher viral loads during HIV-1 primary infection. These findings suggest that the 1650G was associated with enhanced HIV-1 replication as reflected by significantly higher viral loads and lower $CD4^+$ T cell counts during primary infection in the CAPRISA AI 002 cohort. These findings are consistent with the data obtained from Swiss Caucasian HIV-1 cohort, where the 1650G was reported to be associated with a rapid $CD4^+$ T cell depletion [61]. A trend towards association between 1650G and rapid $CD4^+$ T cell depletion was reported among African American [73]. Therefore, in this study we tested the effect of the 1650G on clinical outcomes upon HIV-1 exposure in black South Africans (CAPRISA AI 002 cohort) and our results showed a borderline association between rapid $CD4^+$ T cell depletion and the 1650G. Lack of significant results in our study could be attributed to a small sample size with only 24 HIV-1 positive participants who had the 1650G in the CAPRISA AI cohort.

The 1650G in the promoter region of the *PPIA* gene was significantly associated with elevated levels of CypA mRNA expression post HIV-1 infection, suggesting that the 1650G

promotes HIV-1 replication *in vivo* and rapid disease progression by inducing higher expression of CypA mRNA levels. Pre-HIV-1 infection levels of CypA mRNA expression were significantly higher in PBMC genotyped for the 1650G compared to PBMC harboring wild type genotyped (AA) at the SNP locus. This could suggest that 1650G might be indirectly associated with susceptibility to HIV-1 infection. Results obtained from the *ex vivo* experiment in this study suggested that CypA mRNA expression might be dysregulated by HIV-1 infection. This was evident from the observation that there was a sudden drop in CypA mRNA expression levels upon HIV-1 infection and CypA mRNA levels were restored to significantly higher levels after a week post HIV-1 infection. The 1650G might be directly involved in promoting proper uncoating of HIV-1 capsid to make RNA genome available for reverse transcription to take place in order to synthesize viral copy DNA (cDNA).

Following cDNA synthesis, viral cDNA has to be integrated into the host chromosome which resides within the nucleus which is separated from the cytoplasm by a double lipid bilayer. Therefore HIV-1 relies on import proteins such as TRN-SR2 for nuclear import [160]. Numerous *in vitro* studies have shown that TRN-SR2 is an important HIV-1 cofactor [32, 67, 160, 161], however, this has not been shown *in vivo*. In this study, we undertook a systematic investigation of the association of SNPs in the *TNPO3* gene with susceptibility to HIV-1 infection and disease progression in two South African HIV-1 cohorts.

The minor allele (G) of rs1154330 (rs1154330G) in intron 12 of the *TNPO3* gene was associated with faster acquisition of HIV-1 infection, higher viral load and lower $CD4^+$ T cell count in the CAPRISA AI 002 cohort. However, rs1154330G was not associated with disease

progression in the CAPRISA AI cohort probably due to a small number of HIV-1 positive participants (13 out of 52 participants) who carried rs1154330G. Therefore the analysis of the rs1154330G was extended to the Sinikithemba cohort comprising 450 HIV-1 chronically infected participants from the same geographical area. Since the Sinikithemba cohort is a seroprevalent cohort with unknown dates of infection, the influence of the rs1154330G could only be assessed on $CD4^+$ T cell levels and viral loads at study enrollment (baseline) and on the trajectory of $CD4^+$ T cell decline over the 5 years of follow-up

Interestingly, the rs1154330G was significantly associated with a faster rate of $CD4^+$ T cell decline in the Sinikithemba cohort. The results obtained from the Sinikithemba cohort were consistent with the data obtained from the CAPRISA AI 002 cohort suggesting that the rs1154330G gives HIV-1 selective advantage for efficient replication and productive infection in the South African population. The effect of the rs1154330G could not be assessed on the protein functionality of TNR-SR2 protein because this SNP is located in the intron region of the *TNPO*3 gene. It is therefore presumed that the rs1154330G is associated with TRN-SR2 mRNA expression levels, regulation of possible splice variants or post-translational protein modifications. At baseline, in the Sinikithemba cohort, the rs1154330G was associated with neither $CD4^+$ T cell count nor viral loads. This lack of association could be attributed to the fact that participants in the Sinikithemba cohort were recruited while they were at different stages of HIV-1 infection.

In addition to the SNP rs1154330, there were other SNPs in the *TNPO3* gene that were associated with clinical outcomes in the CAPRISA AI 002 cohort. However, these SNPs did

not show consistent association with clinical outcomes during primary infection. The minor allele (A) of rs2305325 (rs2305325A) was associated with higher $CD4^+$ T cell counts but not with viral loads during HIV-1 primary infection. The association between the rs2305325A and higher $CD4^+$ T cell counts may probably be due to alterations in the cytokine milieu such as variation in IL-7 effect since the levels of $CD4^+$ T cell counts were similar between participants carrying the wild type genotype and mutant genotype prior to HIV-1 infection. On the other hand, the minor allele (A) of the SNP rs35060568 (rs35060568A) was associated with higher viral load during the primary infection. However, the rs35060568A was only associated with lower $CD4^+$ T cell counts during the early chronic phase of infection.

The minor allele (C) of rs6957529 (rs6957529C) showed inconsistent association with both viral loads during primary infection, where it was associated with lower viral load during acute phase of infection and higher viral load during the early chronic phase of infection. The plausible explanation for this could be that the virus developed the escape mutation in the later stages of infection in order to overcome the inhibiting effect of the rs6957529C. The minor allele (C) of rs10229001 was associated with higher $CD4^+$ T cell counts and lower viral load during the early chronic phase of infection only. To the best of our knowledge this study demonstrated for the first time that the genetic variants in the *TNPO3* gene are associated with disease outcome confirming that TRN-SR2 is indeed an important HIV-1 replication cofactor.

LEDGF/p75 has been shown to play a central role in promoting HIV-1 replication *in vitro* [171, 187, 189, 197, 201, 202, 291]. In this study, we investigated the association between genetic variation in the *PSIP1* gene and clinical outcomes. As the preliminary study, we screened part of the C-terminal region of the *PSIP*1 gene which included the IBD in order to identify SNPs that could attenuate HIV-1 infection. The preliminary study was done only in fewer individuals (n = 126) from CAPRISA AI 002 cohort. Sequencing revealed 14 previously described SNPs, 13 of which were allocated in the intron regions of the *PSIP*1 gene and one SNP rs61744944 was located in exon 13 outside the IBD of the *PSIP*1 gene though. The SNP rs61744944 is a non-synonymous SNP resulting in an amino acid change (Glutamine to Leucine) at position 472 of LEDGF/p75 (Q472L).

The preliminary analysis revealed an association between the minor allele (T) of the SNP rs61744944 (rs61744944T) and increased susceptibility to HIV-1 infection. Surprisingly, the rs61744944T was associated with consistently higher $CD4^+$ T cell counts and lower viral loads post HIV-1 infection. Consequently, the analysis of the rs61744944T was therefore extended to a larger number (n = 247) of participants from the same cohort in order to elucidate its role in HIV-1 infection. In addition to this SNP four additional intronic haplotype tagging (ht) SNPs were analyzed in these 247 participants from CAPRISA AI 002 cohort. It is necessary to note that only 52 of the 247 participants analyzed were HIV-1 positive.

Interestingly, the minor allele (A) of the ht SNP rs2277191 (rs2277191A) was found to be associated with higher likelihood of HIV-1 acquisition, lower $CD4^+$ T cell counts during the early chronic phase of infection and rapid $CD4^+$ T cell depletion in CAPRISA AI002 cohort.

Whereas the minor allele (C) of the other ht SNP rs12339417 (rs12339417C) was associated with higher $CD4^+$ T cell count and lower viral load during the early chronic phase of infection in the CAPRISA AI 002 cohort. Although the rs61744944T was not associated with higher $CD4^+$ T cell counts in these 52 HIV-1 positive participants, the rs61744944T was consistently associated with lower viral load during primary infection. Therefore SNPs rs2277191 and rs12339417 were further analyzed in a larger chronic infection Sinikithemba cohort based on their association with $CD4^+$ T cell counts during the early chronic phase of infection. The rs61744944 was included in this analysis based on the preliminary findings that suggested an association with clinical outcomes and based on the fact that this was the only exonic SNP in the *PSIP1* gene that was analyzed in this study.

The Sinikithemba cohort comprises chronically HIV-1 infected individuals with unknown date of infection and therefore we assessed the effect of these three SNPs on trajectory of $CD4^+$ T cell decline over 6 years of follow-up. The rs12339417C was associated with delayed disease progression as was reflected by significantly a slower rate of $CD4^+$ T cell decline in the Sinikithemba cohort. The protective effect of the rs12339417C was consistent between the CAPRISA AI 002 and Sinikithemba cohorts suggesting that the rs12339417C may be associated with reduced HIV-1 replication.

In an attempt to elucidate the mechanism that the rs12339417C used to inhibit HIV-1 replication, we investigated the association between the rs12339417C and LEDGF/p75 mRNA expression levels. The other two SNPs rs2277191 and rs61744944 were also included in this analysis. Interestingly, the rs12339417C was associated with lower expression of LEDGF/p75 mRNA levels suggesting that the rs12339417C inhibited HIV-1 replication through suppressing the mRNA levels of LEDGF/p75 *in vivo*. These findings are consistent with *in vitro* knockdown studies [67, 296]. The PBMCs obtained from seroconverters had higher mRNA expression levels of LEDGF/p75 compared to nonseroconverters suggested that high levels of LEDGF/p75 may increase the likelihood of HIV-1 acquisition and the rate of disease progression. This observation is consistent with findings that LEDGF/p75 is indeed an important HIV-1 replication cofactor [201]. The other two SNPs, rs2277191 and rs61744944 were neither associated with disease progression in the Sinikithemba cohort nor with mRNA expression levels of LEDGF/p75 in the CAPRISA AI 002 cohort.

Lastly, we analyzed the impact of the mutant LEDGF/p75 (Q472L) on LEDGF/p75-IN interaction and plausible effects on cellular binding factors of LEDGF/p75, JPO2 and pogZ. The mutation Q472L did not alter the binding affinity of LEDGF/p75 for IN, JPO2 and pogZ. The back complementation of LEDGF/p75-depleted cells with the Q472L rescued HIV-1 replication to near wild-type levels suggesting that Q472L did not alter the functionality of LEDGF/p75. Lack of association between Q472L and binding affinity of LEDGF/p75 for IN,

JPO2 and pogZ could be explained by the fact that Q472L occurs outside the IBD of LEDGF/p75. However, it is not inconceivable that this mutation may affect other LEDGF/p75 functions such as integration site targeting.

As discussed above, this study extended genetic association studies in HIV-1 research from immunological factors or proteins involved in the immune response to HIV-1 replication cofactors. The novel aspect of this study was this study was conducted in a population from Sub-Saharan Africa, South Africa in particular which is heavily burdened by HIV-1 epidemic.

It is interesting to note that these SNPs were not found to be important in HIV-1 pathogenesis in previous studies that used genomewide association study approach [300, 301]. The reasons for this could include ethnic differences in allele frequencies because they were studied in non-African populations. Other reasons could be cohort design and phenotype used, Pereyra *et al*. looked at elite controllers [301] and Petrovski *et al*. looked at viral set point in African Americans [302], respectively. The other plausible reason could be that the effect of genetic variation within HIV-1 replication cofactors on HIV pathogenesis is modest because these proteins are cellular house-keeping genes that tend to be evolutionarily conserved.

5.2 Conclusions

This study provides *in vivo* evidence that CypA, TRN-SR2 and LEDGF/p75 are crucial replication cofactors for efficient HIV-1 replication, which has been well established through extensive *in vitro* experiments [32, 134, 135, 160, 161, 171, 187, 201, 202, 273].

This study had two significant limitations: (1) there were few individuals with HIV-1 primary infection whose date of infection was known and; (2) The date of infection was not known for HIV-1 chronically infected individuals, and therefore the clinical data of these individuals could not to be categorized and analyzed according to the different phases of HIV-1 infection. The acute phase of infection is characterized by spike in viral load whereas, at viral set point, the viral load drops and at the chronic phase of infection, high viral loads resume. HIV-1 chronic infection cohort made it difficult to assess the influence of genetic variations at baseline (study entry) because participants in this study cohort were at different stages of the disease when they were recruited into this cohort.

Despite the challenges mentioned above, the results of this study suggest genetic variation in these select HIV-1 replication cofactors may be associated with susceptibility to HIV-1 infection and disease progression. Although the association between genetic variant A1650G of CypA and HIV-1 disease outcomes has been previously reported in populations from the developed countries [61, 73, 74], this is the first study to show this association in the population from developing world which is heavily burden by HIV-1 epidemic. Furthermore,

this is the first study to report the association of genetic polymorphisms in the *TNPO3* and *PSIP1* genes with susceptibility to HIV-1 infection and disease progression, confirming that TRN-SR2 and LEDGF/p75 are indeed important HIV-1 replication cofactors. This study suggests that targeting protein-protein interactions (PPIs) between HIV-1 and replication cofactors could be a new antiviral target against which small-molecule protein-protein interaction inhibitors (SMPPIIs) could be designed to inhibit HIV-1 infection. However, due to small sample size and heterogeneous nature of our cohorts, our findings should be interpreted with caution and will need to be replicated in additional studies.

Future studies should investigate the association of genetic variation in other replication cofactors with HIV-1 clinical outcomes. Future studies should also include more studies to understand mechanisms of viral interaction with HIV replication cofactors and design of therapeutic interventions targeting replication cofactors and whether the effectiveness of these new therapeutics will be influenced by population genetic variation of replication cofactors.

REFERENCES

1. CDC. Pneumocystis pnuemonia - Los Angels. *MMWR* 1981 July 4,**30**:306-308.

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

3. Buchacz K, Baker RK, Palella FJ, Jr., Chmiel JS, Lichtenstein KA, Novak RM*, et al.* AIDS-defining opportunistic illnesses in US patients, 1994-2007: a cohort study. *AIDS* 2010,**24**:1549-1559.

4. Cann AJ, Rosenblatt JD, Wachsman W, Shah NP, Chen IS. Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. *Nature* 1985,**318**:571-574.

5. Gouws E, Stanecki KA, Lyerla R, Ghys PD. The epidemiology of HIV infection among young people aged 15-24 years in southern Africa. *AIDS* 2008,**22 Suppl 4**:S5-16.

6. Urnovitz HB, Murphy WH. Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clin Microbiol Rev* 1996,**9**:72-99.

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

8. Geyer M, Fackler OT, Peterlin BM. Structure--function relationships in HIV-1 Nef. *EMBO Rep* 2001,**2**:580-585.

9. Wiegers K, Rutter G, Schubert U, Grattinger M, Krausslich HG. Cyclophilin A incorporation is not required for human immunodeficiency virus type 1 particle maturation and does not destabilize the mature capsid. *Virology* 1999,**257**:261-274.

10. Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* 2003,**12**:591-601.

11. Chiu YL, Soros VB, Kreisberg JF, Stopak K, Yonemoto W, Greene WC. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature* 2005,**435**:108-114.

12. Schrofelbauer B, Hakata Y, Landau NR. HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. *Proc Natl Acad Sci U S A* 2007,**104**:4130-4135.

13. Sherman MP, De Noronha CM, Williams SA, Greene WC. Insights into the biology of HIV-1 viral protein R. *DNA Cell Biol* 2002,**21**:679-688.

14. King JE, Eugenin EA, Buckner CM, Berman JW. HIV tat and neurotoxicity. *Microbes Infect* 2006,**8**:1347-1357.

15. Li JC, Yim HC, Lau AS. Role of HIV-1 Tat in AIDS pathogenesis: its effects on cytokine dysregulation and contributions to the pathogenesis of opportunistic infection. *AIDS* 2010,**24**:1609-1623.

16. Yim HC, Li JC, Lau JS, Lau AS. HIV-1 Tat dysregulation of lipopolysaccharide-induced cytokine responses: microbial interactions in HIV infection. *AIDS* 2009,**23**:1473-1484.

17. Huffman KM, Arrigo SJ, Schmidt MG. HIV-1 Rev promotes the nuclear export of unspliced and singly spliced RNAs in a mammalian cell-free export system. *J Biomed Sci* 1999,**6**:194-205.

18. Hout DR, Mulcahy ER, Pacyniak E, Gomez LM, Gomez ML, Stephens EB. Vpu: a multifunctional protein that enhances the pathogenesis of human immunodeficiency virus type 1. *Curr HIV Res* 2004,**2**:255-270.

19. Wainberg MA. HIV-1 subtype distribution and the problem of drug resistance. *AIDS* 2004,**18 Suppl 3**:S63-68.

20. Mansson F, Biague A, da Silva ZJ, Dias F, Nilsson LA, Andersson S*, et al.* Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. *AIDS* 2009,**23**:1575-1582.

21. Gadelha SR, Shindo N, Cruz JN, Morgado MG, Galvao-Castro B. Molecular epidemiology of human immunodeficiency virus-1 in the state of Ceara, Northeast, Brazil. *Mem Inst Oswaldo Cruz* 2003,**98**:461-463.

22. Papathanasopoulos MA, Hunt GM, Tiemessen CT. Evolution and diversity of HIV-1 in Africa--a review. *Virus Genes* 2003,**26**:151-163.

23. Lal RB, Chakrabarti S, Yang C. Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development. *Indian J Med Res* 2005,**121**:287-314.

24. Girard MP, Osmanov SK, Kieny MP. A review of vaccine research and development: the human immunodeficiency virus (HIV). *Vaccine* 2006,**24**:4062-4081.

25. Al-Mawsawi LQ, Neamati N. Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy. *Trends Pharmacol Sci* 2007,**28**:526-535.

26. McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, Emerman M*, et al.* Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol* 2002,**159**:441-452.

27. Nermut MV, Fassati A. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *J Virol* 2003,**77**:8196-8206.

28. Collin M, Gordon S. The kinetics of human immunodeficiency virus reverse transcription are slower in primary human macrophages than in a lymphoid cell line. *Virology* 1994,**200**:114-120.

29. Bukrinskaya A, Brichacek B, Mann A, Stevenson M. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *J Exp Med* 1998,**188**:2113-2125.

30. Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 1997,**71**:5382-5390.

31. Lim RY, Aebi U, Stoffler D. From the trap to the basket: getting to the bottom of the nuclear pore complex. *Chromosoma* 2006,**115**:15-26.

32. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X*, et al.* The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *J Virol* 2009,**84**:397-406.

33. Bukrinsky MI, Sharova N, Dempsey MP, Stanwick TL, Bukrinskaya AG, Haggerty S*, et al.* Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci U S A* 1992,**89**:6580-6584.

34. Fouchier RA, Malim MH. Nuclear import of human immunodeficiency virus type-1 preintegration complexes. *Adv Virus Res* 1999,**52**:275-299.

35. Gallay P, Hope T, Chin D, Trono D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci U S A* 1997,**94**:9825-9830.

36. Fauci AS, Desrosiers RC. Pathogenesis of HIV and SIV. In: *Retroviruses*. Edited by Coffin JM, Hughes SH, Varmus HE. Cold Springs Harbor; 1997. pp. 587 - 637.

37. Schuitemaker H, Koot M, Kootstra NA, Dercksen MW, de Goede RE, van Steenwijk RP*, et al.* Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. *J Virol* 1992,**66**:1354-1360.

38. Bashirova AA, Thomas R, Carrington M. HLA/KIR restraint of HIV: surviving the fittest. *Annu Rev Immunol* 2011,**29**:295-317.

39. Kumar V, Prakash O, Manpreet S, Sumedh G, Medhi B. Genetic basis of HIV-1 resistance and susceptibility: an approach to understand correlation between human genes and HIV-1 infection. *Indian J Exp Biol* 2006,**44**:683-692.

40. Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* 2009,**37**:D417-422.

41. Pinney JW, Dickerson JE, Fu W, Sanders-Beer BE, Ptak RG, Robertson DL. HIV-host interactions: a map of viral perturbation of the host system. *AIDS* 2009,**23**:549-554.

42. Ptak RG, Fu W, Sanders-Beer BE, Dickerson JE, Pinney JW, Robertson DL*, et al.* Cataloguing the HIV type 1 human protein interaction network. *AIDS Res Hum Retroviruses* 2008,**24**:1497-1502.

43. Lama J, Planelles V. Host factors influencing susceptibility to HIV infection and AIDS progression. *Retrovirology* 2007,**4**:52.

44. O'Brien SJ, Nelson GW. Human genes that limit AIDS. *Nat Genet* 2004,**36**:565-574.

45. Telenti A, Ioannidis JP. Susceptibility to HIV infection--disentangling host genetics and host behavior. *J Infect Dis* 2006,**193**:4-6.

46. Fellay J. Host genetics influences on HIV type-1 disease. *Antivir Ther* 2009,**14**:731-738.

47. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R*, et al.* 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* 1996,**273**:1856-1862.

48. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T*, et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* 1996,**2**:1240-1243.

49. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R*, et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996,**86**:367-377.

50. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM*, et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 1996,**382**:722-725.

51. Paxton WA, Liu R, Kang S, Wu L, Gingeras TR, Landau NR*, et al.* Reduced HIV-1 infectability of CD4+ lymphocytes from exposed-uninfected individuals: association with low expression of CCR5 and high production of beta-chemokines. *Virology* 1998,**244**:66-73.

52. Saha K, Bentsman G, Chess L, Volsky DJ. Endogenous production of beta-chemokines by CD4+, but not CD8+, T-cell clones correlates with the clinical state of human immunodeficiency virus type 1 (HIV-1)-infected individuals and may be responsible for blocking infection with non-syncytium-inducing HIV-1 in vitro. *J Virol* 1998,**72**:876-881.

53. An P, Nelson GW, Wang L, Donfield S, Goedert JJ, Phair J*, et al.* Modulating influence on HIV/AIDS by interacting RANTES gene variants. *Proc Natl Acad Sci U S A* 2002,**99**:10002-10007.

54. Duggal P, Winkler CA, An P, Yu XF, Farzadegan H, O'Brien SJ*, et al.* The effect of RANTES chemokine genetic variants on early HIV-1 plasma RNA among African American injection drug users. *J Acquir Immune Defic Syndr* 2005,**38**:584-589.

55. Gonzalez E, Dhanda R, Bamshad M, Mummidi S, Geevarghese R, Catano G*, et al.* Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic. *Proc Natl Acad Sci U S A* 2001,**98**:5199-5204.

56. Liu H, Chao D, Nakayama EE, Taguchi H, Goto M, Xin X*, et al.* Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proc Natl Acad Sci U S A* 1999,**96**:4581-4585.

57. McDermott DH, Beecroft MJ, Kleeberger CA, Al-Sharif FM, Ollier WE, Zimmerman PA*, et al.* Chemokine RANTES promoter polymorphism affects risk of both HIV infection and disease progression in the Multicenter AIDS Cohort Study. *AIDS* 2000,**14**:2671-2678.

58. Modi WS, Lautenberger J, An P, Scott K, Goedert JJ, Kirk GD*, et al.* Genetic variation in the CCL18-CCL3-CCL4 chemokine gene cluster influences HIV Type 1 transmission and AIDS disease progression. *Am J Hum Genet* 2006,**79**:120-128.

59. Dolan MJ, Kulkarni H, Camargo JF, He W, Smith A, Anaya JM*, et al.* CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. *Nat Immunol* 2007,**8**:1324-1336.

60. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G*, et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005,**307**:1434-1440.

61. Bleiber G, May M, Martinez R, Meylan P, Ott J, Beckmann JS*, et al.* Use of a combined ex vivo/in vivo population approach for screening of human genes involved in the human immunodeficiency virus type 1 life cycle for variants influencing disease progression. *J Virol* 2005,**79**:12674-12680.

62. Magierowska M, Theodorou I, Debre P, Sanson F, Autran B, Riviere Y*, et al.* Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 1999,**93**:936-941.

63. Mummidi S, Ahuja SS, Gonzalez E, Anderson SA, Santiago EN, Stephan KT*, et al.* Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. *Nat Med* 1998,**4**:786-793.

64. van Rij RP, Broersen S, Goudsmit J, Coutinho RA, Schuitemaker H. The role of a stromal cell-derived factor-1 chemokine gene variant in the clinical course of HIV-1 infection. *AIDS* 1998,**12**:F85-90.

65. Winkler C, Modi W, Smith MW, Nelson GW, Wu X, Carrington M*, et al.* Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). *Science* 1998,**279**:389-393.

66. Ioannidis JP, Rosenberg PS, Goedert JJ, Ashton LJ, Benfield TL, Buchbinder SP*, et al.* Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A alleles on HIV-1 disease progression: An international meta-analysis of individual-patient data. *Ann Intern Med* 2001,**135**:782-795.

67. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ*, et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008,**319**:921-926.

68. Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT*, et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 2008,**135**:49-60.

69. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC*, et al.* Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 2008,**4**:495-504.

70. Freed EO. HIV-1 and the host cell: an intimate association. *Trends Microbiol* 2004,**12**:170-177.

71. Bashirova AA, Bleiber G, Qi Y, Hutcheson H, Yamashita T, Johnson RC*, et al.* Consistent effects of TSG101 genetic variability on multiple outcomes of exposure to human immunodeficiency virus type 1. *J Virol* 2006,**80**:6757-6763.

72. Bosco DA, Eisenmesser EZ, Pochapsky S, Sundquist WI, Kern D. Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. *Proc Natl Acad Sci U S A* 2002,**99**:5247-5252.

73. An P, Wang LH, Hutcheson-Dilks H, Nelson G, Donfield S, Goedert JJ*, et al.* Regulatory polymorphisms in the cyclophilin A gene, PPIA, accelerate progression to AIDS. *PLoS Pathog* 2007,**3**:e88.

74. Rits MA, van Dort KA, Kootstra NA. Polymorphisms in the regulatory region of the Cyclophilin A gene influence the susceptibility for HIV-1 infection. *PLoS One* 2008,**3**:e3975.

75. Yamashita M, Emerman M. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J Virol* 2004,**78**:5670-5678.

76. Do H, Vasilescu A, Diop G, Hirtzig T, Heath SC, Coulonges C*, et al.* Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. *J Infect Dis* 2005,**191**:159-163.

77. Goldschmidt V, Bleiber G, May M, Martinez R, Ortiz M, Telenti A. Role of common human TRIM5alpha variants in HIV-1 disease progression. *Retrovirology* 2006,**3**:54.

78. Javanbakht H, An P, Gold B, Petersen DC, O'Huigin C, Nelson GW*, et al.* Effects of human TRIM5alpha polymorphisms on antiretroviral function and susceptibility to human immunodeficiency virus infection. *Virology* 2006,**354**:15-27.

79. Nakayama EE, Carpentier W, Costagliola D, Shioda T, Iwamoto A, Debre P*, et al.* Wild type and H43Y variant of human TRIM5alpha show similar anti-human immunodeficiency virus type 1 activity both in vivo and in vitro. *Immunogenetics* 2007,**59**:511-515.

80. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* 2008,**4**:249-259.

81. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002,**418**:646-650.

82. van Manen D, Rits MA, Beugeling C, van Dort K, Schuitemaker H, Kootstra NA. The effect of Trim5 polymorphisms on the clinical course of HIV-1 infection. *PLoS Pathog* 2008,**4**:e18.

83. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S*, et al.* AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med* 2005,**11**:1290-1292.

84. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S*, et al.* Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 2004,**432**:769-775.

85. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L*, et al.* HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* 2000,**97**:2709-2714.

86. Dalmasso C, Carpentier W, Meyer L, Rouzioux C, Goujard C, Chaix ML*, et al.* Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *PLoS One* 2008,**3**:e3907.

87. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M*, et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007,**317**:944-947.

88. Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O*, et al.* Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 2009,**199**:419-426.

89. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA*, et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med* 1997,**3**:212-217.

90. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S*, et al.* Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol* 2007,**81**:12382-12393.

91. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ*, et al.* HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 1999,**283**:1748-1752.

92. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R*, et al.* Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 2001,**344**:1668-1675.

93. Dorak MT, Tang J, Tang S, Penman-Aguilar A, Coutinho RA, Goedert JJ*, et al.* Influence of human leukocyte antigen-B22 alleles on the course of human immunodeficiency virus type 1 infection in 3 cohorts of white men. *J Infect Dis* 2003,**188**:856-863.

94. Flores-Villanueva PO, Hendel H, Caillat-Zucman S, Rappaport J, Burgos-Tiburcio A, Bertin-Maghit S*, et al.* Associations of MHC ancestral haplotypes with resistance/susceptibility to AIDS disease development. *J Immunol* 2003,**170**:1925-1929.

95. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ*, et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* 1996,**2**:405-411.

96. Steel CM, Ludlam CA, Beatson D, Peutherer JF, Cuthbert RJ, Simmonds P*, et al.* HLA haplotype A1 B8 DR3 as a risk factor for HIV-related disease. *Lancet* 1988,**1**:1185-1188.

97. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E*, et al.* Advantage of rare HLA supertype in HIV disease progression. *Nat Med* 2003,**9**:928-935.

98. Tang J, Costello C, Keet IP, Rivers C, Leblanc S, Karita E*, et al.* HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection. *AIDS Res Hum Retroviruses* 1999,**15**:317-324.

99. Goulder PJ, Bunce M, Luzzi G, Phillips RE, McMichael AJ. Potential underestimation of HLA-C-restricted cytotoxic T-lymphocyte responses. *AIDS* 1997,**11**:1884-1886.

100. Carrington M, Martin MP, van Bergen J. KIR-HLA intercourse in HIV disease. *Trends Microbiol* 2008,**16**:620-627.

101. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F*, et al.* Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nat Genet* 2007,**39**:733-740.

102. Barbour JD, Sriram U, Caillier SJ, Levy JA, Hecht FM, Oksenberg JR. Synergy or independence? Deciphering the interaction of HLA Class I and NK cell KIR alleles in early HIV-1 disease progression. *PLoS Pathog* 2007,**3**:e43.

103. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ*, et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* 2002,**31**:429-434.

104. Qi Y, Martin MP, Gao X, Jacobson L, Goedert JJ, Buchbinder S*, et al.* KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathog* 2006,**2**:e79.

105. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, Schneidewind A*, et al.* Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *J Exp Med* 2007,**204**:3027-3036.

106. Long BR, Ndhlovu LC, Oksenberg JR, Lanier LL, Hecht FM, Nixon DF*, et al.* Conferral of enhanced natural killer cell function by KIR3DS1 in early human immunodeficiency virus type 1 infection. *J Virol* 2008,**82**:4785-4792.

107. Winkler C, An P, O'Brien SJ. Patterns of ethnic diversity among the genes that influence AIDS. *Hum Mol Genet* 2004,**13 Spec No 1**:R9-19.

108. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 1992,**9**:552-569.

109. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature* 1987,**325**:31-36.

110. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J*, et al.* Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet* 1998,**63**:595-612.

111. Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000,**408**:708-713.

112. Stringer C. Human evolution: Out of Ethiopia. *Nature* 2003,**423**:692-693, 695.

113. White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G*, et al.* Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* 2003,**423**:742-747.

114. Nei M. Effective size of human populations. *Am J Hum Genet* 1970,**22**:694-696.

115. Nei M, Takahata N. Effective population size, genetic diversity, and coalescence time in subdivided populations. *J Mol Evol* 1993,**37**:240-244.

116. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM*, et al.* Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 2002,**32**:135-142.

117. Fortin A, Stevenson MM, Gros P. Susceptibility to malaria as a complex trait: big pressure from a tiny creature. *Hum Mol Genet* 2002,**11**:2469-2478.

118. Garte S. Locus-specific genetic diversity between human populations: an analysis of the literature. *Am J Hum Biol* 2003,**15**:814-823.

119. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002,**12**:1805-1814.

120. Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. Council of State and Territorial Epidemiologists; AIDS Program, Center for Infectious Diseases. *MMWR Morb Mortal Wkly Rep* 1987,**36 Suppl 1**:1S-15S.

121. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU*, et al.* Early origin and recent expansion of Plasmodium falciparum. *Science* 2003,**300**:318-321.

122. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G*, et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001,**293**:455-462.

123. Weatherall DJ. The phenotypic diversity of monogenic disease: lessons from the thalassemias. *Harvey Lect* 1998,**94**:1-20.

124. Hill AV. Malaria resistance genes: a natural selection. *Trans R Soc Trop Med Hyg* 1992,**86**:225-226, 232.

125. Hill AV, Jepson A, Plebanski M, Gilbert SC. Genetic analysis of host-parasite coevolution in human malaria. *Philos Trans R Soc Lond B Biol Sci* 1997,**352**:1317-1325.

126. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med* 1976,**295**:302-304.

127. Galvani AP, Slatkin M. Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. *Proc Natl Acad Sci U S A* 2003,**100**:15276-15279.

128. Schliekelman P, Garner C, Slatkin M. Natural selection and resistance to HIV. *Nature* 2001,**411**:545-546.

129. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI*, et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 1996,**87**:1285-1294.

130. Ke HM, Zydowsky LD, Liu J, Walsh CT. Crystal structure of recombinant human T-cell cyclophilin A at 2.5 A resolution. *Proc Natl Acad Sci U S A* 1991,**88**:9483-9487.

131. Mikol V, Kallen J, Pflugl G, Walkinshaw MD. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. *J Mol Biol* 1993,**234**:1119-1130.

132. Fischer G, Wittmann-Liebold B, Lang K, Kiefhaber T, Schmid FX. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* 1989,**337**:476-478.

133. Takahashi N, Hayano T, Suzuki M. Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. *Nature* 1989,**337**:473-475.

134. Franke EK, Yuan HE, Luban J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 1994,**372**:359-362.

135. Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 1993,**73**:1067-1078.

136. Ott DE, Coren LV, Johnson DG, Sowder RC, 2nd, Arthur LO, Henderson LE. Analysis and localization of cyclophilin A found in the virions of human immunodeficiency virus type 1 MN strain. *AIDS Res Hum Retroviruses* 1995,**11**:1003-1006.

137. Thali M, Charles M, Furman C, Cavacini L, Posner M, Robinson J*, et al.* Resistance to neutralization by broadly reactive antibodies to the human immunodeficiency virus type 1 gp120 glycoprotein conferred by a gp41 amino acid change. *J Virol* 1994,**68**:674-680.

138. Hatziioannou T, Perez-Caballero D, Cowan S, Bieniasz PD. Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. *J Virol* 2005,**79**:176-183.

139. Sokolskaja E, Sayah DM, Luban J. Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *J Virol* 2004,**78**:12800-12808.

140. Agresta BE, Carter CA. Cyclophilin A-induced alterations of human immunodeficiency virus type 1 CA protein in vitro. *J Virol* 1997,**71**:6921-6927.

141. Braaten D, Franke EK, Luban J. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J Virol* 1996,**70**:3551-3560.

142. Saphire AC, Bobardt MD, Gallay PA. Host cyclophilin A mediates HIV-1 attachment to target cells via heparans. *EMBO J* 1999,**18**:6771-6785.

143. Sherry B, Zybarth G, Alfano M, Dubrovsky L, Mitchell R, Rich D*, et al.* Role of cyclophilin A in the uptake of HIV-1 by macrophages and T lymphocytes. *Proc Natl Acad Sci U S A* 1998,**95**:1758-1763.

144. Steinkasserer A, Harrison R, Billich A, Hammerschmid F, Werner G, Wolff B*, et al.* Mode of action of SDZ NIM 811, a nonimmunosuppressive cyclosporin A analog with activity against human immunodeficiency virus type 1 (HIV-1): interference with early and late events in HIV-1 replication. *J Virol* 1995,**69**:814-824.

145. Braaten D, Luban J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. *EMBO J* 2001,**20**:1300-1309.

146. Luban J. Absconding with the chaperone: essential cyclophilin-Gag interaction in HIV-1 virions. *Cell* 1996,**87**:1157-1159.

147. Farnet CM, Haseltine WA. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J Virol* 1991,**65**:1910-1915.

148. Saphire AC, Bobardt MD, Zhang Z, David G, Gallay PA. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J Virol* 2001,**75**:9187-9200.

149. Grattinger M, Hohenberg H, Thomas D, Wilk T, Muller B, Krausslich HG. In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* 1999,**257**:247-260.

150. Saphire AC, Bobardt MD, Gallay PA. Human immunodeficiency virus type 1 hijacks host cyclophilin A for its attachment to target cells. *Immunol Res* 2000,**21**:211-217.

151. Chen Y, Maguire T, Hileman RE, Fromm JR, Esko JD, Linhardt RJ*, et al.* Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat Med* 1997,**3**:866-871.

152. Compton T, Nowlin DM, Cooper NR. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* 1993,**193**:834-841.

153. Jackson T, Ellard FM, Ghazaleh RA, Brookes SM, Blakemore WE, Corteyn AH*, et al.* Efficient infection of cells in culture by type O foot-and-mouth disease virus requires binding to cell surface heparan sulfate. *J Virol* 1996,**70**:5282-5287.

154. Krusat T, Streckert HJ. Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells. *Arch Virol* 1997,**142**:1247-1254.

155. Summerford C, Samulski RJ. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* 1998,**72**:1438-1445.

156. WuDunn D, Spear PG. Initial interaction of herpes simplex virus with cells is binding to heparan sulfate. *J Virol* 1989,**63**:52-58.

157. Kataoka N, Bachorik JL, Dreyfuss G. Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol* 1999,**145**:1145-1152.

158. Lai MC, Lin RI, Huang SY, Tsai CW, Tarn WY. A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem* 2000,**275**:7950-7957.

159. Fried H, Kutay U. Nucleocytoplasmic transport: taking an inventory. *Cell Mol Life Sci* 2003,**60**:1659-1688.

160. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D*, et al.* Transportin-SR2 imports HIV into the nucleus. *Curr Biol* 2008,**18**:1192-1202.

161. Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T*, et al.* HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog* 2011,**7**:e1001313.

162. Farnet CM, Haseltine WA. Integration of human immunodeficiency virus type 1 DNA in vitro. *Proc Natl Acad Sci U S A* 1990,**87**:4164-4168.

163. Lee YM, Coffin JM. Efficient autointegration of avian retrovirus DNA in vitro. *J Virol* 1990,**64**:5958-5965.

164. Bushman FD, Craigie R. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci U S A* 1991,**88**:1339-1343.

165. Craigie R, Fujiwara T, Bushman F. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. *Cell* 1990,**62**:829-837.

166. Farnet CM, Bushman FD. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* 1997,**88**:483-492.

167. Lin CW, Engelman A. The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J Virol* 2003,**77**:5030-5036.

168. Chen H, Engelman A. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci U S A* 1998,**95**:15270-15274.

169. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* 1994,**266**:2002-2006.

170. Yung E, Sorin M, Wang EJ, Perumal S, Ott D, Kalpana GV. Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance. *J Virol* 2004,**78**:2222-2231.

171. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y*, et al.* HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 2003,**278**:372-381.

172. Ge H, Si Y, Roeder RG. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J* 1998,**17**:6723-6729.

173. Singh DP, Kimura A, Chylack LT, Jr., Shinohara T. Lens epithelium-derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. *Gene* 2000,**242**:265-273.

174. Sharma P, Singh DP, Fatma N, Chylack LT, Jr., Shinohara T. Activation of LEDGF gene by thermal-and oxidative-stresses. *Biochem Biophys Res Commun* 2000,**276**:1320-1324.

175. Shinohara T, Singh DP, Fatma N. LEDGF, a survival factor, activates stress-related genes. *Prog Retin Eye Res* 2002,**21**:341-358.

176. Singh DP, Fatma N, Kimura A, Chylack LT, Jr., Shinohara T. LEDGF Binds to Heat Shock and Stress-Related Element to Activate the Expression of Stress-Related Genes. *Biochemical and Biophysical Research Communications* 2001,**283**:943-955.

177. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D*, et al.* Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J Biol Chem* 2005,**280**:25517-25523.

178. Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat Struct Biol* 2002,**9**:217-224.

179. Stec I, Nagl SB, van Ommen GJ, den Dunnen JT. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett* 2000,**473**:1-5.

180. Cherepanov P, Devroe E, Silver PA, Engelman A. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem* 2004,**279**:48883-48892.

181. Maertens G, Cherepanov P, Debyser Z, Engelborghs Y, Engelman A. Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *J Biol Chem* 2004,**279**:33421-33429.

182. Vanegas M, Llano M, Delgado S, Thompson D, Peretz M, Poeschla E. Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering. *J Cell Sci* 2005,**118**:1733-1743.

183. Ge H, Si Y, Wolffe AP. A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol Cell* 1998,**2**:751-759.

184. Van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z. Cellular co-factors of HIV-1 integration. *Trends Biochem Sci* 2006,**31**:98-105.

185. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci U S A* 2005,**102**:17308-17313.

186. Cherepanov P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res* 2007,**35**:113-124.

187. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z*, et al.* LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* 2003,**278**:33528-33539.

188. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M*, et al.* LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J Virol* 2004,**78**:9524-9537.

189. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F*, et al.* The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* 2005,**280**:17841-17847.

190. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. *Virology* 2006,**346**:415-426.

191. Zielske SP, Stevenson M. Modest but reproducible inhibition of human immunodeficiency virus type 1 infection in macrophages following LEDGFp75 silencing. *J Virol* 2006,**80**:7275-7280.

192. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002,**110**:521-529.

193. Bushman FD. Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 2003,**115**:135-138.

194. Sandmeyer S. Integration by design. *Proc Natl Acad Sci U S A* 2003,**100**:5586-5588.

195. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P*, et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 2005,**11**:1287-1289.

196. Ciuffi A, Bushman FD. Retroviral DNA integration: HIV and the role of LEDGF/p75. *Trends Genet* 2006,**22**:388-395.

197. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH*, et al.* An essential role for LEDGF/p75 in HIV integration. *Science* 2006,**314**:461-464.

198. Suzuki Y, Craigie R. The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol* 2007,**5**:187-196.

199. Shun MC, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P*, et al.* LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 2007,**21**:1767-1778.

200. Sutherland HG, Newton K, Brownstein DG, Holmes MC, Kress C, Semple CA*, et al.* Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations. *Mol Cell Biol* 2006,**26**:7201-7210.

201. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J*, et al.* Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J Virol* 2006,**80**:11498-11509.

202. Hombrouck A, De Rijck J, Hendrix J, Vandekerckhove L, Voet A, De Maeyer M*, et al.* Virus evolution reveals an exclusive role for LEDGF/p75 in chromosomal tethering of HIV. *PLoS Pathog* 2007,**3**:e47.

203. Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res* 2006,**34**:1653-1675.

204. Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M*, et al.* LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. *Mol Ther* 2010,**18**:552-560.

205. Shun MC, Daigle JE, Vandegraaff N, Engelman A. Wild-type levels of human immunodeficiency virus type 1 infectivity in the absence of cellular emerin protein. *J Virol* 2007,**81**:166-172.

206. De Rijck J, Vandekerckhove L, Christ F, Debyser Z. Lentiviral nuclear import: a complex interplay between virus and host. *Bioessays* 2007,**29**:441-451.

207. Cherepanov P, Sun ZY, Rahman S, Maertens G, Wagner G, Engelman A. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol* 2005,**12**:526-532.

208. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D*, et al.* Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol* 2010,**6**:442-448.

209. Yu F, Jones GS, Hung M, Wagner AH, Macarthur HL, Liu X*, et al.* HIV-1 Integrase Preassembled on Donor DNA Is Refractory to Activity Stimulation by LEDGF/p75. *Biochemistry* 2007,**46**:2899-2908.

210. Murray JM, Emery S, Kelleher AD, Law M, Chen J, Hazuda DJ*, et al.* Antiretroviral therapy with the integrase inhibitor raltegravir alters decay kinetics of HIV, significantly reducing the second phase. *AIDS* 2007,**21**:2315-2321.

211. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M*, et al.* Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J Med Chem* 2008,**51**:5843-5855.

212. Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M*, et al.* Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob Agents Chemother* 2008,**52**:1351-1358.

213. Engle LJ, Simpson CL, Landers JE. Using high-throughput SNP technologies to study cancer. *Oncogene* 2006,**25**:1594-1601.

214. Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. *Br J Cancer* 2004,**90**:747-751.

215. Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G*, et al.* Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 1999,**22**:164-167.

216. Mooser V, Waterworth DM, Isenhour T, Middleton L. Cardiovascular pharmacogenetics in the SNP era. *J Thromb Haemost* 2003,**1**:1398-1402.

217. Bernig T, Chanock SJ. Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. *Expert Rev Mol Diagn* 2006,**6**:319-331.

218. Cantor CR. The use of genetic SNPs as new diagnostic markers in preventive medicine. *Ann N Y Acad Sci* 2005,**1055**:48-57.

219. Sadee W. Pharmacogenomics: harbinger for the era of personalized medicine? *Mol Interv* 2005,**5**:140-143.

220. Divne AM, Allen M. A DNA microarray system for forensic SNP analysis. *Forensic Sci Int* 2005,**154**:111-121.

221. Ota M, Fukushima H, Kulski JK, Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nat Protoc* 2007,**2**:2857-2864.

222. Zhang R, Zhu Z, Zhu H, Nguyen T, Yao F, Xia K*, et al.* SNP Cutter: a comprehensive tool for SNP PCR-RFLP assay design. *Nucleic Acids Res* 2005,**33**:W489-492.

223. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* 2007,**35**:D269-270.

224. Chuang LY, Yang CH, Cheng YH, Gu DL, Chang PL, Tsui KH*, et al.* V-MitoSNP: visualization of human mitochondrial SNPs. *BMC Bioinformatics* 2006,**7**:379.

225. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977,**74**:5463-5467.

226. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977,**74**:560-564.

227. Hindley J, Staden R. Laboratory techniques in biochemistry and molecular biology In: *DNA sequencing*. Edited by Work TS, Burdon RH. U.S.A and Canada: Elsevier, Science Publishing Co., IN. 72 Vanderbilt avenue, new york, N.Y. 10017; 1983. pp. 73-86.

228. Stellwagen NC. Nucleic Acid Electrophoresis. In. Edited by ETietz D. New York: Springer-Verlag Berlin Heidelberg; 1998. pp. 1-26.

229. Owen MJ. DNA sequence determination using dideoxy analogs. *Methods Mol Biol* 1985,**2**:351-366.

230. Brown TA. Molecular Biology. In: *Gene Analysis*. Edited by Hames BD, Rickwood D. Second Edition ed. Califonia: 525 B Sreet, Suite 1900, San Diego, Califonia 92101-4495, USA; 2005. pp. 39-52.

231. Summer H, Gramer R, Droge P. Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *J Vis Exp* 2009,**32**:1485-1503.

232. Garrett R, Grisham CM. Biochemistry. In: *Biochemistry*. Edited by Edition F. Third ed. Belmont: Thomson Brooks/Cole; 2003. pp. 338-344.

233. Parker LT, Zakeri H, Deng Q, Spurgeon S, Kwok PY, Nickerson DA. AmpliTaq DNA polymerase, FS dye-terminator sequencing: analysis of peak height patterns. *Biotechniques* 1996,**21**:694-699.

234. Tabor S, Richardson CC. A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc Natl Acad Sci U S A* 1995,**92**:6339-6343.

235. Kornberg A, Baker TA. Replication Mechanisms and Operations. In: *DNA Replication*. Second ed: Bruce Armbruster University Science Books; 1992. pp. 471-510.

236. Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorolysis and metal ions. *J Biol Chem* 1990,**265**:8322-8328.

237. Lee LG, Connell CR, Woo SL, Cheng RD, McArdle BF, Fuller CW*, et al.* DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. *Nucleic Acids Res* 1992,**20**:2471-2483.

238. Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR*, et al.* New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res* 1997,**25**:4500-4504.

239. Sgueglia JB, Geiger S, Davis J. Precision studies using the ABI prism 3100 genetic analyzer for forensic DNA analysis. *Anal Bioanal Chem* 2003,**376**:1247-1254.

240. Paulus A, Ohms JI. Analysis of oligonucleotides by capillary gel electrophoresis. *J Chromatogr* 1990,**507**:113-123.

241. Fogle T. Information metaphors and the human genome project. *Perspect Biol Med* 1995,**38**:535-547.

242. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 1987,**155**:335-350.

243. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1986,**51**:263-273.

244. Gibbs RA. DNA amplification by the polymerase chain reaction. *Anal Chem* 1990,**62**:1202-1214.

245. Arnheim N, Erlich H. Polymerase chain reaction strategy. *Annu Rev Biochem* 1992,**61**:131-156.

246. Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. *Science* 1991,**252**:1643-1651.

247. Walker JM, Rapley R. Molecular Biology-Principles and Practices. In: *Molecular biomethods handbook* Second Edition ed. Totowa, USA: Humana Press; 2008. pp. 17-28.

248. Berashevich JA, Chakraborty T. Energy contribution of the solvent to the charge migration in DNA. *J Chem Phys* 2007,**126**:035104.

249. Zilberstein G, Korol L, Znaleziona J, Sebastiano R, Righetti PG, Shlar I*, et al.* DNA separation methodology based on charge neutralization in a polycationic gel matrix. *Anal Chem* 2008,**80**:5031-5035.

250. Sharp PA, Sugden B, Sambrook J. Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis. *Biochemistry* 1973,**12**:3055-3063.

251. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 1999,**14**:143-149.

252. McGuigan FE, Ralston SH. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. *Psychiatr Genet* 2002,**12**:133-136.

253. Afonina I, Zivarts M, Kutyavin I, Lukhtanov E, Gamper H, Meyer RB. Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder. *Nucleic Acids Res* 1997,**25**:2657-2660.

254. Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM*, et al.* The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J Biomol Tech* 2005,**16**:398-406.

255. De la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. *Mutat Res* 2005,**573**:111-135.

256. Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES*, et al.* 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 2000,**28**:655-661.

257. De La Vega FM, Dailey D, Ziegle J, Williams J, Madden D, Gilbert DA. New generation pharmacogenomic tools: a SNP linkage disequilibrium Map, validated SNP assay resource, and high-throughput instrumentation system for large-scale genetic studies. *Biotechniques* 2002,**Suppl**:48-50, 52, 54.

258. Taouji S, Dahan S, Bosse R, Chevet E. Current Screens Based on the AlphaScreen Technology for Deciphering Cell Signalling Pathways. *Curr Genomics* 2009,**10**:93-101.

259. Eglen RM, Reisine T, Roby P, Rouleau N, Illy C, Bosse R*, et al.* The use of AlphaScreen technology in HTS: current status. *Curr Chem Genomics* 2008,**1**:2-10.

260. Ullman EF, Kirakossian H, Singh S, Wu ZP, Irvin BR, Pease JS*, et al.* Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc Natl Acad Sci U S A* 1994,**91**:5426-5430.

261. Ullman EF, Kirakossian H, Switchenko AC, Ishkanian J, Ericson M, Wartchow CA*, et al.* Luminescent oxygen channeling assay (LOCI): sensitive, broadly applicable homogeneous immunoassay method. *Clin Chem* 1996,**42**:1518-1526.

262. Peppard J, Glickman F, He Y, Hu SI, Doughty J, Goldberg R. Development of a high-throughput screening assay for inhibitors of aggrecan cleavage using luminescent oxygen channeling (AlphaScreen). *J Biomol Screen* 2003,**8**:149-156.

263. Wilson J, Rossi CP, Carboni S, Fremaux C, Perrin D, Soto C*, et al.* A homogeneous 384-well high-throughput binding assay for a TNF receptor using alphascreen technology. *J Biomol Screen* 2003,**8**:522-532.

264. Beaudenon S, Huibregtse JM. HPV E6, E6AP and cervical cancer. *BMC Biochem* 2008,**9 Suppl 1**:S4.

265. Liu Y, Baleja JD. Structure and function of the papillomavirus E6 protein and its interacting proteins. *Front Biosci* 2008,**13**:121-134.

266. Sehr P, Pawlita M, Lewis J. Evaluation of different glutathione S-transferase-tagged protein captures for screening E6/E6AP interaction inhibitors using AlphaScreen. *J Biomol Screen* 2007,**12**:560-567.

267. Hou Y, McGuinness DE, Prongay AJ, Feld B, Ingravallo P, Ogert RA*, et al.* Screening for antiviral inhibitors of the HIV integrase-LEDGF/p75 interaction using the AlphaScreen luminescent proximity assay. *J Biomol Screen* 2008,**13**:406-414.

268. Reddy K, Winkler CA, Werner L, Mlisana K, Abdool Karim SS, Ndung'u T. APOBEC3G expression is dysregulated in primary HIV-1 infection and polymorphic variants influence CD4+ T-cell counts and plasma viral load. *AIDS* 2010,**24**:195-204.

269. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM*, 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**:e1954.

270. Brumme Z, Wang B, Nair K, Brumme C, de Pierres C, Reddy S*, et al.* Impact of select immunologic and virologic biomarkers on CD4 cell count decrease in patients with chronic HIV-1 subtype C infection: results from Sinikithemba Cohort, Durban, South Africa. *Clin Infect Dis* 2009,**49**:956-964.

271. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E*, et al.* CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* 2007,**13**:46-53.

272. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT*, et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999,**96**:9212-9217.

273. Braaten D, Ansari H, Luban J. The hydrophobic pocket of cyclophilin is the binding site for the human immunodeficiency virus type 1 Gag polyprotein. *J Virol* 1997,**71**:2107-2113.

274. Thali M, Bukovsky A, Kondo E, Rosenwirth B, Walsh CT, Sodroski J*, et al.* Functional association of cyclophilin A with HIV-1 virions. *Nature* 1994,**372**:363-365.

275. Sokolskaja E, Luban J. Cyclophilin, TRIM5, and innate immunity to HIV-1. *Curr Opin Microbiol* 2006,**9**:404-408.

276. Sewram S, Singh R, Kormuth E, Werner L, Mlisana K, Karim SS*, et al.* Human TRIM5alpha expression levels and reduced susceptibility to HIV-1 infection. *J Infect Dis* 2009,**199**:1657-1663.

277. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986,**42**:121-130.

278. Mattaj IW, Englmeier L. Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem* 1998,**67**:265-306.

279. Fassati A. HIV infection of non-dividing cells: a divisive problem. *Retrovirology* 2006,**3**:74.

280. Fassati A, Gorlich D, Harrison I, Zaytseva L, Mingot JM. Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. *EMBO J* 2003,**22**:3675-3685.

281. Hearps AC, Jans DA. HIV-1 integrase is capable of targeting DNA to the nucleus via an importin alpha/beta-dependent mechanism. *Biochem J* 2006,**398**:475-484.

282. Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW*, et al.* Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem* 2007,**282**:13456-13467.

283. Zaitseva L, Cherepanov P, Leyens L, Wilson SJ, Rasaiyaah J, Fassati A. HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome. *Retrovirology* 2009,**6**:11.

284. Woodward CL, Prakobwanakit S, Mosessian S, Chow SA. Integrase interacts with nucleoporin NUP153 to mediate the nuclear import of human immunodeficiency virus type 1. *J Virol* 2009,**83**:6522-6533.

285. Rain JC, Cribier A, Gerard A, Emiliani S, Benarous R. Yeast two-hybrid detection of integrase-host factor interactions. *Methods* 2009,**47**:291-297.

286. Agrawal L, Lu X, Qingwen J, VanHorn-Ali Z, Nicolescu IV, McDermott DH*, et al.* Role for CCR5Delta32 protein in resistance to R5, R5X4, and X4 human immunodeficiency virus type 1 in primary CD4+ cells. *J Virol* 2004,**78**:2277-2287.

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

288. Zimmerman PA, Buckler-White A, Alkhatib G, Spalding T, Kubofcik J, Combadiere C*, et al.* Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med* 1997,**3**:23-36.

289. Greene WC, Debyser Z, Ikeda Y, Freed EO, Stephens E, Yonemoto W*, et al.* Novel targets for HIV therapy. *Antiviral Res* 2008,**80**:251-265.

290. MacNeil A, Sankale JL, Meloni ST, Sarr AD, Mboup S, Kanki P. Genomic sites of human immunodeficiency virus type 2 (HIV-2) integration: similarities to HIV-1 in vitro and possible differences in vivo. *J Virol* 2006,**80**:7316-7321.

291. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D*, et al.* Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* 2007,**2**:e1340.

292. Daelemans D, Lu R, De Clercq E, Engelman A. Characterization of a replication-competent, integrase-defective human immunodeficiency virus (HIV)/simian virus 40 chimera as a powerful tool for the discovery and validation of HIV integrase inhibitors. *J Virol* 2007,**81**:4381-4385.

293. An P, Duggal P, Wang LH, O'Brien SJ, Donfield S, Goedert JJ*, et al.* Polymorphisms of CUL5 are associated with CD4+ T cell loss in HIV-1 infected individuals. *PLoS Genet* 2007,**3**:e19.

294. Bartholomeeusen K, Christ F, Hendrix J, Rain JC, Emiliani S, Benarous R*, et al.* Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. *J Biol Chem* 2009,**284**:11467-11477.

295. Bartholomeeusen K, De Rijck J, Busschots K, Desender L, Gijsbers R, Emiliani S*, et al.* Differential interaction of HIV-1 integrase and JPO2 with the C terminus of LEDGF/p75. *J Mol Biol* 2007,**372**:407-421.

296. Vandekerckhove L, Christ F, Van Maele B, De Rijck J, Gijsbers R, Van den Haute C*, et al.* Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus. *J Virol* 2006,**80**:1886-1896.

297. Pauwels R, Andries K, Debyser Z, Kukla MJ, Schols D, Breslin HJ*, et al.* New tetrahydroimidazo[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one and -thione derivatives are potent inhibitors of human immunodeficiency virus type 1 replication and are synergistic with 2',3'-dideoxynucleoside analogs. *Antimicrob Agents Chemother* 1994,**38**:2863-2870.

298. Appay V, Sauce D. Immune activation and inflammation in HIV-1 infection: causes and consequences. *J Pathol* 2008,**214**:231-241.

299. Finkel TH, Tudor-Williams G, Banda NK, Cotton MF, Curiel T, Monks C*, et al.* Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nat Med* 1995,**1**:129-134.

300. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M*, et al.* HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *N Engl J Med* 2011,**364**:1134-1143.

301. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD*, et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 2010,**330**:1551-1557.

302. Petrovski S, Fellay J, Shianna KV, Carpenetti N, Kumwenda J, Kamanga G*, et al.* Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS* 2011,**25**:513-518.