Genetic Variation and Differential Expression of p21 (WAF1/Cip1) in the Context of HIV-1 Control

Gemma Whitney Koor

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

Supervisors:

Professor Caroline T. Tiemessen

Dr Maria Paximadis

Johannesburg, 2016

Declaration

I, Gemma Whitney Koor (Student number: 968332) am a student registered for the degree of MSc (Med) Virology in the academic year 2016.

I hereby declare the following:

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

Signature: _____ Date: _____

Dedication

This work is dedicated to my parents, whose love and support have made me who I am

Abstract

A recent study has shown variable p21 expression levels linked to individuals displaying different levels of HIV-1 control, with elite controllers (ECs) and viraemic controllers (VCs) exhibiting higher p21 expression when compared to both healthy HIV-1 negative individuals and HIV-1-infected progressors. The role of p21 in HIV-1 control in a sub-Saharan African population has not been established.

Polymorphisms in the regulatory regions of p21, as well as in the microRNAs (miRNAs) that affect p21 regulation can contribute to differential p21 expression. In this study we developed real-time PCR assays to genotype the *p21* exonic rs1801270 and 3'UTR rs1059234 SNPs, in addition to the p21-associated miRNA (miR-106b) rs999885 SNP. We determined their allelic and genotypic frequencies in Black South African HIV-1 negative individuals (n=72), HIV-1 controllers (HICs) (n=52) further subdivided into ECs (n=11), VCs (n=30) and high viral load long term non-progressors (HVL LTNPs) (n=11), and HIV-1 infected progressors (n=74). We sequenced a region of the *p21* 5'UTR and 3'UTR in a subset of these individuals (HICs: n=52, progressors: n=44) to identify variants that may be modulating p21 expression. We compared levels of *p21* mRNA, a marker for p21 expression, in a smaller group of individuals (n=50) with similar clinical phenotypes to determine if p21 upregulation was associated with natural control of HIV-1. Lastly, we developed a real-time PCR assay to genotype a *p21* 5'UTR SNP, rs733590, that alone, and together with HLA-B*2705, was recently shown to directly impact on p21 expression in Caucasians. This SNP was genotyped and analysed in the individuals with *p21* mRNA expression data.

The *p21* rs1801270 and rs1059234 SNPs were found to occur in partial linkage disequilibrium (LD) (r^2=0.61). Although ECs had markedly less representation of the 3'UTR rs1059234 mutant allele (T) and heterozygosity (CT) compared to progressors (T allele: 9.1% ECs vs. 25% progressors; CT genotype: 18.2% ECs vs. 42% progressors), this did not reach significance (p=0.11, OR=3.33; p=0.19, OR=3.49, respectively). Interestingly, HIV-1 controllers with <400 HIV-1 RNA copies/ml (<400 HICs) also had less representation of the CT genotype when compared to progressors (20% vs. 42%, respectively; p=0.11, OR=2.91). In silico analysis of this 3'UTR SNP suggested that there are functional implications in terms of miRNA regulation, however when *p21* mRNA expression was analysed with respect to this SNP, no effect was seen. The role of this 3'UTR SNP on p21 expression and/or function and HIV-1 control requires

further investigation. The *p21* exonic rs1801270 SNP showed no difference in representation among the clinical phenotypic groups and no effect was seen on *p21* mRNA expression.

When comparing HIV-1 controllers with >400 HIV-1 RNA copies/ml (>400 HICs) to progressors, the >400 HICs had significantly lower representation of the minor allele (A) of the miR-106b rs999885 SNP (p=0.04, OR=2.28). In addition, heterozygosity for this SNP (GA) was found in a much lower frequency in >400 HICs when compared to progressors (p=0.05; OR=2.56). Stratification of individuals according to their miR-106b rs999885 SNP genotype and *p21* mRNA expression revealed the GA genotype to be associated with a trend to higher *p21* mRNA expression (p=0.066). A role for the miR-106b rs999885 SNP in HIV-1 control in individuals with higher viraemia needs to be validated in larger cohorts.

Characterisation of the *p21* regulatory regions, namely a region of the 5'UTR and the 3'UTR, identified 19 polymorphisms (18 SNPs and one indel) and 12 SNPs in the respective regions. A prevalent, previously uncharacterised 11-SNP haplotype (LD: r^2=1) was detected in the *p21* promoter region at a frequency of 39.42% in the HIV-1 controllers and 48.86% in the progressor cohort. In addition, a 2-SNP haplotype was identifed and was found to be in moderate LD with the 11-SNP haplotype (r^2=0.67). The ECs were found to have a trend of less representation of the 2-SNP haplotype minor allele when compared to progressors (p=0.08, OR=2.83). Other than the rs1059234 SNP, no other SNPs in the 3'UTR were differentially represented in any of our studied groups.

p21 mRNA expression analysis showed significant correlations between *p21* mRNA expression and markers of disease progression (HIV-1 viral load: r=0.69, p<0.0001 and CD4+ T cell count: r=-0.53, p=0.0005). In our study, ECs and VCs had significantly lower *p21* mRNA expression compared to progressors (p=0.002 and p=0.001, respectively). Furthermore, in our Black South African population (n=50), the *p21* 5'UTR rs733590 SNP CT and TT genotypes were not associated with higher *p21* mRNA expression as has been shown in Caucasians. This, together with the absence of HLA-B*2705 in our Black South African population, points to host genetic differences as the likely contributors to the different results seen in our study with respect to p21 expression and HIV-1 control when compared to reported literature.

Future work with larger sample sizes and varied population groups will be highly informative in determining the role of p21 and natural control of HIV-1 in the Black South African population.

Acknowledgements

This study was made possible by funding from the Poliomyelitus Research Foundation and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa.

To the volunteers who took part in this study, ‘thank you’ is not enough. You have allowed us to gain infinite knowledge and understanding.

To my awe-inspiring supervisors, Professor Caroline Tiemessen and Dr Maria Paximadis, you have made me love science again and I cannot thank you both enough. Professor Tiemessen, your immense knowledge has always been available to me and your passion and commitment to science is inspirational. I always leave your office feeling newly inspired and excited. Maria, your boundless energy and enthusiasm has made my day countless times, and your genuine care and interest in me not only as a student, but as a person, has made my experience at the lab so wonderful. You sacrifice your time so willingly to help not only myself, but everyone in need of assistance, and that will be forever appreciated. You have truly been a life-changing supervisor.

Shayne, thank you for always being ready and willing to help. Thank you for letting me pick your brain and for all your assistance and time. To all the members of the Tiemessen lab, thank you for the magical environment that has been created through all of you. I couldn’t ask for a better place to learn, work and be inspired.

To my amazingly supportive family, I can never thank you enough for your continual support and love and your constant interest in what I am learning. This support has been the backbone in getting me to this point. Thank you for always pushing me to follow my dreams, whatever they may be. I don’t know how I got so lucky. Shannon, thank you for being my best friend and partner in crime. Gags, you have taught me the wonders of imagination and to question the world around me and have encouraged me from the start. I hope to be as energetic and full of life as you are and to embrace all the things life has to offer.

David, thank you for your everlasting support and for going through this MSc journey with me from day one, through different countries, jobs and apartments. Your support means the world to me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADCC	Antibody dependent cell-mediated cytotoxicity
ARV	Anti-retroviral
APOBEC3G	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
bp	Base pair
BSA	Bovine serum albumin
CNS	Central nervous system
CCR5	Chemokine (C-C motif) receptor 5
CD4	Cluster of differentiation 4
cDNA	Complementary DNA
CI	Confidence interval
CNV	Copy number variant
CT	Cycle threshold
CDK9	Cyclin dependent kinase 9
CDKN1A	Cyclin dependent kinase inhibitor 1A
DC	Dendritic cell
dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
EC	Elite controller
Env	Envelope protein
EDTA	Ethylene-diamine-tetraacetic acid
FKBPL	FK506-binding protein like
gp120	Glycoprotein 120
Gag	Group-specific antigen
HC	Healthy control
Hsp	Heat shock protein
HVL LTNP	High viral load long term non-progressor
HAART	Highly active antiretroviral therapy
HIC	HIV-1 controller
HIV-1	Human immunodeficiency virus type 1
HLA	Human leukocyte antigen

HPV	Human papilloma virus
KIR	Killer cell immunoglobulin-like receptors
kb	Kilobase
LNA	Lock nucleic acid
LTNP	Long term non-progressor
MgCl2	Magnesium chloride
MHC	Major histocompatibility complex
mRNA	Messenger RNA
miRNA	Micro RNA
MCM7	Mini-chromosome maintenance 7
MGB	Minor groove binder
Mt	Mutant
NIAID	National Institute of Allergy and Infectious Diseases
NTC	No template control
OR	Odds ratio
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
Pol	Polymerase
PCR	Polymerase chain reaction
P-TEFb	Positive transcription elongation factor
PCNA	Proliferating cell nuclear antigen
RT	Real-time
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RIN	RNA integrity number
SAMHD1	SAM domain and HD domain-containing protein 1
SIV	Simian immunodeficiency virus
SNP	Single nucleotide polymorphism
SAA	South African Africans
SAC	South African Caucasians
TCR	T cell receptor
TPR	Tetratricopeptide repeat

TF	Transcription factor
UTR	Untranslated region
VC	Viraemic controller
Vif	Viral infectivity factor
VL	Viral load
VLS	Viral load setpoint
Vpu	Viral protein unique
Vpx	Virion-associated protein
WT	Wild type
WHO	World Health Organization

1 INTRODUCTION

1.1 Background

There are two major types of Human Immunodeficiency Virus (HIV), namely HIV-1 and HIV-2. Both types are similar in their gene arrangement, mode of transmission, replication pathways and clinical consequences, but HIV-2 is known to have lower transmissibility and a slower progression to Acquired Immune Deficiency Syndrome (AIDS) (reviewed by Nyamweya et al. 2013). HIV-2 is mostly endemic in West Africa, while HIV-1 accounts for 95% of HIV infections worldwide.

The HIV-1 pandemic continues to affect millions of people worldwide. The World Health Organization (WHO) estimated there to be 35 million people living with HIV-1 in 2013. It is estimated that approximately 71% of HIV-1 infected individuals worldwide live in sub-Saharan Africa. (WHO report, 2013: http://www.who.int/gho/hiv/en/index.html). In 2013, the WHO reported that over 2 million people became newly infected with HIV-1. In sub-Saharan Africa, most countries have an infection rate of >1%, with more concentrated epidemics in particular populations. The four sub-Saharan African countries with the highest adult HIV-1 prevalence rates as of 2013 are shown in Table 1.1.

South Africa alone has over 6 million people infected with HIV-1 as of 2013. The South African National HIV Prevalence, Incidence and Behaviour Survey by the Human Science Research Council (HSRC) in 2012, surveying 38 000 South Africans, showed that only 26.8% surveyed knew how HIV-1 was transmitted or how to prevent getting the virus, which is down from 30.3% in 2008.

Individuals living in sub-Saharan African countries are at an increased risk of becoming infected with HIV-1, with women being at the highest risk. Most sub-Saharan African countries have slowing economic growth and this has impacted on government spending on social services like healthcare and education. This translates to a lower ability to promote HIV-1 prevention. Poverty also has a dramatic effect on transmission of HIV-1 in sub-Saharan Africa. It was shown that the greater the national level income inequality, the higher the HIV-1 prevalence (Gillespie et al., 2007). Low economic status has also been found to be associated with earlier age of first sexual experience, lower use of condoms, having a larger number of sexual partners and a higher probability of having non-consensual sex (Mabala, 2006).

Table 1.1 - Top four sub-Saharan African countries with highest HIV-1 prevalence

Sub-Saharan Country	HIV-1 Prevalence Rate (%)
Swaziland	26.5
South Africa	17.9
Namibia	13.3
Mozambique	11.1

Reproduced and modified from UNAIDS: Global Report on the Global AIDS epidemic. Geneva, Switzerland: Joint United Nations Program on HIV/AIDS (UNAIDS); 2013

Despite these figures, infected individuals in sub-Saharan Africa are understudied when compared with other HIV-1 infected populations around the world. Given that populations differ remarkably with regard to their immunogenetic backgrounds and ultimately their response to viral infections, and that sub-Saharan populations are predominantly infected with HIV-1 subtype C, which differs from subtypes in most other regions of the world (Figure 1.1), it is important for more research to be done on these populations.

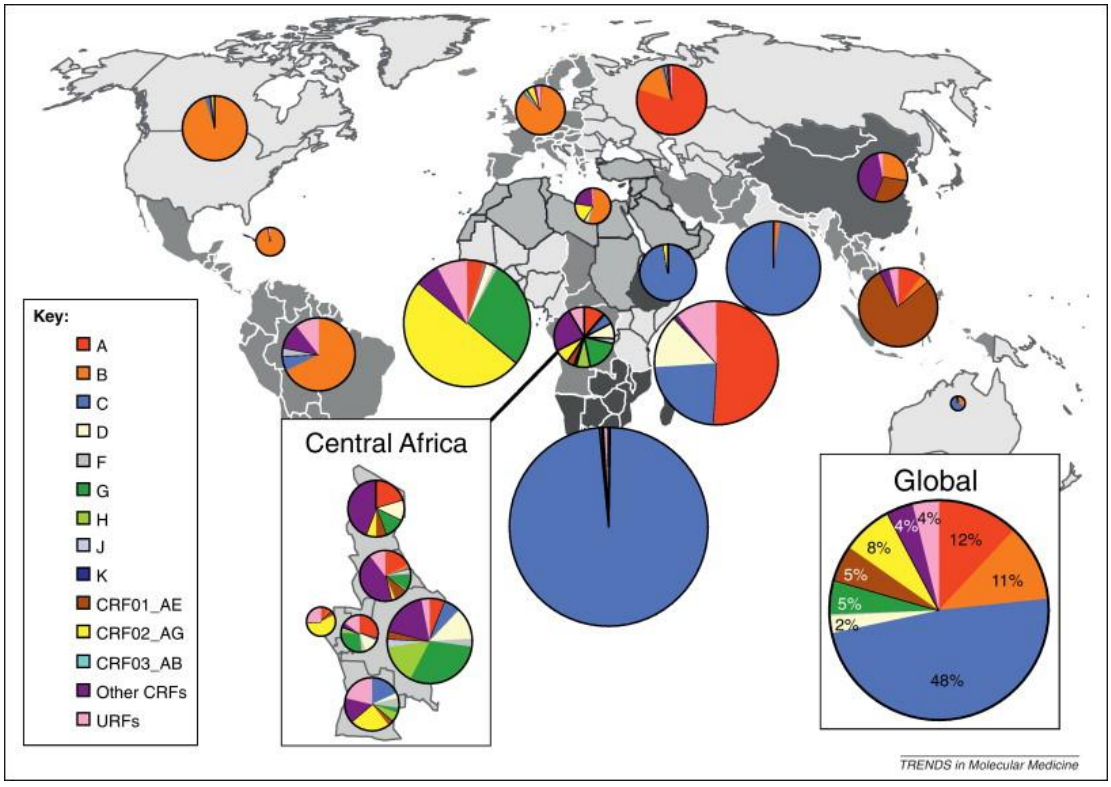


Figure 1.1 - Global map showing distribution of different HIV-1 subtypes (2012)

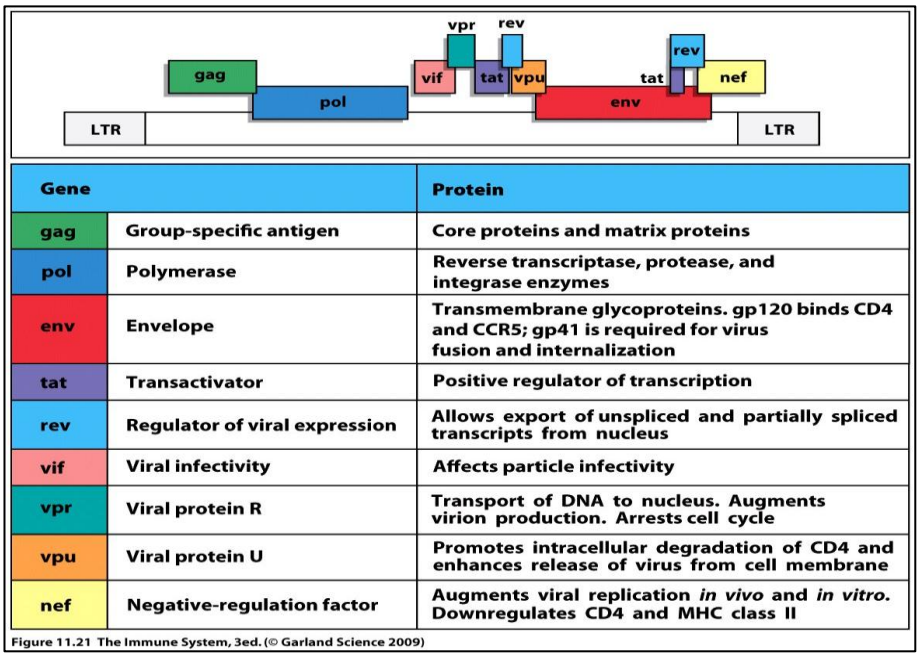
Reproduced from Hemelaar (2012)

1.2 The HIV-1 life cycle

HIV-1 is a lentivirus of the Retroviridae family and like other lentiviruses, has a relatively long incubation period. Uniquely, lentiviruses are able to infect non-dividing cells nearly as effectively as dividing cells (reviewed by Yamashita and Emerman, 2006). HIV-1, in addition to having the major genes found in all retroviruses that code for enzymes and structural proteins (i.e. *gag, pol, env*), has accessory genes that are exclusively found in HIV-1 (Figure 1.2).

HIV-1 uses the process of reverse transcription in order to replicate inside the host cell. Upon reaching the surface of a cell, HIV-1 binds and primarily uses the CD4 cell surface receptors, recognized by the HIV-1 envelope glycoprotein gp41/gp120 trimers, to gain entry into the cytoplasm (reviewed by Nisole and Saib, 2004). CD4 binding to gp120 causes a conformational change in gp120, exposing the co-receptor binding site. When CXCR4 or CCR5 bind, gp41 mediated fusion results, thereby allowing entry into the host cell.

HIV-1 ribonucleic acid (RNA) and proteins, within the viral capsid, enter into the cytoplasm of the host cell, where HIV-1's reverse transcriptase transcribes the HIV-1 RNA to complementary DNA (cDNA), and then to double stranded DNA (dsDNA). The viral DNA is then translocated into the host cell's nucleus and is integrated into the host DNA, where it is then able to undergo transcription and translation, forming new viruses (reviewed by Barre-Sinoussi et al., 2013). Cleavage of the viral polyproteins by the protease enzyme inside the immature virion allows for the formation of mature Gag proteins, leading to the production of infectious virions that start the cycle again (reviewed by Barre-Sinoussi et al., 2013).



Gene		Protein
gag	Group-specific antigen	Core proteins and matrix proteins
pol	Polymerase	Reverse transcriptase, protease, and integrase enzymes
env	Envelope	Transmembrane glycoproteins. gp120 binds CD4 and CCR5; gp41 is required for virus fusion and internalization
tat	Transactivator	Positive regulator of transcription
rev	Regulator of viral expression	Allows export of unspliced and partially spliced transcripts from nucleus
vif	Viral infectivity	Affects particle infectivity
vpr	Viral protein R	Transport of DNA to nucleus. Augments virion production. Arrests cell cycle
vpu	Viral protein U	Promotes intracellular degradation of CD4 and enhances release of virus from cell membrane
nef	Negative-regulation factor	Augments viral replication *in vivo* and *in vitro*. Downregulates CD4 and MHC class II

Figure 11.21 The Immune System, 3ed. (© Garland Science 2009)

Figure 1.2 - HIV-1 genes and encoded proteins. Schematic shows how the genes are arranged in the HIV-1 genome

Reproduced from The Immune System 3ed, Garland Science (2009)

1.3 How HIV-1 evades the immune system

The immune system employs multiple mechanisms in order to protect the body from disease. The immune system is generally divided into two main arms: the innate immune system and the adaptive immune system. The innate immune system is the first line of defence against pathogens and is necessary for the control of common bacterial/viral infections. Several different cell types form part of the innate immune system, including natural killer cells, mast cells, eosinophils, basophils, macrophages, neutrophils and dendritic cells. Receptors that are expressed on macrophages recognize a wide range of molecular patterns that are specific to pathogens (Medzhitov and Janeway, 2000).

The adaptive immune system has evolved to provide protection against reinfection with the same pathogen. This protection depends on the generation of antigen receptors, namely T-cell receptors (TCRs) and immunoglobulins, which result from somatic rearrangement processes in blast cells (Hansson et al., 2002). Foreign antigen is presented to the immune system in association with human leukocyte antigen (HLA) class I and II molecules on antigen presenting cells. When T cells are presented with foreign antigen, the adaptive immune response is initiated against the specific antigen presented. The initiated adaptive responses

include direct targeting and killing of cells containing the antigen, mediated by cytotoxic T lymphocytes, and the production of antibodies against the antigen, mediated by stimulated B cells (Hansson et al., 2002).

Because HIV-1 has an error-prone reverse transcriptase (Boyer et al., 1992), the virus frequently acquires mutations in its genome that either aid or hinder its survival via altered viral protein production. Through this evolutionary journey, HIV-1 has gained the ability to evade the immune system via immune escape mutations, as well as by using other distinct mechanisms (Table 1.2).

HIV-1 infects and destroys CD4+ T cells and therefore interferes with the functioning of the immune system. The immune system has mechanisms in place to attack the HIV-1 infected CD4+ T cell (Figure 1.3) but as more and more CD4+ T cells are infected with HIV-1, the massive destruction leads to a loss of immune function. Due to this inability of the immune system to effectively neutralize or kill HIV-1, researchers have had to focus on different strategies in order to be able to control HIV-1 infection.

Table 1.2 - Host defences and mechanisms of HIV-1 evasion

Immune Response	Host Defence *	Antiviral Effect	Viral Evasion or Antagonistic Mechanism	Viral Factor(s) or Properties
INNATE	NK cells	Lysis of infected cells	Selective downmodulation of HLA-A and -B, but not HLA-C and -E	Nef
INTRINSIC	ABOBEC 3G	Lethal hypermutations	Polyubiquitination and degradation	Vif
	TRIM5α	Untimely uncoating	Variation in capsid	High variability
	Tetherin	Blocks virion release	Sequestration from the site of virion budding	Vpu, Nef, Env
ADAPTIVE	Cytotoxic CD8+ T cells	Lysis of infected cells Inhibitory cytokines	MHC-I downmodulation, escape mutations, latent infection	Nef, high variability
	CD4+ helper T cells	Helper function to promote antibody and CTL responses	Destruction by infection or bystander apoptosis; downmodulation of CD4, CD3, CD28, and CXCR4	Nef, Vpu, viral cytopathicity
	B cells, antibodies	Neutralization	Antigenic variation, glycosylation, shielding of functional epitopes, inhibition of IgG2, and IgA switching	High variability, N-linked glycosylation sites, Env structure, Nef

Reproduced and adapted from Kirchhoff (2010)

**This table describes only some of the more well studied mechanisms*

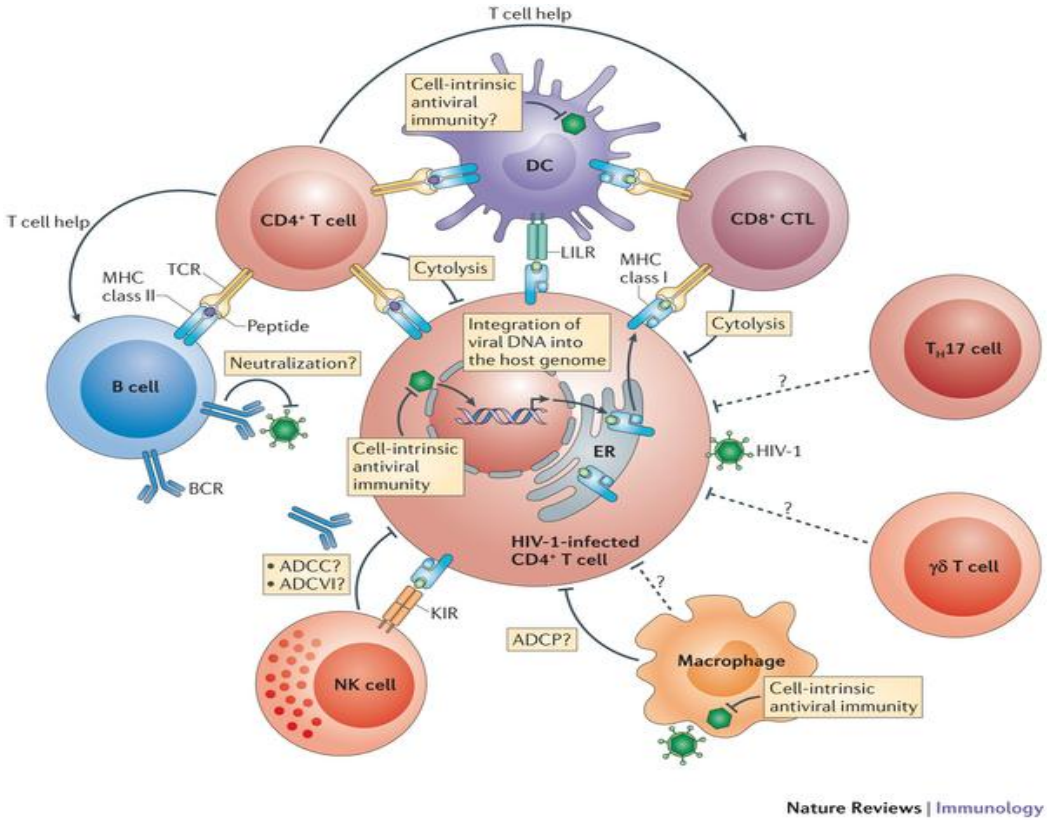


Figure 1.3 - Schematic of the immune response against HIV-1 infected CD4+ T cells. ADCC: antibody-dependant cell-mediated cytotoxicity. MHC: major histocompatibility complex. TCR: T cell receptor. DC: dendritic cell. T_h: T helper cell. ER: endoplasmic reticulum.

Reproduced from Walker and Yu (2013)

1.4 Anti-retrovirals and consequences on HIV-1 infection

Anti-retroviral drugs (ARVs) are the only current treatment available for HIV-1 infected individuals. There have been discoveries of different drugs which target different stages of HIV-1 replication (Figure 1.4). During the HIV-1 replication process, for every 1000 to 10 000 nucleotides synthesized, approximately one mutation is generated (Abram et al., 2010). With the HIV-1 genome being 10 000 nucleotides in length, this results in one to ten mutations being introduced in each viral genome with every replication cycle. As a result, these random inserted mutations contribute to ARV resistance. Therefore, using only one ARV results in a much higher probability of HIV-1 developing resistance. To overcome this, three or more different drugs are used together to treat an infected individual. This form of treatment is known as highly active anti-retroviral therapy (HAART).

ARVs have allowed for substantial improvements in the management of HIV-1 in infected patients, but there are several disadvantages associated with ARV therapy. These include complex regimens, the need for and cost of lifelong treatment, ARV resistance and the development of side effects including, but not limited to, gastrointestinal issues, central nervous system problems and haematological disturbances (Rudorf and Krikorian, 2005). Additionally, ARVs do not allow for full eradication of HIV-1 due to the persistence of latent viral reservoirs occurring in long-lived memory CD4+ T cells and immune cells in the central nervous system (CNS) that contain integrated provirus within host cellular DNA (Archin et al., 2014). These reservoirs are hidden from the immune system and unaffected by ARV therapy.

In sub-Saharan Africa, improved HIV-1 care and treatment has resulted in an increased distribution of ARVs from 5% to 30% during 2004-2007, but for each new individual started on ARVs, approximately 2-3 new infections have been reported (reviewed by Braunstein et al., 2009), emphasizing the critical need for research into preventative and therapeutic treatments including vaccines for HIV-1.

Recommendations regarding when to initiate HAART are frequently changing. The current guidelines in South Africa, as outlined by the WHO, are to commence treatment in an individual when their CD4+ T cell count is <500 cells/µl (*http://www.who.int/hiv/mediacentre/news/niaid_start/en/)*. Interim results from the United

States National Institute of Allergy and Infectious Diseases (NIAID) suggest that HIV-1 infected individuals will benefit from ARV treatment regardless of CD4+ T cell count.

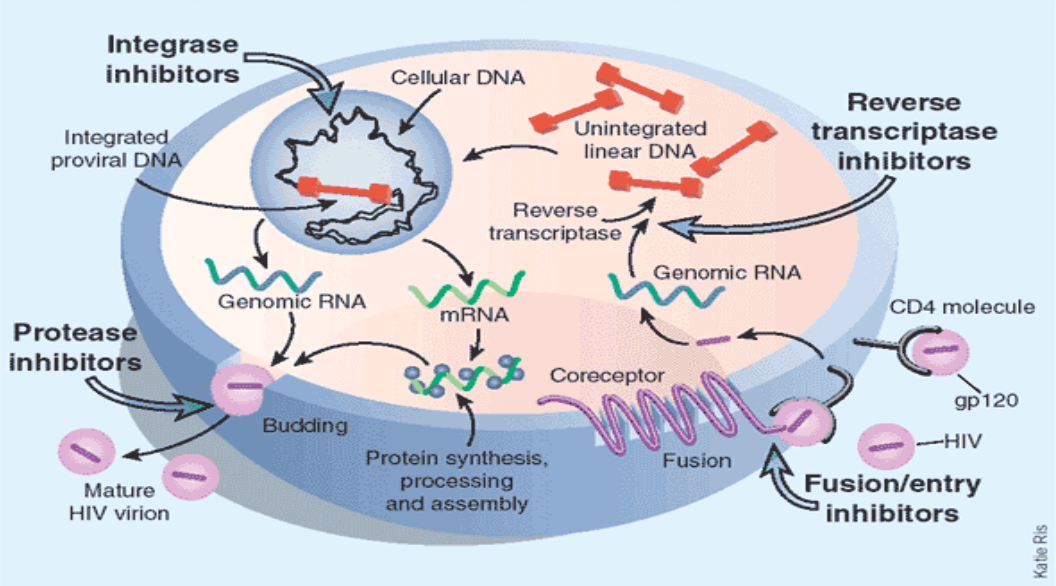


Figure 1.4 - Schematic showing anti-HIV mechanisms of available anti-retroviral drugs

Reproduced from http://kesehatan.kompasiana.com/medis/2011/03/29/sekilas-tentang-pengobatan-hivAIDS-350452.html. Accessed 24/02/2015

1.5 Natural control of HIV-1

Individuals infected with HIV-1 show substantial variation in rates of disease progression. Most commonly, progression of HIV-1 results in lowered CD4+ T cell count and increased viral load over time. Peak levels of HIV-1 viral RNA occur after initial infection, after which levels decrease until a viral load set point (VLS) is reached where viral RNA levels are maintained for months to years (Kelley et al., 2007). While it is still unclear what determines VLS, a number of factors, both viral and immunological, have been proposed to have an influence. The VLS has been found to be a good predictor of disease progression (Figure 1.5) (Mellors et al., 1997, Richardson et al., 2003), with a higher viral load at the VLS being associated with faster disease progression to AIDS and consequently death (Lavreys et al., 2006, Koehler et al., 2010).

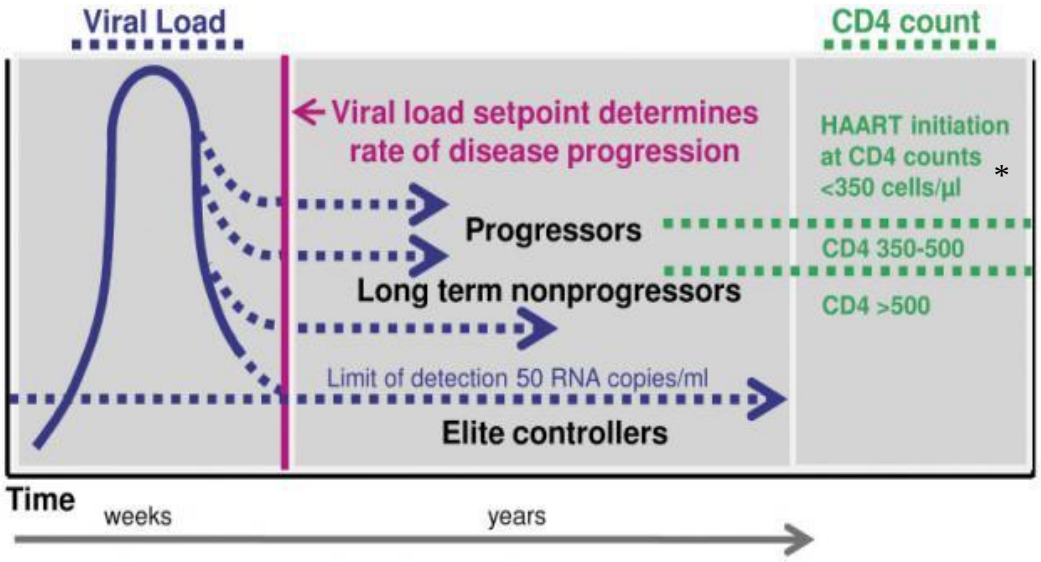


Figure 1.5 – Graph showing varying viral set points and consequence on HIV-1 progression

* Current WHO guidelines recommend initiation of HAART at <500 cells/µl, with interim data from NIAID suggesting HIV-1 infected individuals start ARV treatment regardless of CD4+ T cell count

Reproduced from Tiemessen and Martinson (2012)

Among HIV-1 infected individuals, there are unique individuals who are able to naturally suppress the virus and exhibit slow progression of the disease, without the use of ARVs. HIV-1-infected viraemic controllers (VCs) are normally defined as maintaining high CD4+ T cell counts and low detectable viral loads (<2000 RNA copies/ml), while rare individuals [less than 1% of HIV-1 infected individuals in previously studied cohorts (Okulicz et al., 2009)] who maintain high CD4+ T cell counts and undetectable viral loads (<50 RNA copies/ml) after infection with HIV-1 are termed elite controllers (ECs) (Saez-Cirion et al., 2007). There has been much interest in the idea of a functional cure for HIV-1. For an individual to be 'functionally cured' of HIV-1, they would need to exhibit long-term control of viral replication to undetectable levels without the aid of anti-retrovirals (ARVs) (Katlama et al., 2013). Essentially, they would exhibit clinical features characteristic of ECs.

It is not, however, always simple defining an individual in the different controller subgroups. The many different clinical phenotypes of HIV-1 controllers (HICs) in the literature differ in terms of naming of the different controller groups, as well as how each controller group is defined. What complicates matters is that not all HICs are necessarily long term non-progressors, as some individuals who are able to maintain low levels of viraemia can still

have significant decreases in their CD4+ T cells and may even experience AIDS events (Hunt et al., 2008). On the other end of the spectrum are HICs who do not necessarily suppress HIV-1 viral load [who are named high viral load long term non-progressors (HVL LTNPs) in this study]. These individuals, while maintaining CD4+ T cell counts of >500 cells/μl long term without ARVs, tend to have relatively high viral loads of >10 000 RNA copies/ml, a similar phenotype to that of SIV-infected sooty mangabeys who have high viral loads yet show no AIDS-related events. Interestingly, there was a report of an individual infected with HIV-1 who had a viral load of >150 000 RNA copies/ml yet maintained normal CD4+ T cell counts without the use of ARVs for over 10 years (Okulicz et al., 2009).

Thus, this variability in levels of control in HICs necessitates stratification of the controller groups into more specific subgroups, as the outcomes of disease vary significantly between HICs falling within these different definitions. VCs were shown to have a much faster disease progression than ECs, and HICs maintaining CD4+ T cell counts of around 1500 cells/μl over 7 years had increased mortality rates when compared to HICs maintaining these high CD4+ T cell counts for 10 years or more (Okulicz et al., 2009). Thus, grouping all HICs into one group may mask the identification of factors that may be of importance, as different immune factors may be at play in the different 'modes' of HIV-1 control.

Further study into how ECs and other individuals classified as HICs naturally control HIV-1 or the effects of viraemia will be beneficial in determining the correlates of protective immunity which are essential for the rational design of vaccines and novel therapies.

1.6 Mechanisms of Control

Both the HIV-1 pathogen and the host have an inherent variability that result in differing levels of individual control with regards to the progression of HIV-1 (reviewed by Santa-Marta et al., 2013). It has been hypothesized that less virulent strains of HIV-1 may be resulting in the perceived control an individual may exhibit (reviewed by Tiemessen and Martinson, 2012). However, it was shown that in the vast majority of cases, viruses isolated from HIV-1 controllers were replication-competent viruses, suggesting that host-mediated control is the main mechanism of viral suppression leading to HIV-1 control (reviewed by Tiemessen and Martinson, 2012).

Previous research has shown multiple ways in which HIV-1 controllers may suppress viral replication including chemokine receptor variation, human leukocyte antigen (HLA) and killer cell immunoglobulin-like (KIR) variation, and intrinsic host proteins, to name a few. It is important to note that only a subset of ECs and HICs previously studied have an identified protective characteristic (viral and/or host) (reviewed by Okulicz, 2012), suggesting that there are other as yet unidentified factors involved that allow for the control of HIV-1 by these individuals.

Variation in chemokine receptor gene 5 (*CCR5*), *HLA* and *KIR* genes are some of the most commonly studied host factors with regards to differential HIV-1 control.

1.6.1 Variation in chemokine receptor gene *CCR5*

One of the most studied genetic variations with regards to a host protein is the Δ32 mutation in *CCR5* (*CCR5Δ32*). CCR5, together with the CD4+ T cell receptor, is responsible for allowing HIV-1 entry into the cell. Individuals homozygous for the *CCR5Δ32* mutation tend to be resistant to R5 HIV-1 strains as these strains use the CCR5 receptor to gain entry to CD4+ T cells and macrophages. HIV-1 infected individuals heterozygous for the *CCR5Δ32* polymorphism have been found to have a significantly slower disease progression (Mummidi et al., 1998, Mulherin et al., 2003). The polymorphism results in the CCR5 co-receptor being rendered non-functional and not expressed on the cell surface due to a 32 bp deletion in the gene sequence which introduces a premature stop codon (Santa-Marta et al., 2013). This polymorphism is predominantly found in European populations and is essentially absent in African, East Asian, and American Indian populations (Sabeti et al., 2005). Single nucleotide polymorphisms (SNPs) identified in the *CCR5* cis-regulatory region have also been attributed to differential control of HIV-1, particularly the *CCR5-59029* A/G polymorphism. Individuals who have a G/G genotype were found to progress slower to AIDS and/or death than individuals with an A/A genotype at the SNP position (Santa-Marta et al., 2013).

There are also various *CCR5* haplotypes that have been shown to have an effect on HIV-1 disease progression. In the study by Gonzalez et al. (1999), findings showed that the *CCR5* haplotypes that are shown to be associated with altered rates of HIV-1 disease progression are different in Caucasians as compared to African Americans (i.e. individuals with African descent).

A study was done on a population of HIV-1 uninfected South African Africans (SAA) and South African Caucasians (SAC) looking at haplogroups that were previously defined by Gonzalez et al. (1999) (Picton et al., 2010). Figure 1.6 shows these haplogroups as well as their frequencies as reported by Picton et al. (2010). Results showed that within the SAA population, there were 5 predominant haplotypes, compared to 3 found within the SAC population. Interestingly, the one shared haplotype, SAA/C-HHC (Figure 1.6) was found to be the most frequent haplotype in the SAC study group and conversely, was found least frequently in the SAA study group (Picton et al., 2010).

It was shown that HHA haplotypes (Figure 1.6) were linked with slower disease progression in African individuals of African descent (Gonzalez et al., 1999). Gonzalez et al. (1999) also showed that individuals with the HHF*2 haplotype had a slower progression to AIDS (p=0.01) as well as a slower progression to death (p=0.005). This was only found in the African American individuals studied and not in Caucasians. Individuals homozygous and heterozygous for the HHF*1 haplotypes were found to be more likely to have a faster progression to AIDS (p=0.05) and death (p=0.04) (Gonzalez et al., 1999). This was found to be true for the entire cohort (p=0.05) and in African Americans (p=0.04).

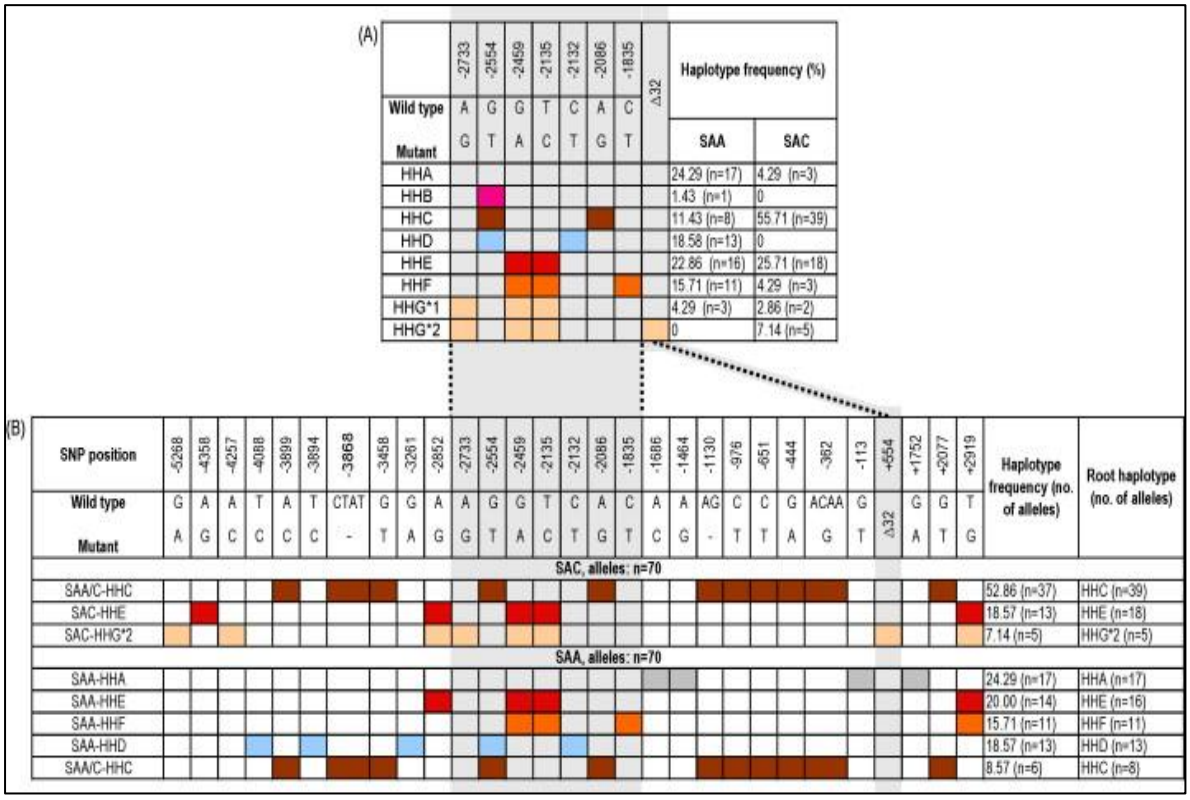


Figure 1.6 – Schematic showing haplogroups found in *CCR5* gene. (A) Schematic showing haplotypes defined by Gonzalez et al. (1999) as well as the frequencies of the haplotypes within the South African African and South African Caucasian groups in the study by Picton et al. (2010). (B) Schematic showing the haplotypes that were identified within the *CCR5* gene in the South African African and South African Caucasian study populations. The coloured boxes show the SNPs or indels that form the haplotype.

Reproduced from Picton et al. 2010

1.6.2 Variation in HLA and KIR molecules

The HLA and KIR molecules are both highly non-conserved and prone to multiple mutations and variation (Heinrichs and Orr, 1990, Yawata et al., 2002). NK cells express the KIR molecules on their surface which aid in the killing ability of NK cells by interacting with HLA molecules as ligands in order to detect infected cells or tumour cells (Middleton et al., 2002). HLA molecules also present peptide antigen to T cells, and if the antigen presented is recognized as non-self (i.e. a pathogen), the immune cells will be prompted to activate in order to neutralize or kill the invading microbe. Both of these molecules are therefore involved in important immune processes including antigen presentation to T cells and regulation of natural killer (NK) cell responses, interacting in both innate and adaptive immune responses.

There are two classes of HLA molecules, class I (consisting of HLA-A, HLA-B and HLA-C alleles) and class II (consisting of HLA-DP, HLA-DQ and HLA-DR alleles), each presenting mainly to CD8+ T cells and CD4+ T cells respectively (Figure 1.7). The most variable of these alleles are the HLA-B and HLA-DRB1 alleles.

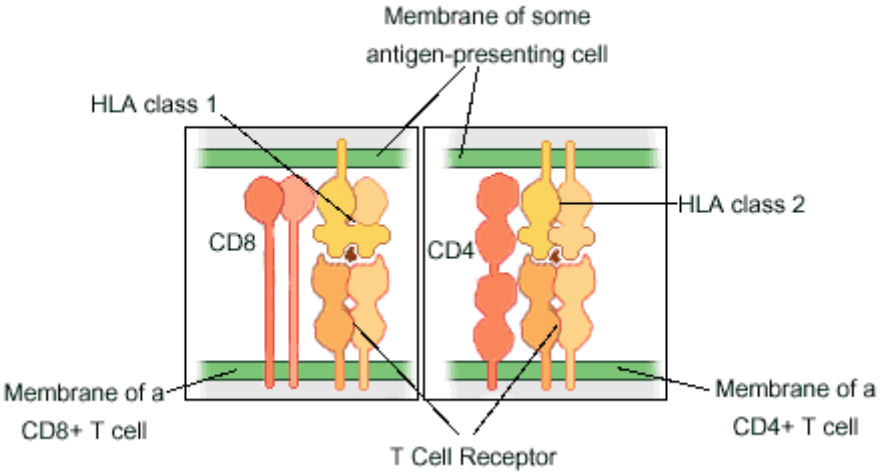


Figure 1.7 – Schematic showing a class I HLA molecule presenting antigen to a CD8+ T cell and a class II HLA molecule presenting antigen to a CD4+ T cell.

Reproduced from http://www.mcld.co.uk/hiv/?q=HLA

The extreme diversity of HLA and KIR molecules has a significant impact on disease susceptibility and severity experienced among different individuals (Bashirova et al., 2011).

HIV-1 controllers have been shown to have an overrepresentation of HLA-B*57 and -B*27 alleles (Kiepiela et al., 2004, Pereyra et al., 2008). When studying HIV-1 positive individuals during acute infection, Altfeld et al. (2003) found that virus-specific CD8+ T-cell responses were dominated by HLA-B*57-restricted responses. Using stepwise regression modelling in a European sample of controllers and progressors, HLA-B*57:01, -B*27:05, -B*14/Cw*08:02, -B*52, and -A*25 were shown to have a protective effect on HIV-1 control and HLA-B*35 and -Cw*07 were shown to negatively impact control of HIV-1 (reviewed by Goulder and Walker, 2012) .

When studying HIV-1 cohorts during viral infection, it was found that there is a robust synergistic effect of *KIR* and *HLA* genes/alleles (Martin et al., 2002). HLA-B molecules express the Bw4 epitope at amino acid position 80, either containing an isoleucine (referred to as HLA-Bw4^{80I}) or a threonine residue (referred to as HLA-Bw4^{80T}) (Cella et al., 1994, Gumperz et al., 1995). The KIR3DL1 inhibitory receptor binds to this Bw4 epitope, more strongly at HLA-Bw4^{80I}. A protective effect of a joint KIR-HLA genotype has been found, consisting of KIR3DS1 and Bw4^{80I,} with regards to natural killer (NK) cell mediated killing of HIV-1 infected cells (Figure 1.8) (Martin et al., 2002). This particular genotype was also shown to be associated with a lower viral load (Qi et al., 2006).

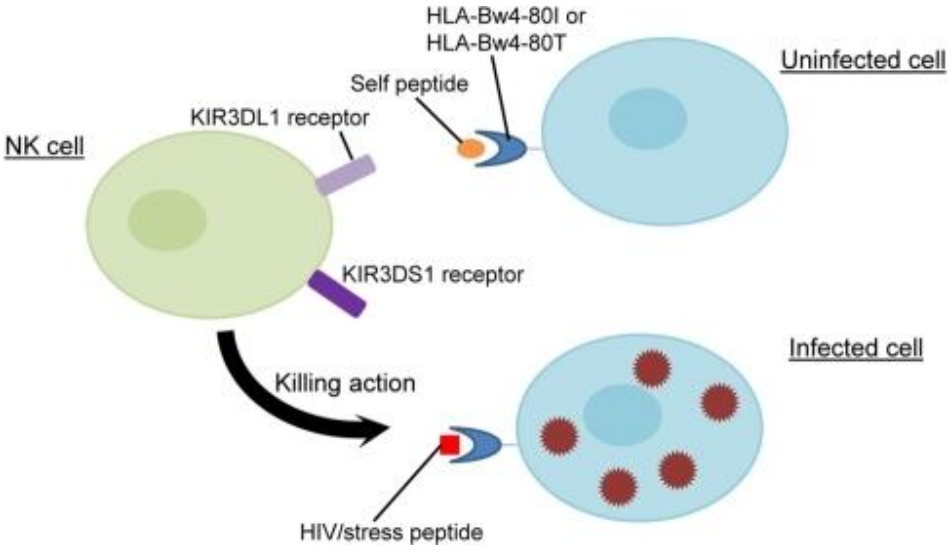


Figure 1.8 - Schematic showing NK cell with KIR3DS1 and killing an HIV-1 infected cell with HLA-Bw4^{80I} presenting a viral epitope. The HIV-1 infected cell presents a stress molecule to the NK cell that prompts the killing of that cell by the NK cell.

Reproduced from Pelak et al. (2011)

1.6.3 Intrinsic factors

There are pre-existing proteins that intrinsically exert an antiviral effect on HIV-1, including but not limited to tetherin, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G (APOBEC3G), and SAM domain and HD domain-containing protein 1 (SAMHD1) (Figure 1.9). These intrinsic antiviral factors target different steps of the HIV-1 replication cycle. However, they are not necessarily effective at combatting HIV-1, as HIV-1 has evolved accessory proteins that counteract these factors, in addition to using other not yet understood mechanisms.

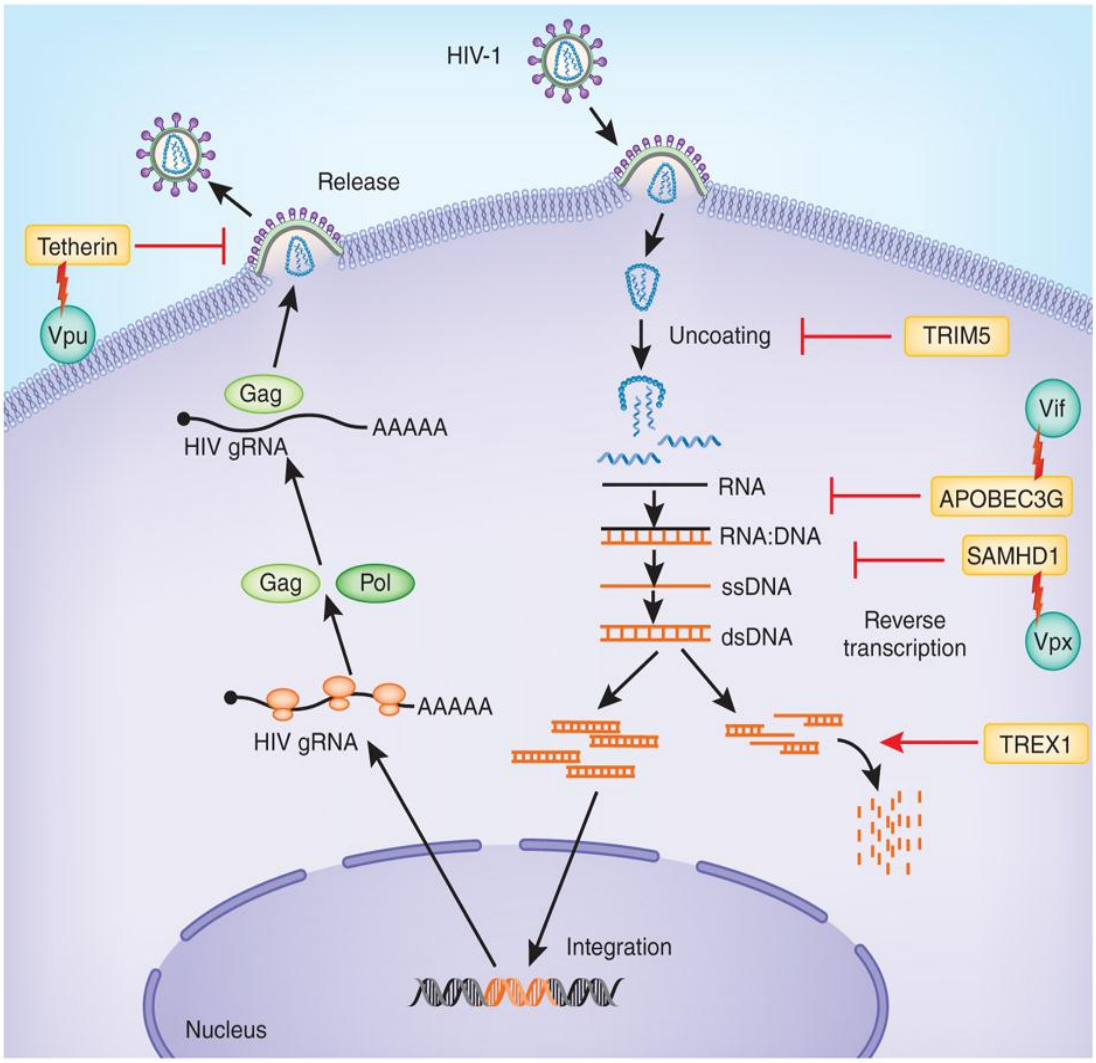


Figure 1.9 - Schematic showing antiviral intrinsic factors at different steps in the HIV-1 replication cycle. Intrinsic factors are shown in yellow boxes and HIV-1 accessory proteins that target each intrinsic factor are shown in dark green circles.

Reproduced from Yan and Chen (2012)

1.6.3.1 Tetherin

Tetherin is an interferon-inducible membrane protein that tethers nascent HIV-1 viral particles and inhibits their release from the cell surface (Venkatesh and Bieniasz, 2013). Tetherin therefore aids in the restriction of cell to cell infection of HIV-1. Tetherin is comprised of a short N-terminal cytosolic tail, a single pass transmembrane helix and an extracellular domain that is primarily alpha helical (Hinz et al., 2010). The HIV-1 accessory protein, viral protein unique (Vpu), downmodulates the cell surface expression of tetherin, leading to its degradation, and therefore allows for increased viral release (Kuhl et al., 2010).

1.6.3.2 APOBEC3G

APOBEC3G is a cytidine deaminase protein that acts to halt HIV-1 infection by converting cytosine residues into uracil within the minus strand of HIV-1 viral DNA, thereby interfering with the reverse transcription process (Wang et al., 2012b). APOBEC3G is predominantly expressed in cells that are targets of HIV-1 infection, such as CD4+ T cells, macrophages and dendritic cells (Chiu et al., 2005). The HIV-1 encoded viral infectivity factor (Vif) disrupts ABOPEC3G activity by targeting APOBEC3G for ubiquitination and degradation (Chiu et al., 2005).

1.6.3.3 SAMHD1

SAMHD1 is a cellular enzyme that is able to block replication of the HIV-1 virus in dendritic cells, macrophages and monocytes (Laguette et al., 2011, Hrecka et al., 2011). SAMHD1 does this by converting nucleotide triphosphates into triphosphate and a nucleoside, thus depleting the amount of deoxynucleotide triphosphates (dNTPs) in the cell. HIV-1 needs dNTPs in order to synthesise viral cDNA (Lahouassa et al., 2012). Recently, SAMHD1's nuclease activity has also been attributed to its ability to restrict HIV-1, as it was shown that ribonuclease activity is required for HIV-1 restriction (Ryoo et al., 2014). The Vpx (virion-associated protein) accessory protein is encoded by HIV-2 and SIV, but is absent in HIV-1. Vpx is able to counteract SAMHD1 by inducing its degradation via the ubiquitin-proteasome pathway (Laguette et al., 2012). The degradation of SAMHD1 results in an increased availability of dNTPs which in turn allows for effective viral reverse transcription. Vpx is found in HIV-2 virions in large amounts due to its requirement in HIV-2 reverse transcription (Wu et al., 1994, Fujita et al., 2008). It has been found that HIV-1 infected ECs have a higher expression of SAMHD1 when compared to healthy controls or progressors (Riveira-Munoz et al., 2014).

1.7 A role for a cyclin-dependent kinase inhibitor p21 in control of HIV-1 infection

A protein proposed to play a role in HIV-1 disease control is the cellular protein p21 (also known as CDKN1A/WAF-1/Cip-1) (Zhang et al., 2007, Bergamaschi et al., 2009, Chen et al., 2011). p21 is a 164 amino acid (18 kDa) *cip/kip* family inhibitor encoded by the *CDKN1A* gene on chromosome 6 and acts as a regulator of the cell cycle by inhibiting cyclin-dependent kinases and regulating the transition of replicating cells from G_1 - S (Arias et al., 2007). The

CDKN1A gene is comprised of three exons and two introns (Figure 1.10), with exon 2 being the main translation region (Sivonova et al., 2013).

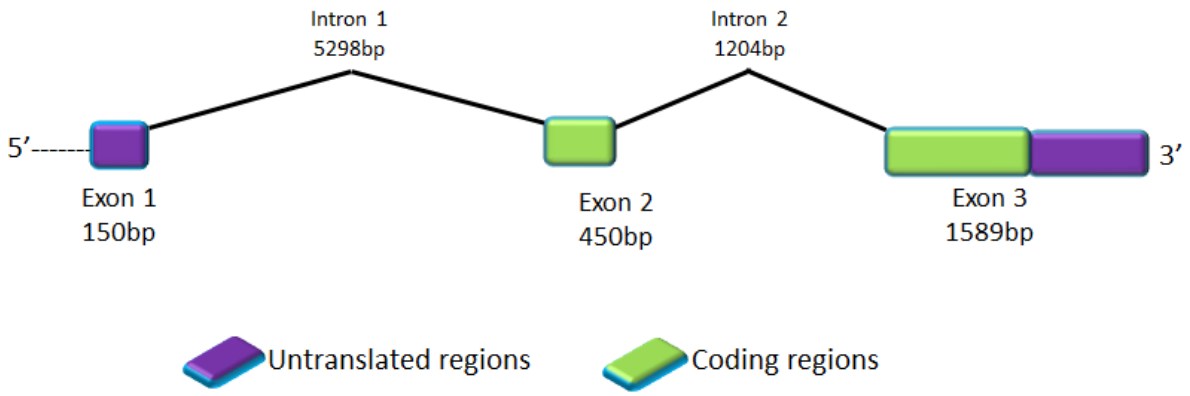


Figure 1.10 – Schematic showing the *CDKN1A* gene composed of two introns and three exons. The coding region is shown in green. Translation begins in Exon 2.

p21 is a well-known tumour suppressor (Poole et al., 2004, Kawamura et al., 2009) and mostly acts in conjunction with another significant tumour suppressor p53, where expression of p21 is induced in both a p53-dependent and independent manner after DNA damage (Macleod et al., 1995, reviewed by Gartel and Tyner, 2002). The *p21* gene contains a p53 binding site located 2.4 kb upstream of the translational start site (el-Deiry et al., 1993). p21 also has the ability to bind proliferating cell nuclear antigen (PCNA) and in this way, PCNA-dependent DNA polymerase activity is affected, resulting in inhibition of DNA replication and affecting multiple PCNA-dependent DNA repair processes (Li et al., 1994, Abbas and Dutta, 2009).

1.7.1 Upregulation of p21 in elite controllers

CD4+ T cells of HIV-1 infected ECs have been shown to have a noticeable upregulation of p21 and its associated mRNA (Chen et al., 2011). p21 has been found to inhibit HIV-1 replication in specific cell types such as macrophages (Bergamaschi et al., 2009, Allouch et al., 2013) and hematopoietic stem cells (Zhang et al., 2007) and has been shown to inhibit mRNA transcription by the binding to and inhibition of CDK9 (Chen et al., 2011), a cyclin-dependent kinase that is associated with the P-TEFb (positive transcription elongation factor

b) complex (an elongation factor for transcription directed by RNA polymerase II). By lowering the deoxyribonucleotide (dNTP) stores in macrophages enough to prevent viral cDNA synthesis, p21 has been found to suppress reverse transcription of both HIV-1 and SIV (Allouch et al., 2013). Allouch et al. (2013) showed that this reduction in dNTPs is a result of p21 repressing the expression of ribonucleotide reductase subunit R2, an enzyme necessary for the biosynthesis of dNTPs, via a SAMHD1-independent pathway.

In hematopoietic stem cells, p21 complexes with HIV-1 integrase and consequently prevents chromosomal integration of HIV-1 (Zhang et al., 2007). The ability of p21 to inhibit multiple replication events of HIV-1 suggests that it may be very efficient at limiting HIV-1 progression. When studying CD4+ T cells with regards to susceptibility to HIV-1 infection, Chen et al. (2011) found a highly significant inverse correlation ($p<0.0001$) between *p21* mRNA expression levels and ability of HIV-1 to infect the cell, thereby indicating that p21 may play a pivotal role in how CD4+ T cells are able to resist infection by HIV-1 in ECs.

Another study showed broad p21 expression regardless of individual phenotype (de Pablo et al., 2015), however they did show a trend to higher p21 expression in ECs. In the context of protection from HIV-1 infection, Herbeck et al. (2015) concluded that if p21 expression does contribute to the HIV-1 resistance in HIV-exposed seronegatives (HESN), it most likely only plays a minor role and possibly only in those that have had high levels of exposure to HIV-1.

1.7.2 Genetic variation in *p21*

Multiple SNPs have been identified in the *p21* gene with varying associated outcomes (Li et al., 2005, Gravina et al., 2009, Wang et al., 2012a, Ma et al., 2013), but have not been investigated in the context of HIV-1 infection. The two most studied *p21* SNPs, rs1801270 (also referred to as C98A or Ser31Arg) and rs1059234 (also known as *p21*C70T), have been implicated in cancer, and frequencies of these SNPs differ greatly depending on ethnicity and geography of studied populations (Sivonova et al., 2013). The frequency and representation of the two SNPs in the sub-Saharan African population as well as other reference populations are shown in Table 1.3.

Table 1.3 - SNP frequencies and representation in sub-Saharan Africans

***p21* SNP frequency in differing populations**						
SNP	**Location in gene**	**Nucleotide change**	**MINOR ALLELE FREQUENCY (sample number)**			
			Caucasian	**African**	**Asian**	**Global**
rs1801270	+93	C/A	0.040 (N=224)	0.308 (N=224)	0.394 (N=170)	0.244 (N=172)
rs1059234	+1719	C/T	0.035 (N=226)	0.241 (N=224)	0.401 (N=172)	0.128 (N=180)
Frequency and representation of SNP genotypes in sub-Saharan Africans (N=220)						
SNP			**Mt/Mt**	**WT/Mt**		**WT/WT**
rs1801270			0.100	0.407		0.493
rs1059234			0.045	0.393		0.562

Source: NCBI dbSNP database

The rs1801270 mutation results in a serine to arginine substitution at codon 31 of *p21* (Wang et al., 2012a). It is still not clear as to whether this polymorphism is beneficial or deleterious with regards to cancer as there are conflicting results in a number of studies. This can be attributed to the fact that p21 not only functions as a growth inhibitor, but can also act as an inhibitor of apoptosis, leading to paradoxical results (Roninson, 2002). The rs1059234 SNP is found within the 3' untranslated region (UTR) of the *p21* gene, and results in a single C-T substitution 20 nucleotides downstream of the exon 3 stop codon.

In a population of Chinese women, Wang et al. (2012a) found that the *p21* rs1801270A allele, in addition to an a three SNP haplotype formed by linkage disequilibrium (LD) with the 3'UTR rs1059234 T allele and another *p21* SNP (rs3176352G), were associated with a protective effect with regards to the development of cervical cancer. rs1801270 has also been associated with increased risk of endometrial cancer in Korean populations (Roh et al., 2004). Li et al. (2005) found rs1801270 in addition to rs1059234 to be associated with susceptibility to squamous cell carcinoma of the head and neck. Interestingly, these two *p21* variants have been found to be significantly underrepresented in Italian centenarians (p=0.020 for rs1801270; p=0.026 for rs1059234), suggesting that these SNPs may have a negative effect on longevity (Gravina et al., 2009). Chen et al. (2002) found rs1801270 to be associated with a higher risk of bladder cancer in Taiwanese patients. In Caucasian populations, the rs1801270 polymorphism was found to be significantly linked ($p<0.001$) to an overall

increased cancer risk (Liu et al., 2011). No studies, to our current knowledge, have been conducted to determine whether two these reported *p21* SNPs influence HIV-1 susceptibility and/or disease progression.

More recently, de Pablo et al., (2015) showed a link between possession of a 5'UTR *p21* rs733590 SNP and/or the HLA-B*2705 allele (a well known protective HLA allele found in Caucasians) and higher relative p21 expression in HIV-1 infected individuals. The HLA-B*2705 allele is rarely found in African populations, and the prevalence of the rs733590 SNP in the Black South African population has not been determined. Furthermore, this SNP has not been investigated for its effect on p21 expression in a sub-Saharan African population.

1.7.3 Regulation of p21

Regulation of a gene occurs at all levels during the process of converting DNA into a protein product. The core promoter is generally defined to be the DNA region that directs the initiation of transcription by RNA polymerase II (reviewed by Juven-Gershon et al., 2008). Regulatory elements are contained within the promoter sequence and at the transcription start site of a gene. A regulatory sequence acts to modulate the expression of genes. Promoters usually contain a TATA box (named for its core DNA sequence of 5'-TATAAA-3'), a transcription factor II B (TFIIB) recognition site, an initiator, and the downstream core promoter element (Butler and Kadonaga, 2002).

Regulation can also occur after transcription in the 3'UTR of the mRNA, resulting in differing mRNA stability and having an effect on translation (Merritt et al., 2008). Regulatory elements within the 3'UTR are mostly expected to function post-transcriptionally at the mRNA level, but they can also function at the DNA level as distal enhancers to control transcription (Merritt et al., 2008).

p21 is an unstable protein when it is found within actively dividing cells and has a relatively short half-life of 20-60 minutes. FK506-binding protein like (FKBPL), also known as WISp39, is a tetratricopeptide repeat (TPR) protein that works to recruit heat shock protein (Hsp) 90 to p21. FKBPL works to protect newly synthesised p21 from proteasomal degradation (Jascur et al., 2005). Cells that are depleted of this protein are unable to upregulate p21 when DNA damage is detected. It was found that SNPs in the C-terminal TPR domain of WISp39 cause a loss of interaction between WISp39 and Hsp90, resulting in the inability to stabilize p21 (Jascur et al., 2005). Other variants of p21, including phospho-site-

deficient or phospho-site-mimicking mutants, have also been shown to regulate p21 stability by either affecting interaction with p21 binding proteins or by moving p21 into the cytoplasm (Jascur et al., 2005). This provides evidence that transcriptional control alone may not be sufficient to upregulate p21 when there is no p21 stabilisation occurring in the cell (Jascur et al., 2005).

Viral proteins are also able to influence post-transcriptional control of p21, resulting in an effect on cellular proliferation (reviewed by Abbas and Dutta, 2009). E6 is a human papilloma virus (HPV) protein that has been shown to downregulate p21 independently of p53 (Burkhart et al., 1999, Fan et al., 2005). It was also found that the core protein in hepatitis C virus has an inhibitory effect on p21 post-transcriptionally, resulting in the activation of CDK2 and consequent tumorigenesis (Yoshida et al., 2001). This post-transcriptional regulation of p21 by different viruses provides some insight into how viruses may regulate cell cycle progression and apoptosis. As to precisely how these viral proteins mediate regulation of p21 still needs to be elucidated.

1.7.4 miRNA regulation of p21

MicroRNAs (miRNAs) are short RNA molecules of +/- 22 nucleotides that bind post-transcriptionally to the 3'UTR of target mRNAs, thereby regulating gene expression by silencing translation and/or degrading the targeted transcript (Lu et al., 2005).

miRNAs in the miR-106b family have been shown to directly target p21, reducing *p21* mRNA levels by 38% and p21 protein levels by 46%, and anti-miR-106b increasing p21 protein levels by 53% (Ivanovska et al., 2008). The coding sequences for miR-106b miRNAs are located in intron 13 of the mini-chromosome maintenance 7 (MCM7) gene. A SNP, rs999885, has been reported in the promoter area of MCM7, causing an A to G base change and having an effect on transcription of miR-106b. Like rs1801270 and rs1059234, this SNP has been implicated in cancer progression, with one study showing that the variant genotypes of rs999885 were associated with a significantly decreased risk of death for intermediate or advanced hepatocellular carcinoma in a Chinese population (Qi et al., 2014). This SNP has not, to our knowledge, been studied in the context of HIV-1 control. Given that increased expression levels of *p21* mRNA have been found in ECs, thereby suggesting a role for p21 on HIV-1 control, and that miR-106b has been shown to directly target the *p21* 3'UTR, the rs999885 SNP could be playing an indirect role on p21 expression and consequently, HIV-1

control. The frequency and representation of this SNP in different populations are shown in Table 1.4.

Table 1.4 – Frequency and representation of the miR-106b rs999885 SNP

SNP	Nucleotide Change	Minor Allele Frequency (sample number)		
		Caucasian	African	Asian
rs999885	C/T (REV)	0.549 (N=224)	0.243 (N=218)	0.831 (N=172)
Frequency and representation of SNP genotypes in sub-Saharan Africans (N=220)				
SNP	Mt/Mt		WT/Mt	WT/WT
rs999885	0.083		0.321	0.596

Source: NCBI dbSNP database. REV: Reverse

Multiple miRNAs, specifically those in the miR-106b family (Ivanovska et al., 2008), have the ability to downregulate p21, which is helpful in understanding p21 function. As well as being associated with the regulation of HIV-1, p21 has been implicated in other diseases. Downregulation of p21 results in varied disease phenotypes, including increased spontaneous tumours and lupus (Arias et al., 2007). Additionally, downregulation of p21 was shown to result in an increased number of stem cells under normal cellular conditions i.e. when cells are not under stress (Cheng et al., 2000, Stier et al., 2003, Zhang et al., 2007), having potential consequences on the efficiency of gene therapy. miRNA-mediated p21 downregulation in HIV-1 infected, activated CD4+ T cells resulted in increased levels of p24 capsid protein (Chen et al., 2011).

These miRNA studies show that p21 has serious implications on multiple diseases including AIDS, and therefore strategies to modulate p21 levels, or levels of other proteins upstream or downstream in the biochemical pathway, could be effective in modifying multiple disease phenotypes.

1.8 Study rationale

To our knowledge, nothing is currently known about the genetic variability of the *p21* gene in the Black South African population, and the role of p21 in elite control as well as HIV-1 related control warrants investigation. Given that p21 expression may play a role in HIV-1 control, SNPs and other variations within two areas of the *p21* gene that are expected to influence expression, namely the 5'UTR and 3'UTR regions, were characterised. In addition, the possible influence of a SNP found in the miR-106b miRNA and the consequent effect on p21 expression was assessed.

Given that Chen et al. (2011) found p21 to be significantly upregulated in HIV-1 controllers in their cohort, it is important to determine if this is true for Black South Africans, thereby determining the potential importance in our local setting. The study of this gene through comparison of different clinical phenotypes of HIV-1 infected individuals will provide important insights into pathways involved in HIV-1 control that could be manipulated to achieve functional cure for HIV-1 infected individuals.

1.8.1 Aim of Study

The overall purpose of this study is to explore the role of p21 (WAF1/Cip1) expression in the context of host ability to control HIV-1 infection. The approach taken involves the characterisation of genetic variation of the *p21* gene and quantifying *p21* transcription in relevant clinical samples.

1.8.2 Specific objectives

1. To design real-time PCR assays in order to detect known *p21* rs1801270 and rs1059234 SNPs, in addition to the miR-106b rs999885 SNP, as well as any novel SNPs identified from sequencing, and describe their frequencies in a larger sample set of Black South African ECs, LTNPs, progressors and HIV-1 negative individuals.
2. In a smaller subset of these individuals, to sequence a region of the *p21* 5'UTR and 3'UTR in order to detect novel SNPs that may be implicated in modulating p21 expression and HIV-1 control.
3. To compare levels of *p21* mRNA expression in a subset of Black South African ECs, LTNPs, progressors and HIV-1 negative individuals.

4. Based on the collective findings, to establish a genetic signature that distinguishes individuals who inherently are high p21-expressors (with predicted good-excellent control) from those who are low p21-expressors (HIV-1 progressors).

2 MATERIALS AND METHODS

2.1 Populations and study samples

2.1.1 Genetic characterization of *p21*

Samples used for the genetic characterisation component of the study included Black South African healthy controls (HCs), elite controllers (ECs), viraemic controllers (VCs), high viral load long term non-progressors (HVL LTNPs) and progressors. Stored DNA samples were available for 72 HCs, 11 ECs, 11 HVL LTNPs, 30 VCs and 74 progressors. A summary of the characteristics of the various cohorts are shown in Table 2.1.

In this study, ECs are defined as having viral loads of <50 RNA copies/ml, with some ECs having long term data. VCs are defined as having documented CD4+ T cell counts of >500 cells/µl with viral loads <2000 RNA copies/ml. HVL LTNPs have CD4+ T cell counts >500 cells/µl for >7 years with viral loads exceeding 10 000 RNA copies/ml. Progressors are defined as having declining CD4+ T cell counts to <350 cells/µl with viral loads >10 000 RNA copies/ml, and who require initiation of antiretroviral therapy.

HIV-1 RNA levels were quantitated using the COBAS® AmpliPrep/COBAS® Taqman® HIV-1 Test, v2.0 ultrasensitive tests (<20 RNA copies/ml) (Roche Diagnostic Systems, Inc, New Jersey, USA) and CD4+ T cell counts were determined using FACSCount System from Becton Dickinson (San Jose, CA).

Table 2.1 – Characteristics of HIV-1 infected study participants in the clinical phenotype groups studied

Cohort	n	Gender (% female)	Age (years) (Mean and range)	CD4 Count (cells/µl) (Median and IQR)	Viral Load (HIV RNA copies/ml) (Median and IQR)
Progressors	74	83.8	42.5 (28-71)	177 (146 – 210)	38 444 (19 853 – 103 042)
VCs	30	90.0	36 (19-46)	651 (555 – 832)*	495 (327 - 965)**
ECs	11	81.8	44 (19–54)	853 (718 - 1022)	<40
HVL LTNPs	11	81.8	41 (31-51)	663 (635 – 749)	54 375 (13 415† – 77 820)

*Two individuals had CD4+ T cell counts <350 cell/µl which at the time was the cutoff for starting ART (lab ID: TG1 and TG9). These individuals were included due to their status as LTNPs (9 years since diagnosis for both).
**One individual (TG9) had a VL>2 000 but was included due to length of infection in the absence of ART (9 years). The next highest VL was 1 445 HIV RNA copies/ml.
†Two individuals, Pru2 and NP26, while considered LTNPs (>7 years infection), had VLs below 10 000 at time of enrolment. However, subsequent to enrolment into this study, VLs for these individuals increased dramatically without a drop in CD4+ T cell counts.

2.1.2 *p21* mRNA expression analysis

For mRNA studies, a selection of existing ECs and LTNPs and healthy controls were recalled to consent for additional blood samples, and a set of new HIV-1 progressors (n=12) were identified and recruited for blood draws. In addition, some newly identified HICs were also included (n=5). Three additional recruited individuals (SHP002, SHP003, SHP010) could not be classified as controllers based on high viral loads (>2000 RNA copies/ml) and a lack of long-term data that would define them as HVL LTNPs, and were therefore excluded from analyses when analysing data stratified according to phenotypic groupings of HIV-1. These three individuals were, however, included in correlation analyses where p21 expression was analysed in terms of CD4+ T cell counts and VL.

To establish the relationships between markers of disease severity and quantitation of *p21* mRNA, participants were grouped according to VL and CD4+ T cell counts at the time of blood draw rather than predefined phenotypic groupings used for earlier genetic studies i.e. a participant defined as an EC (viral loads of <50 RNA copies/ml) with a VL of 1000 RNA copies/ml at the time of blood draw for the mRNA expression study was defined as a VC (Table 2.2).

Informed consent was obtained from all individuals participating in this study. Ethics approval for the greater LTNP and EC study has been obtained from the Human Research Ethics Committee at the University of the Witwatersrand, Johannesburg under clearance certificate M140926 (Prof. C. T. Tiemessen). This particular study has also obtained ethics approval under clearance certificate M140996 from the Human Research Ethics Committee at the University of the Witwatersrand.

Table 2.2 – Characteristics of study participants in the *p21* mRNA expression aspect of the study

Sample ID	Gender	Age (2015)	HIV-1 Viral Load (RNA copies/ml)	CD4+ T Cell Count (cells/µl)	Previously assigned clinical phenotype	Assigned clinical phenotype for expression analysis
SHP001	F	36	141	1344	-	VC
SHP002	F	35	8070	557	-	Unclassified*
SHP003	F	41	32133	752	-	Unclassified*
SHP004	F	39	275106	165	-	Progressor
SHP005	F	34	40370	88	-	Progressor
SHP006	F	40	49	641	-	EC
SHP007	M	31	894183	37	-	Progressor
SHP008	M	35	402192	222	-	Progressor
SHP009	M	39	25969000	10	-	Progressor
SHP010	F	47	4732	1147	-	Unclassified*
SHP011	F	50	278566	121	-	Progressor
SHP012	M	55	2203257	44	-	Progressor
SHP013	F	27	292	1475	-	VC
SHP014	F	20	75634	115	-	Progressor
SHP015	M	41	7676	163	-	Progressor
SHP016	F	37	47207	159	-	Progressor
SHP017	F	37	5577	165	-	Progressor
SHP018	F	53	49	640	-	EC
SHP020	F	26	27210	179	-	Progressor
SM0001	M	49	20	922	-	EC
Toga11	F	39	138	561	EC	EC
Toga4	M	42	212	605	VC	VC
Pru-048	F	58	3349	924	EC	VC
Pru-058	F	40	9897	462	VC	HVL LTNP
Pru-052	M	47	9004	465	HVL LTNP	HVL LTNP
NP1010	F	55	65	1559	EC	EC
NP1029	M	47	105	772	EC	EC
NP1035	F	42	72	778	EC	EC
NP1044	F	47	95	718	EC	EC
NP1055	F	22	155	943	EC	EC
NP1068	F	46	49	760	EC	EC
NP1004	F	44	2149	465	VC	VC
NP1016	F	43	560	701	VC	VC
NP1027	F	39	927	908	VC	VC
NP1065	F	38	1288	609	VC	VC
NP1036	F	54	11967	709	HVL LTNP	HVL LTNP
NP1069	F	37	3771	536	HVL LTNP	VC
NP1024	F	36	39945	323	HVL LTNP	HVL LTNP
NP1005	F	39	3539	747	VC	VC
NP1063	F	51	46	712	EC	EC
BC1	F	23	-	-	-	HC
BC2	F	25	-	-	-	HC
BC3	M	32	-	-	-	HC
BC4	M	36	-	-	-	HC
BC5	F	NA	-	-	-	HC
BC6	M	30	-	-	-	HC
BC7	F	35	-	-	-	HC
BC8	F	NA	-	-	-	HC
BC9	F	32	-	-	-	HC
BC10	F	NA	-	-	-	HC

HC = healthy control; EC = elite controller; VC = viraemic controller; HVL LTNP = high viral load long term non-progressor. *These three individuals were included in correlation analyses where p21 expression was analysed in terms of CD4+ T cell counts and viral load, however were excluded from analyses when analyzing data in terms of control. - Not applicable, NA = not available

2.2 Potential genetic biomarkers for differential p21 expression

Figure 2.1 shows a schematic representation of the *p21* gene with the regions amplified and sequenced as well as the position of the two SNPs (red dots) that were genotyped with respect to the characterisation aspect of the study. The additional *p21* 5'UTR SNP that was genotyped as part of the expression aspect of the study is also shown (blue dot).

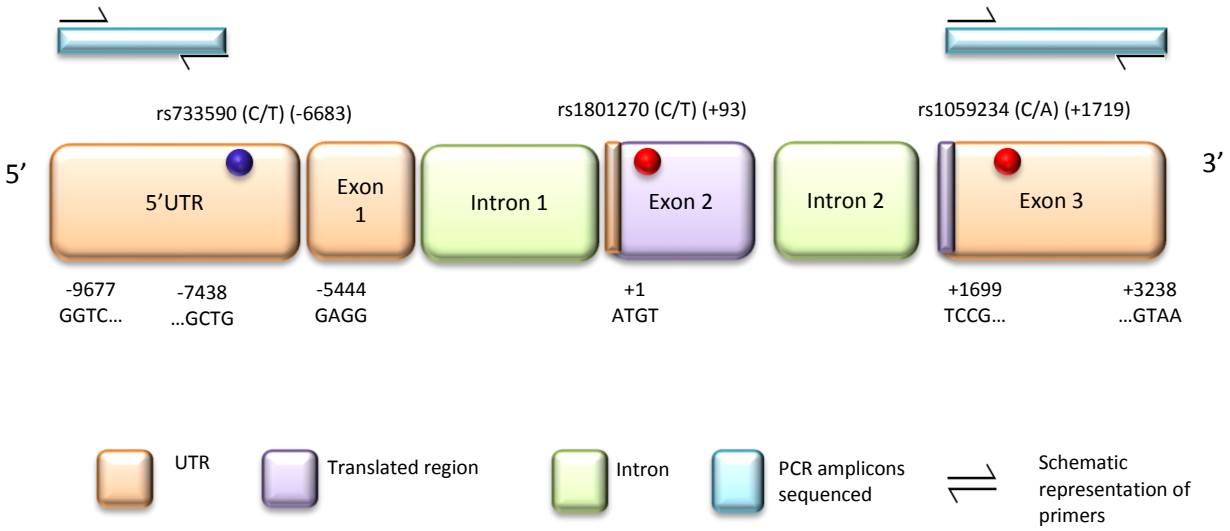


Figure 2.1 - Schematic of the *p21* gene showing an overview of which regions and variations were characterised in the *p21* gene. The regions that were sequenced are shown as blue bars and the three *p21* SNPs genotyped are shown as red dots (rs1801270 and rs1059234) and a blue dot (rs733590) with their position in the *p21* gene shown in brackets

2.2.1 Standard Polymerase Chain Reaction (PCR) Amplification

Genomic DNA that was used in the amplifications for 2.2.3 was previously extracted from either whole blood or buffy coats of patients using the QIAamp DNA Blood Mini Kit (Qiagen). Each region of interest was PCR amplified with predesigned primers designed using PrimerW software and using the Expand High Fidelity PCR System (Roche, Mannheim, Germany) with cycling conditions carried out according to manufacturer's instructions. The mastermix used for all PCR reactions contained a 300 nM final concentration of each primer, 200 µM final concentration of dNTPs and 1.5 mM final concentration of $MgCl_2$. The resulting PCR products were electrophoresed on 1% agarose gels for larger sized products (>500 bp) and 2% agarose gels for smaller fragments (<500 bp) with Fermentas Middle Range Molecular Weight markers (Thermo Fisher Scientific, Massachusettes, USA) used for size referencing. The gels were run for 30-40 minutes at 100V. The PCR amplified products were then purified using either the MSB Spin PCRapace kit (STRATEC Molecular, Berlin-Buch, Germany) or Agencourt Ampure XP (Beckman Coulter, Missouri, USA) magnetic separation kit according to manufacturer's instructions, in preparation for downstream applications.

2.2.2 Sanger sequencing

Sequencing reactions were set up in 96-well plates using BigDye Terminator version 3.1 Cycle Sequencing Chemistries (Applied Biosystems, Foster City, CA, USA). Purified PCR amplicons were used as the DNA template.

After cycling, the sequencing reactions were purified using standard ethanol-sodium acetate precipitation. Briefly, 35 µl of an ethanol-sodium acetate solution was added to the sequencing reactions (10 µl) and the plate was centrifuged at 448g for 30 minutes. The plate was then placed upside down on absorbent paper and centrifuged at 3g for 1 minute. Subsequently, 70% ethanol (50 µl) was added to each well and the plate was centrifuged at 448g for 5 minutes. Again, the plate was placed upside down in the same manner and centrifuged at 3g for 1 minute. The plate was then dried for 3 minutes at 63°C on a GeneAMP PCR System 9700 (Applied Biosystems, Foster City, CA, USA). Dried sequencing pellets were resuspended in 10 µl of HiDi Formamide (Life Technologies, California, USA). The resuspended sequenced fragments were subsequently heated at 95°C for 2 minutes and then electrophoresed on an automated 3100 Genetic Analyser (Applied

Biosystems) using Pop6 polymer and a 36cm capillary array (Life Technologies, California, USA), according to manufacturer's instructions.

Resulting sequence chromatograms were analysed using Sequencher software version 5.1 (Gene Codes Corporation, Ann Arbor, MI, USA). Imported sequences were aligned to a reference sequence obtained from the NCBI database (NCBI ref seq NC_000006.12) and SNPs and indels identified were recorded for downstream analysis. A summary of the sequencing process from DNA extraction to analysis of sequences is shown in Figure 2.2.

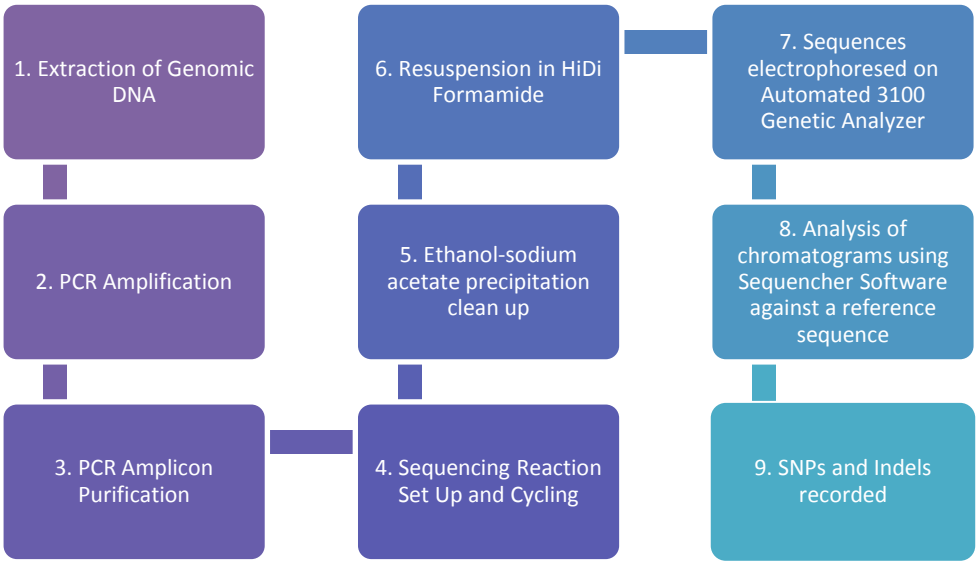


Figure 2.2 – Flow diagram showing steps involved in the sequencing of PCR amplicons from genomic DNA extraction to analysis

2.2.3 SNP Genotyping

A number of different methods were employed in order to genotype select SNPs of interest.

2.2.3.1 Real-time C_T Shift Assay

A real-time cycle threshold (C_T) shift PCR assay was one of the methods used to obtain genotypic data for specific SNPs. It is an assay that uses cycle threshold shifts as a result of mispriming 3' end primer nucleotides to differentiate between the different genotypes. In order to develop the C_T shift PCR assays, it was required to have individuals of known genotype at the three SNP positions (rs1801270, rs1059234 and rs999885 SNPs), determined

by amplifying and sequencing the region harbouring the respective SNPs from control samples. Generally, twenty control samples were PCR amplified and sequenced in order to successfully detect all three possible genotypes. The assays were optimized using two known homozygous WT individuals, two known heterozygous individuals and two known homozygous Mt individuals for each of the SNPs. The PCR primers were designed using PrimerW software. Genomic DNA from the newly recruited individuals [used for the p21 expression component (see 2.3)] was extracted from whole blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), according to manufacturer's instructions, and were genotyped using the same three designed C_T shift assays.

The C_T shift PCR assays were designed using SYBR Green to detect each SNP using allele-specific PCR, with two primers used; one specific for the WT allele and one specific for the Mt allele of each SNP, and one common primer, either in the forward or reverse orientation, depending on the orientation of the allele-specific primers. Thus, for each sample, two reactions were conducted, one with the major allele (WT) primer and one with the minor allele (Mt) primer. Reactions (final volume of 10 μl) were set up in 96-well plates, with each reaction containing 1× Fermentas SYBR Green PCR Master Mix (Thermo Fisher Scientific, Massachusetts, USA), 10 ρmol of each forward and reverse primer, and ~10-50 ng of genomic DNA as template. The reactions were then run in an Applied Biosystems 7500 Real-Time PCR system. The final run settings used for the C_T shift assays were an initial holding stage at 95 °C for 10 minutes, and a cycling stage consisting of 95 °C for 15 seconds, 60 °C for 40 seconds and 70 °C for 1 minute, for a total of 45 cycles. A no template control (NTC) was included and a melt-curve analysis was performed to ensure the absence of primer-dimers. Allele-specific primers with locked nucleic acid (LNA™) modified 3' ends were utilized as they were found to result in more optimal shift patterns. PCR primers and C_T shift assay primers used are shown in Table 2.3.

An LNA is a modified RNA analog, with the ribose ring "locked" in the ideal conformation for binding. The use of highly thermo-stable LNA oligonucleotides in a primer allows for a shorter primer to be designed due to the resulting increased melting temperature (Tm) of the primer. The use of an LNA allows for a high sensitivity, high affinity primer that binds the 3' end more strongly than non-LNA primers. Therefore, using primers with LNA-modified bases at the 3'end makes accidental mispriming less likely to occur, thereby resulting in larger C_T shifts.

Table 2.3 - PCR primers for amplification of regions harbouring single nucleotide polymorphisms of interest and primers used for the C_T shift assays

	PCR primers		
SNP	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**	**Fragment Length (bp)**
rs1801270	GGCCTTCCTTGTATCTCTGC	CTGAGAGTCTCCAGGTCCAC	443
rs1059234	CTTCCTGTTCTCAGCAGTCG	AACCTCTCATTCAACCGCCT	392
rs999885	AGAGCGACTGGACAGGA	AGCGCTGAGAGTACAGGAG	378
	C_T shift assay primers		
SNP	**Common Primer (5'-3')**	**WT Primer * (5'–3')**	**Mt Primer * (5'-3')**
rs1801270	CGGTGACAAAGTCGAAGTTCCAT (R)	TGGACAGCGAGCAGCTGAG[C] (F)	TGGACAGCGAGCAGCTGAG[A] (F)
rs1059234	CAAACGCCGGCTGATCTC (F)	CTCGCGCTTCCAGGACT[A] (R)	CTCGCGCTTCCAGGACT[G] (R)
rs999885	CAGCCCCAAACTGTAAAG (F)	AGGAGGGTGAGGAAAGA[G] (R)	AGGAGGGTGAGGAAAGA[A] (R)

*LNA modified ends indicated by a square bracket and red colour. Primer orientation indicated in round brackets. F: forward, R: reverse

Homozygous WT individuals had a resulting C_T shift (i.e. difference in cycle threshold between the WT reaction and Mt reaction) due to mispriming of the Mt primer in that respective well, heterozygous individuals resulted in both primers binding optimally, resulting in no C_T shift, while homozygous Mt individuals resulted in a shift similar to the homozygous WT individuals, but with the Mt primer binding optimally. Data were analysed using software supplied with the 7500 Real-Time PCR system and the genotype of each sample was recorded.

2.2.3.2 Restriction Fragment Length Polymorphism (RFLP) for genotyping

Due to multiple indels being present in the promoter region of *p21*, sequencing of the full region was not always possible for certain samples. In order to genotype the rs113266348 and the rs113041051 SNPs residing within this unresolvable area of the promoter, we made use of RFLP analysis. The PflFI restriction enzyme (New England BioLabs, Massachusetts, USA) has a restriction pattern (Figure 2.3A) that allows for distinguishing between the WT and Mt allele of the rs113266348 SNP and the DraIII restriction enzyme (Life Technologies, California, USA) was used to genotype the rs113041051 SNP, with restriction pattern shown in (Figure 2.3B). Because the original sequence did not contain a DraIII restriction site

necessary to distinguish between the two alleles, a mutation was introduced with a PCR primer in order to create a cut site for the WT allele (Figure 2.3C).

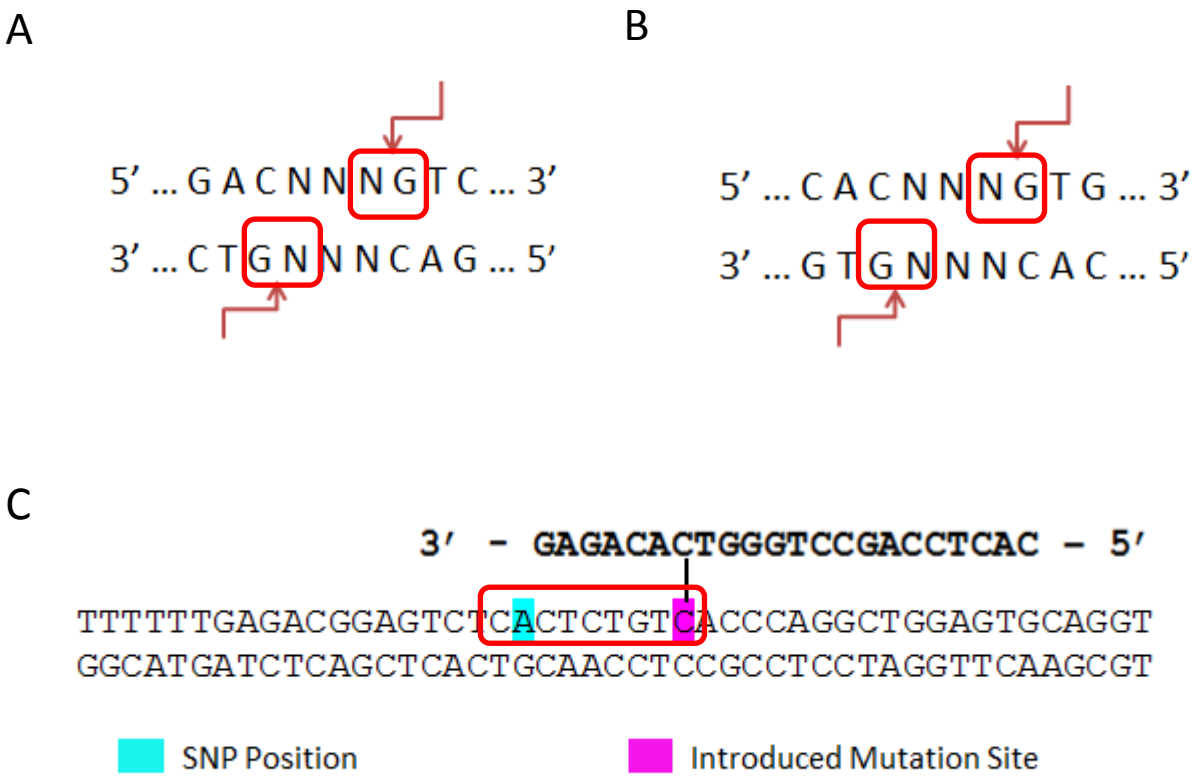


Figure 2.3 - Schematic showing (A) the PflFI restriction enzyme restriction site, (B) the DraIII restriction enzyme restriction site, with the arrows and the red blocks showing where the restriction enzyme cuts and (C) the rs113041051 SNP position and primer with mutation creating a restriction site for the DraIII restriction enzyme, with the red block indicating the restriction site

Primers used for DNA amplification are shown in Table 2.4. A 718 bp region within the promoter region (termed "Promoter Small Amplicon") was PCR amplified (see 2.2.4). For the PflFI digestion, the enzyme was added directly to the "Promoter Small Amplicon" with no additional amplification necessary. Prior to DraIII digestions, a nested PCR was performed in order to amplify a product using the "Promoter Small Amplicon" (Table 2.5) as template.

Table 2.4 – Primers used to amplify DNA to be digested by PflFI restriction enzyme and DraIII restriction enzyme

Restriction Enzyme	F Primer (5' – 3')	R Primer (5' – 3')
PflFI	TTT TTT TGA GAC GGA GTC T	AAA GTT GGA TTT ATT GTT TCT G
DraIII	GGA CAG TTG AAG TTA AAA GGT TTT GA	CAC TCC AGC CTG GGT CA**C** AGA G

Red C: Introduced mutation site

PflFI digestion reactions were set up in 200 µl thin-walled PCR tubes using 15 µl template DNA, 2 µl of 10x NEBuffer, 2 µl H_2O and 1 µl of the PflFI restriction enzyme (10 U/µl) for each sample. The reactions were then incubated using the GeneAmp PCR System 9700 Thermocycler (Applied Biosystems, Foster City, CA, USA) at 37°C for 1 hour followed by a 20 minute inactivation at 65°C. Loading dye was added directly to each PCR tube and the total volume was electrophoresed in a 2% agarose gel in order to visualise digested fragments for genotyping. Samples with a Mt/Mt genotype were expected to result in 2 fragments (161 bp and 679 bp), samples with a WT/Mt genotype were expected to result in 4 fragments (150 bp, 161 bp, 527 bp and 679 bp), and samples with a WT/WT genotype were expected to result in 3 fragments (150 bp, 161 bp and 527 bp) (Figure 2.4).

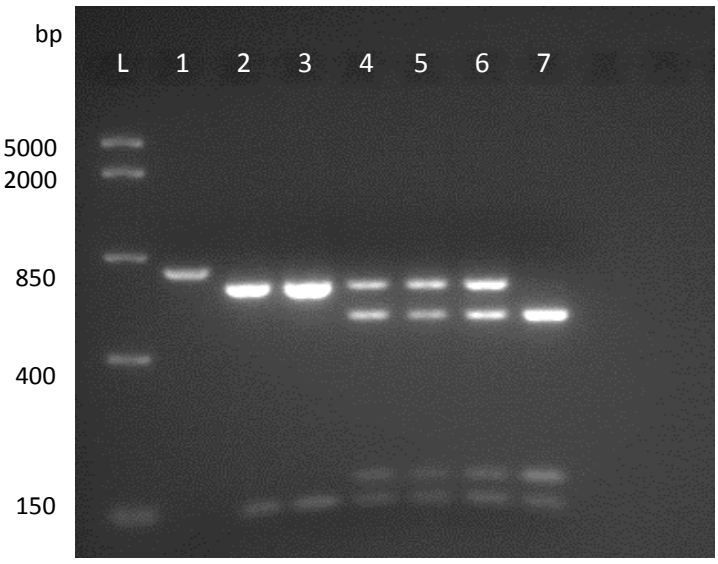


Figure 2.4 – Gel image of PflFI restriction fragments showing samples representing the 3 possible genotypes. L: Ladder with sizes of fragments indicated in base pairs (bp) on the left of image. Lane 1: Unrestricted sample; Lanes 2 and 3: samples showing Mt/Mt genotype restriction pattern; Lanes 4-6: samples showing WT/Mt genotype restriction pattern; Lane 7: sample showing WT/WT genotype restriction pattern

DraIII digestion reactions were set up in 200 µl thin-walled PCR tubes using 10 µl template DNA, 2 µl 6x buffer G, 7 µl H_2O and 1 µl of the DraIII restriction enzyme (10 U/µl) for each sample. The reactions were then incubated using the GeneAmp PCR System 9700 Thermocycler (Applied Biosystems, Foster City, CA, USA) at 37°C for 3 hours. Loading dye was added directly to the sample in the PCR tube and the total volume was then electrophoresed in a 3% superfine resolution (SFR) agarose (Amresco, Ohio, USA) gel in order to genotype samples. SFR agarose was used in order to clearly differentiate between DNA fragments that only had a 15 bp difference in size. The gel was run at 70V for 3 hours. Samples with a Mt/Mt genotype were expected to result in 1 fragment (189 bp – same size as unrestricted sample), samples with a WT/Mt genotype were expected to result in 3 fragments (189 bp, 174 bp and 15 bp) and samples with a WT/WT genotype were expected to result in 2 fragments (174 bp and 15 bp) (Figure 2.5). The 15 bp fragment, due to its small size, could not be visualised on the gel, however this did not impede the genotyping.

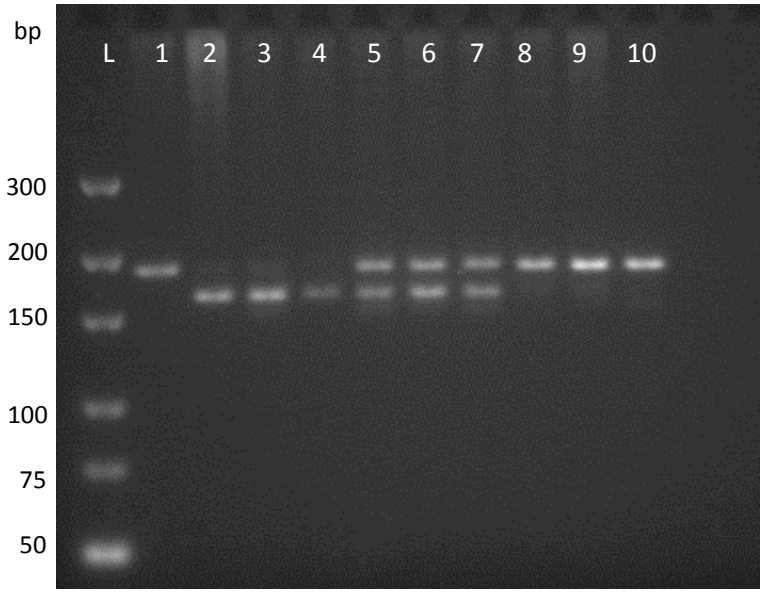


Figure 2.5 – Gel image of DraIII restriction fragments showing samples representing the 3 possible genotypes. L: Ladder with sizes of fragments indicated in base pairs (bp) on the left of image. Lane 1: unrestricted; Lanes 2-4: samples showing WT/WT genotype restriction pattern; Lanes 5-7: samples showing WT/Mt genotype restriction pattern, Lanes 8-10: samples showing Mt/Mt genotype restriction pattern

2.2.4 *p21* 5'UTR sequencing

Since very little literature was available regarding the start site and the region spanning the *p21* promoter, we used data from Gravina et al. (2009), who determined promoter activity using a luciferase assay, to select a region of the 5'UTR for characterisation. We PCR amplified and sequenced a 2248 bp amplicon within the 5' UTR from genomic DNA with primers designed using PrimerW software.

Due to secondary structure within the region as well as multiple long strings of the same nucleotide, a smaller 718 bp region within the promoter region (termed "Promoter Small Amplicon") was also PCR amplified to be sequenced in order to attempt to get clearer sequencing data in this region. Sequencing primers were then designed to span the entire region. Sequencing reactions were set up and analysed as described in 2.2.2, with all variations recorded. PCR and sequencing primers used are shown in Table 2.5.

Table 2.5 – PCR and sequencing primers used to amplify and sequence the *p21* 5'UTR region

PCR Primers			
Amplicon	**Forward PCR primer (5'-3')**	**Reverse PCR primer (5'-3')**	**Fragment Length (bp)**
Promoter	GGTCTTGGATTGAGGAACAG	CAGCACACACTCACACAAGC	2248
Promoter Small Amplicon	TTTTTTTGAGACGGAGTCT	AAAGTTGGATTTATTGTTTCTG	715
Sequencing Primers			
Primer Name		**Primer Sequence (5'-3')**	
P21-Pro-F *		GGTCTTGGATTGAGGAACAG	
P21-Pro-Seq1		TGCTCAGCCATTGTGTCTGC	
P21-Pro-Seq2		CTGGAACAAGCTCTTCGAGG	
P21-Pro-Seq-10		TATTGGCCAGGCTGGT	
P21-Pro-R0		AAGCCAATCAGAGCCACAGC	
P21-Pro-Seq8		GATTGTGCCACTGCTGACTT	
P21-Pro-Seq6		TCCTCACATCCTCCTTCTTC	
P21-Pro-AMP-R **		GAGGCGGAACAAAGATAGAA	
P21-Pro-930-F **		GGAGTCTCACTCTGTCAC	
P21-Pro-930-R **		GCAGATCACAGGGTCAGGAG	

**Primers used for both PCR and sequencing - diluted to 30pmol/ul for PCR and diluted to 3.2pmol/ul for sequencing*

*** Primers used on a secondary smaller amplified region of the promoter*

2.2.5 *p21* 3'UTR sequencing

A 1902 bp amplicon was PCR amplified from genomic DNA with primers designed using PrimerW software. Sequencing primers were designed to span the entire region, with 200-300 bp overlaps. Sequencing reactions were set up and analysed as described previously in 2.2.2. PCR and sequencing primers used are shown in Table 2.6.

Table 2.6 - PCR and sequencing primers used to amplify and sequence the *p21* 3'UTR

Amplicon	Primer Sequence (5'-3')
P21-3UTR-F *	CTTCCTGTTCTCAGCAGT
P21-3UTR-Seq1	ACTAGGCGGTTGAATGAGAG
P21-3UTR-Seq2	CCTGAAGTGAGCACAGCCTA
P21-3UTR-Seq3	CAGCTCAATGGACTGGAAGG
P21-3UTR-Seq4	CAGTAGAGGCTATGGACAGG
P21-3UTR-R *	TTGCAGAGGTGTGACAGT

**Primers used for both PCR and sequencing - diluted to 30 pmol/ul for PCR and diluted to 3.2pmol/ul for sequencing*

2.3 *p21* mRNA expression analysis

2.3.1 Isolation of CD4+ T cells from whole blood

Previous work showed that variable p21 expression, with respect to HIV-1 control, was found to occur in CD4+ T cells. Therefore, CD4+ T cells were isolated from peripheral blood mononuclear cells (PBMCs) to investigate if *p21* mRNA expression, as a marker for p21 protein production, was differentially expressed in relation to markers of disease severity (i.e. CD4+ T cell count and VL) and in terms of differential HIV-1 control (i.e. using a subset of individuals with different clinical phenotypes).

2.3.1.1 Isolation of peripheral blood mononuclear cells (PBMCs)

Approximately 60 ml of whole blood from patients was delivered to the laboratory and was processed on the same day, between 4 and 7 hours after the blood draw. PBMCs were first

isolated utilising a density gradient system. Approximately 20 ml of blood was layered onto 20 ml of Ficoll (Sigma-Aldrich, St. Louis, Missouri, USA) in 50 ml falcon tubes and centrifuged for 30 minutes at 448g. After centrifugation, plasma was removed and stored at -80°C for future studies. The PBMC layer, found between the Ficoll and the plasma layer, was removed and placed into a new 50 ml falcon tube containing phosphate buffered saline (PBS) in a final volume of 50 ml, and washed for a total of 2 washes; the first wash was carried out at 363g for 10 minutes and the second wash at 220g for 10 minutes. After the second wash, the pelleted PBMCs were resuspended in 10-20 ml of PBS, depending on pellet size. A 40 µl aliquot of the cell suspension was diluted 1:1 with 0.4% trypan blue (Sigma, Steinheim, Germany) and cell numbers were then quantified using an Improved Neubauer haemotycytometer. A total of 2 x 10^7 PBMCs were utilized for CD4+ T cell isolation. Due to the progressors having lower CD4+ T cells, 4 x 10^7 PBMCs were used for the CD4+ T cell isolations wherever possible.

2.3.1.2 CD4+ T cell isolation

CD4+ T cells were isolated from PBMCs using MACS® cell separation technology with CD4+ Microbeads and MS columns (Miltenyi Biotec, Germany), and were isolated according to manufacturer's instructions. CD4+ T cells were positively isolated, i.e. the magnetic beads were coated with CD4-specific antibodies and hence bound the CD4+ T cells remaining in the column due to the column being mounted on a magnet, with all other cell types released from the column during the washing steps, using a buffer containing PBS, pH7.2, 0.5% bovine serum albumin (BSA) and 2 mM ethylene-diamine-tetraacetic acid (EDTA). The CD4+ T cells were subsequently eluted from the column after removal from the magnet using the same buffer. The resulting suspension was centrifuged for 10 minutes at 300g and the supernatant was removed. The CD4+ T cell pellet was then resuspended in 150 µl of RNAlater® solution (Life Technologies, California, USA) and stored overnight at 4°C and then at -80°C until needed for mRNA expression analysis.

2.3.1.3 Flow cytometry to determine CD4+ T cell purity

To test the separation efficiency of the MACS® cell separation system as well as determining the purity of our isolated CD4+ T cells, CD4+ T cells were isolated from PBMCs from a control sample as described in 2.3.1.2. The resulting CD4+ T cell pellet was resuspended in the PBS-BSA-EDTA buffer used in the isolation step. The resuspended CD4+ T cell pellet was stained with CD4 FITC and CD3 APC antibodies and incubated for 15 minutes at 4 °C.

This was followed by two washes using the CD4+ T cell isolation buffer. The sample was acquired immediately on a BD FACSCalibur™.

2.3.2 RNA extraction and cDNA synthesis

RNA was extracted from the isolated CD4+ T cells in RNAlater using the mirVana miRNA Isolation Kit (Ambion®, Life Technologies, California, USA), which allows for extraction of mRNA as well as miRNA, according to manufacturer's instructions. Although we did not use the miRNA in this study, future work is planned in which the miRNA component will be studied. Since we had stored our CD4+ T cells in RNAlater®, we needed to pellet out the cells prior to RNA extraction. Ice cold PBS (150 µl) was added to the RNAlater®-cell suspension and the tube was then centrifuged for 6 minutes at 9000g. After extraction, RNA quality was assessed using the Agilent RNA 6000 Nano Kit and the Agilent 2100 Bioanalyzer system. Only RNA with a RNA Integrity Number (RIN) larger than 7 was subsequently used for expression analysis.

RNA concentrations were measured using the Bioanalyzer system (Agilent Technologies, California, USA), the Qubit® 2.0 system for RNA (High Sensitivity) (ThermoFisher Scientific, Massachusetts, USA) and a NanoDrop spectrophotometer (ThermoFisher Scientific, Massachusetts, USA). Measurements were taken twice for each method of quantification in order to determine which method gave the most consistent results. The NanoDrop gave higher concentration values when compared to the other two methods, which could possibly be attributed to the presence of genomic DNA in the RNA sample, since most, if not all RNA extractions cannot exclude all genomic DNA without making use of DNase digestions. We opted not to include a DNase digestion since our quantification assay probes all spanned exon-exon boundaries which ensured that no contaminating genomic DNA in the cDNA would be amplified. Since the Bioanalyzer resulted in the most consistent measurements between repeats, we took an average of the two measurements from the Bioanalyzer system as the final RNA concentration. The RNA was standardised to the sample with the lowest concentration. All samples were diluted to 10 ng/µl concentration in preparation for cDNA synthesis.

cDNA was synthesized using the Invitrogen Superscript III first strand synthesis system (Thermofisher Scientific, Massachusetts, USA), using both oligo dT primers and random hexamers, on an Applied Biosystems 7500 Real-Time PCR system, according to

manufacturer's instructions. Two synthesis reactions were performed for each sample, with 7 µl of RNA used in each reaction. cDNA was stored at -20°C prior to use in the expression analysis.

2.3.3 Real-time relative gene quantification using PCR

Synthesized cDNA was used as the template for gene-specific amplification using a predesigned gene expression TaqMan assay for CDKN1A/p21 from Applied Biosystems (Life Technologies: Hs00355782_m1). In addition, we also amplified all samples using a predesigned Taqman assay for CCR5 (Life Technologies: Hs00152917_m1) as a marker of cell activation. Two endogenous controls were employed for normalisation: ribosomal protein large, PO (RPLPO) (Life Technologies: Hs04189669_g1), and beta-actin (ACTB) (Life Technologies: Hs01060665_g1). As mentioned, all purchased assays were designed with probes spanning exons, thereby avoiding amplification of any carryover genomic DNA.

Reactions (10 µl final volume) were performed in triplicate for each sample and were set up in 96-well plates. For each sample and respective Taqman assay, a mix was made for 4 reactions (one extra reaction included), containing 2 µl of the respective 20x Taqman Gene Expression Assay, 20 µl of Taqman Gene Expression Mastermix (Life Technologies), 4 µl of cDNA and 14 µl of nuclease-free water (Ambion). 10 µl from the mix was added to each well. The reactions were then run on an Applied Biosystems 7500 Real-Time PCR system. The amplification settings included an initial holding stage at 95 °C for 10 minutes and a cycling stage consisting of 95 °C for 15 seconds and 60 °C for 40 seconds. A no template control (NTC) was included for each assay i.e. CDKN1A/p21, CCR5, RPLPO and ACTB.

2.4 *p21* 5'UTR rs733590 SNP genotyping

The group of individuals used for the *p21* mRNA expression experiment (n=50) were also genotyped for the *p21* 5'UTR rs733590 SNP using a C_T shift PCR assay as described in 2.2.3.1. Since the SNP occurs in high frequencies in both Caucasian and Black South African populations, with differing minor alleles, the assay was optimised using a subset of Caucasian and African individuals in order to detect all three genotypes. The primers were designed using PrimerW software and utilized LNA™ modified 3' ends.

The C_T shift common primer used was 5'-CGAGGTCAGCTGCGTTAGAG-3' (F), with the C_T shift WT primer used being 5'-TTCATCTGAACAGAAATCCCACT[G]-3' (R) and the

C_T shift Mt primer used being 5'-TTCATCTGAACAGAAATCCCACT[A]-3' (R). The final amplification settings used for the C_T shift assay was an initial holding stage at 95 °C for 10 minutes, and a cycling stage consisting of 95 °C for 15 seconds, 60 °C for 50 seconds and 72 °C for 1 minute, for a total of 45 cycles.

2.5 Analysis and Statistics

Fisher exact tests were performed using VassarStats (*http://vassarstats.net/odds2x2.html)* to calculate the statistical significance and 95% confidence intervals (CI) of odds ratios (OR) for genotype and allele frequency comparisons. Haploview software (Broad Institute of MIT and Harvard) was used to determine the linkage disequilibrium (LD) between polymorphisms, as well as Hardy-Weinberg equilibrium.

Relative gene expression was calculated using the $2^{\Delta CT}$ method, subtracting the average target gene C_T shift from the average reference gene C_T shift for each individual to get the ΔC_T value. Mann-Whitney U tests were used to compare non-parametric data (i.e. *p21* mRNA expression), and correlations were calculated using Spearman rank correlation to calculate r values. Analyses were performed using GraphPad Prism version 5.0 for Windows (Graphpad Software, San Diego, CA, USA www.graphpad.com).

For all analyses, two-tailed tests were used and statistical significance was set at $p<0.05$.

3 RESULTS

3.1 Representation of the two *p21* SNPs (rs1801270 and rs1059234) and the miR-106b SNP (rs999885) in study groups

3.1.1 SNP genotyping using a CT Shift Assay

A real-time C_T shift assay was developed as outlined in section 2.2.3.1 in order to genotype 72 HCs, 52 HIV-1 controllers (11 ECs, 30 VCs, 11 HVL LTNPs) and 74 progressors for the rs1801270, rs1059234 and rs999885 SNPs.

The rs1801270 and the rs1059234 SNPs are the two most commonly studied *p21* SNPs, largely in the context of cancer research, with the rs1801270 SNP being found in codon 31 of the *p21* gene and the rs1059234 SNP being found in the 3'UTR of the *p21* gene. The miR-106b miRNAs have been shown to affect regulation of p21 and are hosted in intron 13 of the *MCM7* gene, with the rs999885 SNP being found in the promoter area of *MCM7*.

The C_T shift assays were optimised to result in consistent shifts in C_T values (i.e. the number of cycles differing between the WT and Mt reactions in the homozygous individuals) that were consistent and reproducible for each SNP. For heterozygous individuals (WT/Mt), both PCRs should result in very similar C_T values. Figure 3.1 shows representative shift patterns for select individuals, but are representative of all samples genotyped.

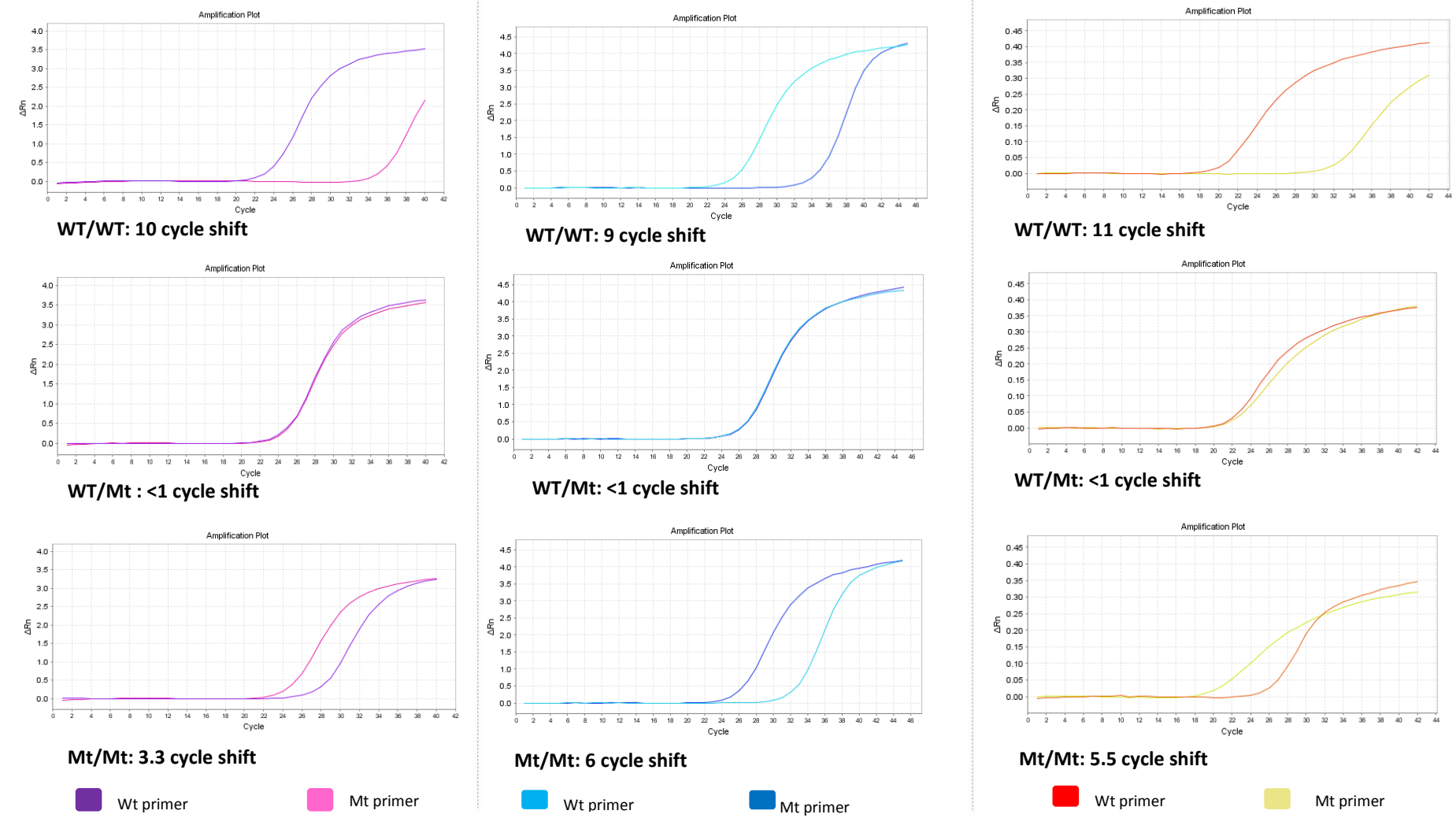


Figure 3.1 – C_T shift patterns for the WT/WT, WT/Mt and Mt/Mt genotypes of the (A) rs1801270 SNP, (B) rs1059234 SNP, and (C) rs999885 SNP. WT/WT represents both alleles being of the ancestral/wild-type, WT/Mt represents one allele being of the ancestral/wild-type and one allele being of the mutated type, and Mt/Mt represents both alleles being of the mutated type. Colours representing the WT primer and the Mt primer are shown at the bottom of the figure.

3.1.2 Test for Hardy-Weinberg Equilibrium

Prior to analysis, the genotypic data generated for the rs1801270, rs1059234 and rs999885 SNP positions were tested for deviation from Hardy-Weinberg equilibrium. There was no significant deviation from Hardy-Weinberg equilibrium for all three SNPs, with the resulting *p* values shown in Table 3.1.

Table 3.1 - Hardy-Weinberg Exact Probability values calculated for the rs1801270, rs1059234 and rs999885 SNPs for the three study groups

	SNP		
	rs1801270	**rs1059234**	**rs999885**
Cohort	**Hardy-Weinberg Exact Probability (Haldane)**		
Healthy Controls	1.00 (n=72)*	1.00 (n=71)*	0.53 (n=70)*
HIV Controllers (n=52)	0.77	0.59	0.85
Progressors (n=74)	0.44	0.53	1.00

*Healthy control sample numbers differed for each SNP as a result of exclusion of some samples due to failure to genotype

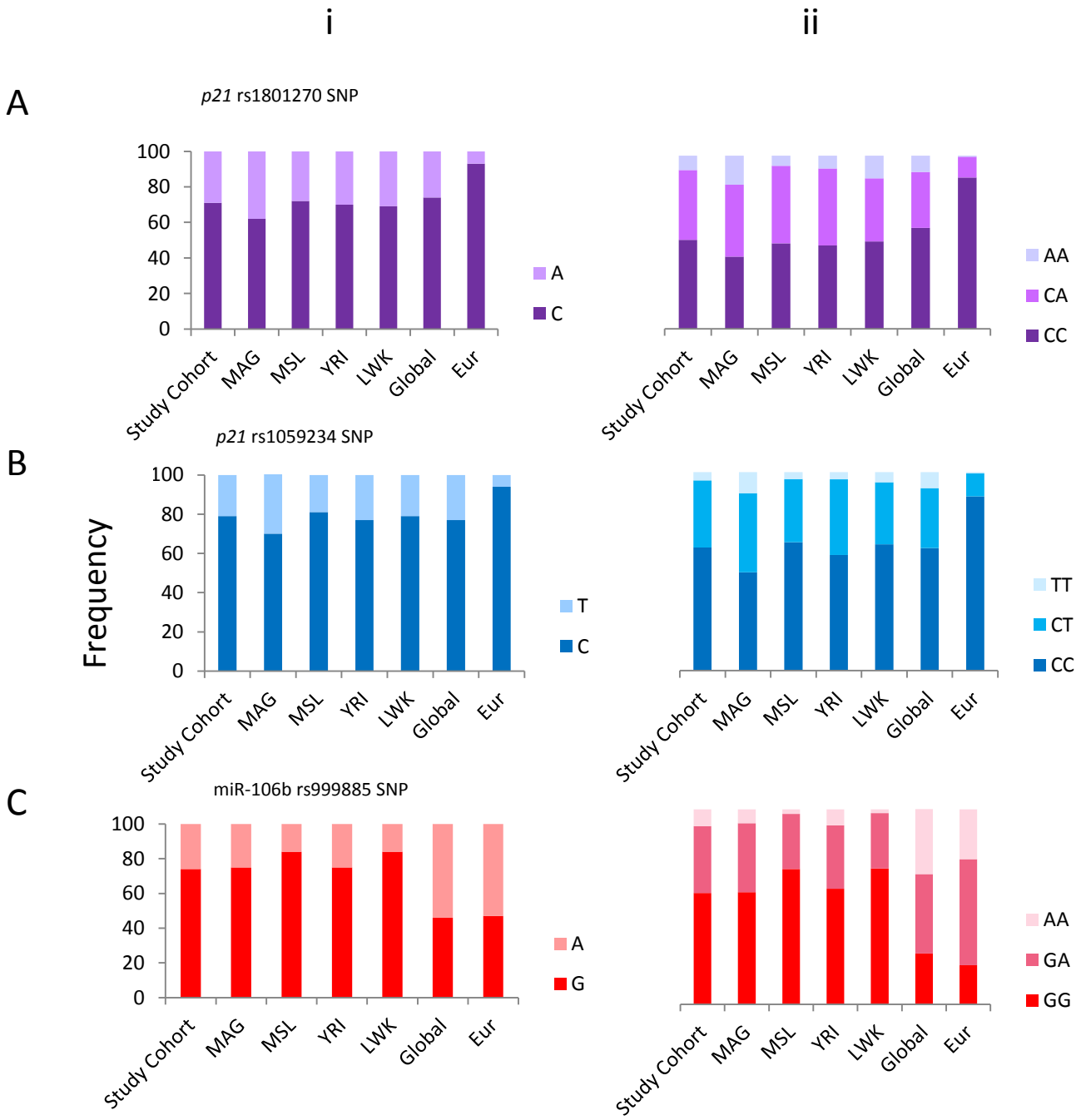
3.1.3 Representation of the rs1801270, rs1059234 and rs999885 SNPs in healthy Black South Africans and in other reference population groups

In order to ascertain how representation of the rs1801270, rs1059234 and rs999885 SNPs differed to other populations of sub-Saharan African ancestry, as well as to populations of European (Caucasian) ancestry, allelic and genotypic frequencies for these SNPs in healthy Black South African individuals were compared to those found in other populations. In addition, the global representation of these variants was also compared. Comparisons are shown in Figure 3.2.

For the two *p21* SNPs, rs1801270 and rs1059234 (Figure 3.2A and Figure 3.2B), Black South Africans compared similarly to four other sub-Saharan African populations (selected from the 1000 Genomes Project: *1000genomes.org*) as well as to the global population data for both allelic and genotypic frequencies. Europe showed marked differences with regards to allelic and genotypic distributions of these two SNPs when compared with the sub-Saharan African populations, although this was not unexpected as Caucasians and Black populations often differ greatly with regards to representation of a number of genetic variants. This again serves

to highlight the importance of studying disease-associated polymorphisms in the context of the specific population group affected. For both the *p21* SNPs, the European population possessed no homozygotes. The sub-Saharan African population that showed the closest representation to our population was that of the Nigerian Yoruba population.

For the miR-106b rs999885 SNP, our study cohort again compared similarly to the sub-Saharan African populations, most closely again with the Nigerian Yoruba population, in both allelic and genotypic representations. Both the global and European allelic frequencies for this SNP varied greatly to that of the African populations, with the ancestral allele (G) (as defined on the NCBI database) showing less representation than the minor A allele (Figure 3.2Ci). With regards to genotypic frequencies, a similar pattern was seen in which the African populations differ greatly to the global and European populations (Figure 3.2Cii).



Population number: Study Cohort: n=72 MAG: n=113 MSL: n=85 YRI: n=108 LWK: n=99 Global: n=2504 Eur: n=503

Figure 3.2 – Bar graphs showing the allelic frequencies (i) and the genotypic frequencies (ii) for the two *p21* SNPS, rs1801270 (A) and rs1059234 (B), and the miR-106b SNP, rs999885 (C) for healthy Black South Africans compared to other populations. MAG: Mandinka in the Gambia; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria; LWK: Luhya in Webuye, Kenya; Eur: European. Data for other populations sourced from *1000genomes.org*

3.1.4 Testing the *p21* rs1801270 and rs1059234 SNPs for linkage disequilibrium

Population representation analysis of the two *p21* SNPs, rs1801270 and rs1059234, suggested that there was linkage between them i.e. the presence of the one SNP was often, but not exclusively, associated with the presence of the second SNP. If two alleles at different loci in a gene or across genes are in linkage disequilibrium (LD), it suggests that they are non-randomly linked and that they may be responsible for producing a cumulative functional effect i.e. may represent a functional haplotype. We thus investigated the LD between the two SNPs.

The LD between the rs1801270 and rs1059234 SNPs was determined using Haploview software. The resulting r^2 value was 0.61, showing a moderate to strong LD, thereby representing a two-SNP intragenic haplotype. Given that the Nigerian Yoruba population showed very similar allelic representation for these two SNPs, we also tested the two SNPs for LD in the Nigerian Yoruba population (genotype data from *1000genomes.org*) with a resulting r^2 value of 0.65, which was very comparable to our population. Therefore, in addition to assessing the effect of the two SNPs separately with respect to representation in clinical phenotype groupings, we also assessed their combined effect i.e. as a haplotype.

3.1.5 Comparison of allelic and genotypic frequencies of the rs1801270, rs1059234 and rs999885 SNPs between healthy individuals, HIV-1 controllers and progressors

The allelic and genotypic frequencies for the rs1801270, rs1059234 and rs999885 SNPs in our study groups are listed in Table 3.2. In addition, the genotypic distributions of the three SNPs in all the groups studied are schematically shown in Figure 3.3.

When comparing the allelic and genotypic frequencies for the two *p21* SNPs, rs1801270 and rs1059234, between HCs and HICs and respective HIC subgroups, as well as when comparing HICs and respective HIC subgroups to progressors, there were no significantly over- or under-represented alleles or genotypes in any of the groups (Table 3.3, Table 3.4, Table 3.5).

Interestingly, however, in the case of the rs1059234 SNP, there was markedly less representation of the minor allele (T) as well as the heterozygous genotype (CT) in ECs compared to progressors (T allele - ECs: 9% vs. progressors: 25%, p=0.11, OR=3.33; CT genotype - ECs: 18% vs. progressors: 42%, p=0.19, OR=3.49). When controllers were grouped and analysed according to VL, the <400 HICs also had markedly less representation

of the CT genotype compared to progressors (<400 HICs: 20% vs. progressors: 42%, p=0.11, OR=2.91). Although these associations were not significant, the high odds ratios may suggest a moderate effect of this SNP, and consequently p21, on control in ECs and in controllers with low VL.

Analysis of the combined effect of the two SNPs i.e. the rs1801270 and the rs1059234 SNPs, as a two-SNP intragenic haplotype did not reveal any significant associations when comparing either HCs to HICs and controller subgroups or HICs and controller subgroups to progressors. Interestingly, when comparing HCs to HICs with VLs >400 RNA copies/ml (>400 HICs), HCs had markedly less representation of the rs1059234 SNP occuring without the rs1801270 SNP (HCs: 0% vs. >400 HICs: 6.25%, p=0.09, OR=0.09). These values are shown in Table 3.6.

With regards to the miR-106b rs999885 SNP, when comparing the >400 HICs to the progressor cohort, the >400 HICs had significantly lower representation of the minor allele of the rs999885 SNP (p=0.04, OR=2.28). In addition, heterozygosity for this SNP was found in a much lower proportion of >400 HICs when compared to progressors (p=0.05; OR=2.56). These results suggest a possible role for the miR-106b miRNA in the control of HIV-1 in individuals with higher viral loads.

When analyzing the genotypic frequencies between groups for the three studied SNPs (Figure 3.3), it is interesting that homozygosity for the minor allele in all three SNPs was absent in both the ECs and HVL LTNPs. However, this is likely due to the rarity of the genotype in addition to the small sample size of both of these groups (n=11 in each).

Although correction for multiple testing, such as the highly conservative Bonferroni correction, is normally used to control for Type I or 'false-positive' results that occur with the repeated use of statistical tests such as in our study, the risk of increasing Type II error in this case was not a primary consideration. Due to the exploratory nature of this study, we have opted to not apply correction for multiple testing and strongly emphasize that any trends or significant associations need to be validated in additional studies.

Table 3.2 – Total number and frequency (%) of rs1801270, rs1059234 and rs999885 SNP alleles and genotypes in healthy controls, HIV-1 controllers and progressors

SNP	Healthy Controls (HCs)	HIV Controllers (HICs)	Progressors	HIV-1 controller sub-groups			HIV-1 controllers stratified according to VL	
				Elite Controllers (ECs)	Viraemic Controllers (VCs)	High Viral Load Controllers (HVL LTNPs)	<400 HICs	>400 HICs
rs1801270	**n=72**	**n=52**	**n=74**	**n=11**	**n=30**	**n=11**	**n=20**	**n=32**
Allele								
C	103 (71.53)	76 (73.08)	106 (71.62)	18 (81.82)	40 (66.67)	18 (81.82)	29 (72.50)	47 (73.44)
A	41 (28.47)	28 (26.92)	42 (28.38)	4 (18.18)	20 (33.33)	4 (18.18)	11 (27.50)	17 (26.56)
Genotype								
CC	37 (51.39)	27 (51.92)	36 (48.65)	7 (63.64)	13 (43.33)	7 (63.64)	10 (50.00)	17 (53.13)
CA	29 (40.28)	22 (42.31)	34 (45.95)	4 (36.36)	14 (46.67)	4 (36.36)	9 (45.00)	13 (40.63)
AA	6 (8.33)	3 (5.77)	4 (5.40)	0 (0.00)	3 (10)	0 (0.00)	1 (5.00)	2 (6.25)
rs1059234	**n=71**	**n=52**	**n=74**	**n=11**	**n=30**	**n=11**	**n=20**	**n=32**
Allele								
C	112 (78.87)	84 (80.77)	111 (75.00)	20 (90.91)	46 (76.67)	18 (81.82)	34 (85.00)	50 (78.13)
T	30 (21.13)	20 (19.23)	37 (25.00)	2 (9.09)	14 (23.33)	4 (18.18)	6 (15.00)	14 (21.87)
Genotype								
CC	44 (61.97)	33 (63.46)	40 (54.05)	9 (81.82)	17 (56.67)	7 (63.64)	15 (75.00)	18 (56.25)
CT	24 (33.80)	18 (34.62)	31 (41.90)	2 (18.18)	12 (40.00)	4 (36.36)	4 (20.00)	14 (43.75)
TT	3 (4.23)	1 (1.92)	3 (4.05)	0 (0.00)	1 (3.33)	0 (0.00)	1 (5.00)	0 (0.00)
rs999885	**n=70**	**n=52**	**n=74**	**n=11**	**n=30**	**n=11**	**n=20**	**n=32**
Allele								
G	104 (74.29)	81 (77.88)	104 (70.27)	17 (77.27)	46 (76.67)	18 (81.82)	27 (67.50)	54 (84.37)
A	36 (25.71)	23 (22.12)	44 (29.73)	5 (22.73)	14 (23.33)	4 (18.18)	13 (32.50)	10 (15.63)
Genotype								
GG	40 (57.14)	31 (59.61)	36 (48.65)	6 (54.55)	18 (60)	7 (63.64)	9 (40.00)	23 (71.87)
GA	24 (34.29)	19 (36.54)	32 (43.24)	5 (45.45)	10 (33.33)	4 (36.36)	11 (55.00)	8 (25.00)
AA	6 (8.57)	2 (3.85)	6 (8.11)	0 (0.0)	2 (6.67)	0 (0.00)	1 (5.00)	1 (3.13)

HCs: Healthy HIV-1-uninfected controls; HICs: HIV-1 controllers; HVL LTNPs: High viral load controllers; VCs: Viraemic controllers; ECs: Elite controllers; <400 HICs: HICs with viral load <400 (RNA copies/ml); >400 HICs: HICs with viral load >400 (RNA copies/ml)

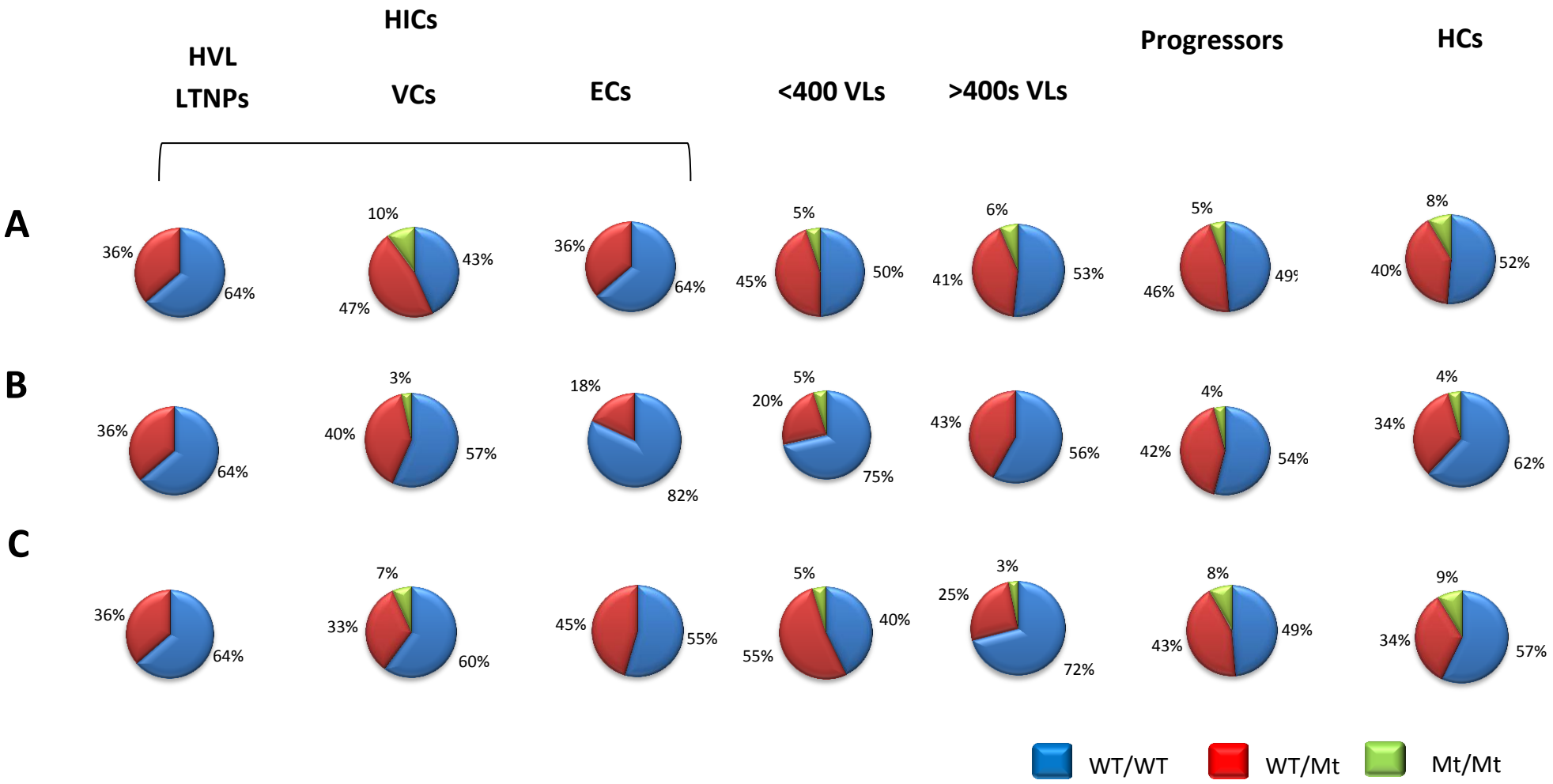


Figure 3.3 – Pie charts showing genotypic representation of the (A) rs1801270 SNP, (B) rs1059234 SNP and (C) rs999885 SNP in HCs, HICs and progressors. WT/WT refers to both alleles being of the ancestral/wild-type, WT/Mt refers to one allele being of the ancestral/wild-type and one allele being of the mutated type, and Mt/Mt refers to both alleles being of the mutated type. HCs – healthy controls, HICs – HIV-1 controllers, HVL LTNPs – high viral load controllers, VCs – viraemic controllers, ECs – elite controllers.

Table 3.3 – Comparison of allelic and genotypic frequencies of the *p21* rs1801270 SNP between both healthy controls and HIV-1 controllers, and HIV-1 controllers and progressors

HCs vs. HICs																		
	HCs vs. HICs			**HCs vs. HVL LTNPs**			**HCs vs. VCs**			**HCs vs. ECs**			**HCs vs. <400 HICs**			**HC vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (C)	0.93	0.53-1.63	0.89	0.56	0.18-1.75	0.44	1.26	0.66-2.40	0.51	0.56	0.18-1.75	0.44	0.95	0.44-2.08	1.00	0.91	0.47-1.76	0.87
Mt (A)	1.08	0.61-1.90	0.89	1.79	0.57-5.61	0.44	0.8	0.42-1.52	0.51	1.79	0.57-5.61	0.44	1.05	0.48-2.30	1.00	1.10	0.57-2.13	0.87
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CA	0.96	0.46-2.02	1.00	1.37	0.37-5.14	0.75	0.73	0.30-1.79	0.5	1.37	0.37-5.14	0.75	0.87	0.31-2.42	0.80	1.02	0.43-2.45	1.00
AA	1.46	0.33-6.36	0.73	∞	NaN-∞	0.58	0.7	0.15-3.22	0.69	∞	NaN-∞	0.59	1.62	0.17-15.07	1.00	1.38	0.25-7.55	1.00
Progressors vs. HICs																		
	Progressors vs. HICs			**Progressors vs. HVL LTNPs**			**Progressors vs. VCs**			**Progressors vs. ECs**			**Progressors vs. <400 HICs**			**Progressors vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (C)	0.93	0.53-1.63	0.89	0.56	0.18-1.76	0.44	1.26	0.66-2.40	0.51	0.56	0.18-1.76	0.44	0.96	0.44-2.09	1.00	0.91	0.47-1.77	0.87
Mt (A)	1.08	0.61-1.89	0.89	1.78	0.57-5.58	0.44	0.79	0.42-1.51	0.51	1.78	0.57-5.58	0.44	1.04	0.48-2.28	1.00	1.09	0.57-2.12	0.87
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CA	1.16	0.56-2.41	0.71	1.65	0.44-6.16	0.53	0.88	0.36-2.13	0.82	1.65	0.44-6.16	0.53	1.05	0.38-2.90	1.00	1.23	0.52-2.92	0.67
AA	1.00	0.21-4.85	1.00	∞	NaN-∞	1.00	0.48	0.09-2.45	0.66	∞	NaN-∞	1.00	1.11	0.11-11.09	1.00	0.94	0.16-5.67	1.00

Table 3.4 - Comparison of allelic and genotypic frequencies of the *p21* rs1059234 SNP between both healthy controls and HIV-1 controllers, and HIV-1 controllers and progressors

HCs vs. HICs																	
	HCs vs. HICs			**HCs vs. HVL LTNPs**			**HCs vs. VCs**			**HCs vs. ECs**			**HCs vs. <400 HICs**			**HC vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (C)	0.89	0.47-1.67	0.75	0.83	0.26-2.64	1.00	1.14	0.55-2.34	0.85	0.37	0.08-1.69	0.25	0.66	0.25-1.72	0.50	1.05	0.51-2.14	1.00
Mt (T)	1.13	0.60-2.12	0.75	1.21	0.38-3.83	1.00	0.88	0.43-1.81	0.85	2.68	0.59-12.11	0.25	1.52	0.58-3.95	0.50	0.96	0.47-1.96	1.00
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.00	0.47-2.14	1.00	0.95	0.25-3.59	1.00	0.77	0.32-1.88	0.65	2.45	0.49-12.29	0.32	2.05	0.61-6.86	0.28	0.70	0.30-1.65	0.51
TT	2.25	0.22-22.62	0.64	∞	NaN-∞	1.00	1.16	0.11-11.93	1.00	∞	NaN-∞	1.00	1.02	0.1-10.59	1.00	∞	NaN-∞	0.55
Progressors vs. HICs																		
	Progressors vs. HICs			**Progressors vs. HVL LTNPs**			**Progressors vs. VCs**			**Progressors vs. ECs**			**Progressors vs. <400 HICs**			**Progressors vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (C)	0.71	0.39-1.32	0.29	0.67	0.21-2.10	0.6	0.92	0.45-1.85	0.86	0.3	0.07-1.35	0.11	0.53	0.21-1.36	0.21	0.84	0.42-1.69	0.73
Mt (T)	1.4	0.76-2.59	0.29	1.5	0.48-4.72	0.6	1.1	0.54-2.22	0.86	3.33	0.74-14.95	0.11	1.89	0.73-4.86	0.21	1.19	0.59-2.40	0.73
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
CT	1.42	0.68-2.98	0.45	1.36	0.36-5.05	0.75	1.1	0.46-2.63	1.00	3.49	0.70-17.31	0.19	2.91	0.88-9.63	**0.11**	0.99	0.43-2.31	1.00
TT	2.48	0.25-24.93	0.63	∞	NaN-∞	1.00	1.28	0.12-13.15	1.00	∞	NaN-∞	1.00	1.13	0.11-11.6	1.00	∞	NaN-∞	0.55

Significant values and values that are trending towards significance are shown in bold and highlighted in blue

Table 3.5 - Comparison of allelic and genotypic frequencies of the miR-106b rs999885 SNP between both healthy controls and HIV-1 controllers, and HIV-1 controllers and progressors

HCs vs. HICs																	
	HCs vs. HICs			**HCs vs. HVL LTNPs**			**HCs vs. VCs**			**HCs vs. ECs**			**HCs vs. <400 HICs**			**HC vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (G)	0.82	0.45-1.50	0.55	0.64	0.20-2.02	0.6	0.88	0.43-1.79	0.86	0.85	0.29-2.47	0.8	1.39	0.65-2.98	0.42	0.54	0.25-1.16	0.15
Mt (A)	1.22	0.67-2.22	0.55	1.56	0.49-4.91	0.6	1.14	0.56-2.31	0.86	1.18	0.41-3.42	0.8	0.72	0.34-1.54	0.42	1.87	0.86-4.05	0.15
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
GA	0.98	0.46-2.10	1.00	1.05	0.28-3.96	1.00	1.08	0.43-2.72	1.00	0.72	0.20-2.62	0.74	0.49	0.18-1.36	0.20	1.73	0.67-4.46	0.36
AA	2.33	0.44-12.32	0.46	∞	NaN-∞	0.58	1.35	0.25-7.35	1.00	∞	NaN-∞	0.60	1.35	0.14-12.64	1.00	3.45	0.39-30.47	0.41
Progressors vs. HICs																		
	Progressors vs. HICs			**Progressors vs. HVL LTNPs**			**Progressors vs. VCs**			**Progressors vs. ECs**			**Progressors vs. <400 HICs**			**Progressors vs. >400 HICs**		
Allele	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
WT (G)	0.67	0.38-1.20	0.19	0.53	0.17-1.64	0.32	0.72	0.36-1.44	0.40	0.7	0.24-2.00	0.62	1.13	0.54-2.41	0.85	0.44	0.20-0.94	**0.04**
Mt (A)	1.49	0.83-2.67	0.19	1.9	0.61-5.95	0.32	1.39	0.69-2.78	0.40	1.44	0.50-4.14	0.62	0.88	0.42-1.86	0.85	2.28	1.07-4.89	**0.04**
Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
GA	1.45	0.69-3.05	0.35	1.56	0.42-5.81	0.54	1.6	0.65-4.00	0.37	1.07	0.30-3.83	1.00	0.73	0.27-1.98	0.61	2.56	1.00-6.51	**0.05**
AA	2.58	0.49-13.73	0.29	∞	NaN-∞	0.57	1.5	0.27-8.19	0.71	∞	NaN-∞	0.58	1.50	0.16-14.08	1.00	3.83	0.43-33.94	0.25

Significant values and values that are trending towards significance are shown in bold and highlighted in blue

Table 3.6 - Comparison of the frequency of the rs1801270 and rs1059234 SNPs when presenting as a haplotype, and for each SNP when found without the other

	HCs vs. HICs																
	HCs vs. HICs			HCs vs. HVL LTNPs			HCs vs. VCs			HCs vs. ECs			HCs vs. <400 HICs			HC vs. >400 HICs		
rs1801270 + rs1059234	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	1.13	0.57-2.24	0.86	1.62	0.43-6.16	0.56	0.75	0.34-1.66	0.53	2.84	0.61-13.25	0.23	1.62	0.56-4.72	0.46	0.99	0.45-2.18	1
	Progressors vs. HICs																	
	Progressors vs. HICs			Progressors vs. HVL LTNPs			Progressors vs. VCs			Progressors vs. ECs			Progressors vs. <400 HICs			Progressors vs. >400 HICs		
rs1801270 + rs1059234	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	0.74	0.38-1.47	0.4	0.52	0.14-1.95	0.39	0.9	0.41-1.98	0.84	3.4	0.73-15.83	0.14	1.94	0.67-5.63	0.24	1.21	0.55-2.67	0.70
	HCs vs. HICs																	
	HCs vs. HICs			HCs vs. HVL LTNPs			HCs vs. VCs			HCs vs. ECs			HCs vs. <400 HICs			HC vs. >400 HICs		
rs1801270 without rs1059234	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	0.59	0.20-1.73	0.41	1.14	0.13-10.06	1.00	0.45	0.13-1.56	0.29	0.66	0.12-3.52	0.64	0.38	0.11-1.32	0.15	0.95	0.23-3.90	1
	Progressors vs. HICs																	
	Progressors vs. HICs			Progressors vs. HVL LTNPs			Progressors vs. VCs			Progressors vs. ECs			Progressors vs. <400 HICs			Progressors vs. >400 HICs		
rs1801270 without rs1059234	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	0.55	0.18-1.69	0.39	1.06	0.12-9.60	1.00	0.42	0.12-1.52	0.28	0.62	0.11-3.38	0.63	0.35	0.09-1.28	0.14	0.88	0.21-3.76	1
	HCs vs. HICs																	
	HCs vs. HICs			HCs vs. HVL LTNPs			HCs vs. VCs			HCs vs. ECs			HCs vs. <400 HICs			HC vs. >400 HICs		
rs1059234 without rs1801270	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	0	0-NaN	0.17	0	0-NaN	0.15	0	0-NaN	0.25	NaN	NaN-NaN	1.00	NaN	NaN-NaN	1	0.09	0- NaN	**0.09**
	Progressors vs. HICs																	
	Progressors vs. HICs			Progressors vs. HVL LTNPs			Progressors vs. VCs			Progressors vs. ECs			Progressors vs. <400 HICs			Progressors vs. >400 HICs		
rs1059234 without rs1801270	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
	0.74	0.10-5.40	1.00	0.35	0.03-4.20	0.40	0.71	0.06-8.14	1.00	∞	NaN-∞	1.00	∞	NaN-∞	1	0.44	0.06-3.28	0.59

Significant values and values that are trending towards significance are shown in bold highlighted in blue

3.2 Characterisation of the regulatory regions of *p21*

A region of the 5'UTR and the entire 3'UTR were sequenced in order to determine if polymorphisms in the regulatory regions of the *p21* gene were associated with HIV-1 control.

3.2.1 Test for Hardy-Weinberg Equilibrium

The genotypic data generated for all the SNPs found in the 5'UTR region and the 3'UTR were tested for deviation from Hardy-Weinberg equilibrium. There was no significant deviation from Hardy-Weinberg equilibrium for any polymorphism.

3.2.2 Variation in the *p21* 5'UTR region

A region of the *p21* 5'UTR was sequenced as outlined in 2.2.4 for 52 HICs and 44 progressors. Eighteen SNPs and one indel were identified in the sequenced 2248 bp promoter region. The positions and nucleotide changes of these polymorphisms are shown in Table 3.7. All of these polymorphisms were previously described and reported in the NCBI dbSNP database. No polymorphisms were found to be significantly over- or under-represented in either the HICs or the progressor cohort.

Due to the presence of numerous indels and long runs of identical nucleotides in regions of the promoter, sequencing was problematic and consequently some data are incomplete. Analysis of the variants identified in the promoter revealed that eleven of the detected polymorphisms were found to consistently occur together. Using Haploview software to determine the LD of these polymorphisms using only complete data sets (n=48), the resulting calculated r^2 value was 1, indicating that these eleven polymorphisms are found in complete LD. This putative intragenic haplotype is referred to as Hap-p21-P1 and was found in both the HICs and the progressor cohort at frequencies of 39.42% and 48.86% respectively. Due to the complete LD of SNPs making up Hap-p21-P1, we were able to assume presence of this putative haplotype even in incomplete data when other SNPs involved in the putative haplotype were also present i.e. we used these SNPs as 'tag' SNPs to predict the presence of the haplotype. All polymorphisms present in the *p21* promoter region were analysed using Haploview software to determine the LD of the polymorphisms present. The Haploview LD plot is shown in Figure 3.4.

The rs11326348 and rs113041051 SNPs (part of putative haplotype Hap-p21-P1) were only resolvable if a downstream indel (rs113749555; also present in Hap-p21-P1) was found in a

homozygous state, resulting in a correct alignment of the DNA strands and allowing the sequencing primer to read through. However, when rs113749555 was present in a heterozygous state, the sequences were no longer in alignment and the rs11326348 and rs113041051 SNPs were unresolvable, as well as all sequence upstream of the indel. Normally, to resolve this, primers are designed in the reverse direction in order to obtain the sequence up until the indel in the reverse direction. In this case, however, sequencing from the other direction was not possible due to an extended run of T nucleotides that affected the ability of the sequencing polymerase to successfully read the sequence. To overcome this, RFLP analysis was made use of as described in 2.2.3.2 in order to genotype the individuals with a WT/Mt rs113749555, to confirm our expectation that the rs11326348 and rs113041051 SNPs would also be found in the WT/Mt form, in fitting with the inferred putative haplotype (Hap-p21-P1).

The rs9368953 and rs9357222 SNPs (also part of Hap-p21-P1) were also found within a particularly difficult area to sequence, which accounts for the majority of the incomplete data. New primers were designed in an attempt to obtain clearer sequence data for the region harbouring the rs9368953 and rs9357222 SNPs (P21-Pro-930-F and P21-Pro-930-R; Table 2.5), however they only provided clear data for a subset of the samples. Due to available data (37/64) for these two SNPs and having representation of individuals with all three potential genotypes (WT/WT, WT/Mt and Mt/Mt), we did not deem it necessary to perform RFLP analysis as we could clearly see that they were in complete linkage with the other polymorphisms in Hap-p21-P1. For the rs11326348 and rs113041051 SNPs discussed above however, we needed to determine whether the heterozygous genotype (WT/Mt) of these SNPs was present when the other nine SNPs in the suspected haplotype were found in the heterozygous state, to determine if they were in fact part of this putative haplotype and in complete LD.

A 2-SNP putative intragenic haplotype (rs9394371 and rs3829963) that was also found to be present in both the HICs and the progressors was found to be associated with Hap-p21-P1. This putative haplotype is referred to as Hap-p21-P2. Hap-p21-P1 was occasionally present in certain individuals without Hap-p21-P2, however Hap-p21-P2 was never present without Hap-p21-P1, suggesting directional or conditional linkage between the two haplotypes. Hap-p21-P2 was found at an allelic frequency of 32.69% in the HICs and 38.64% in the progressor cohort. The rs9394371 and rs3829963 SNPs making up Hap-p21-P2 have an r^2 value of 1 and are therefore in complete LD, and an r^2 value of 0.67 when compared to the

other SNPs present in Hap-p21-P1 (Figure 3.4), i.e. the LD between Hap-p21-P1 and Hap-p21-P2 is strong but not complete. The genotypic frequencies of Hap-p21-P1 and Hap-p21-P2 when occuring together and separately are shown in Table 3.8.

A less prevalent 2-SNP haplotype involving the rs7485748 and rs4135234 SNPs was found to have an r^2 value = 0.74, suggesting relatively strong LD.

The allelic frequencies of Hap-p21-P1 (using the rs12214686 SNP as an identifying SNP) and Hap-p21-P2 were calculated for the HICs and progressors, including the HIC subgroups (Table 3.9). When comparing the allelic frequency of Hap-p21-P2 between ECs and progressors, the ECs had markedly less representation of the Mt allele haplotype (p=0.08, OR=2.83, CI=0.88-9.10). The allelic frequencies for Hap-p21-P1 and Hap-p21-P2 for all groups are shown graphically in Figure 3.5. No significant associations were found between ECs and progressors when comparing the genotypic frequencies of Hap-p21-P1 or Hap-p21-P2.

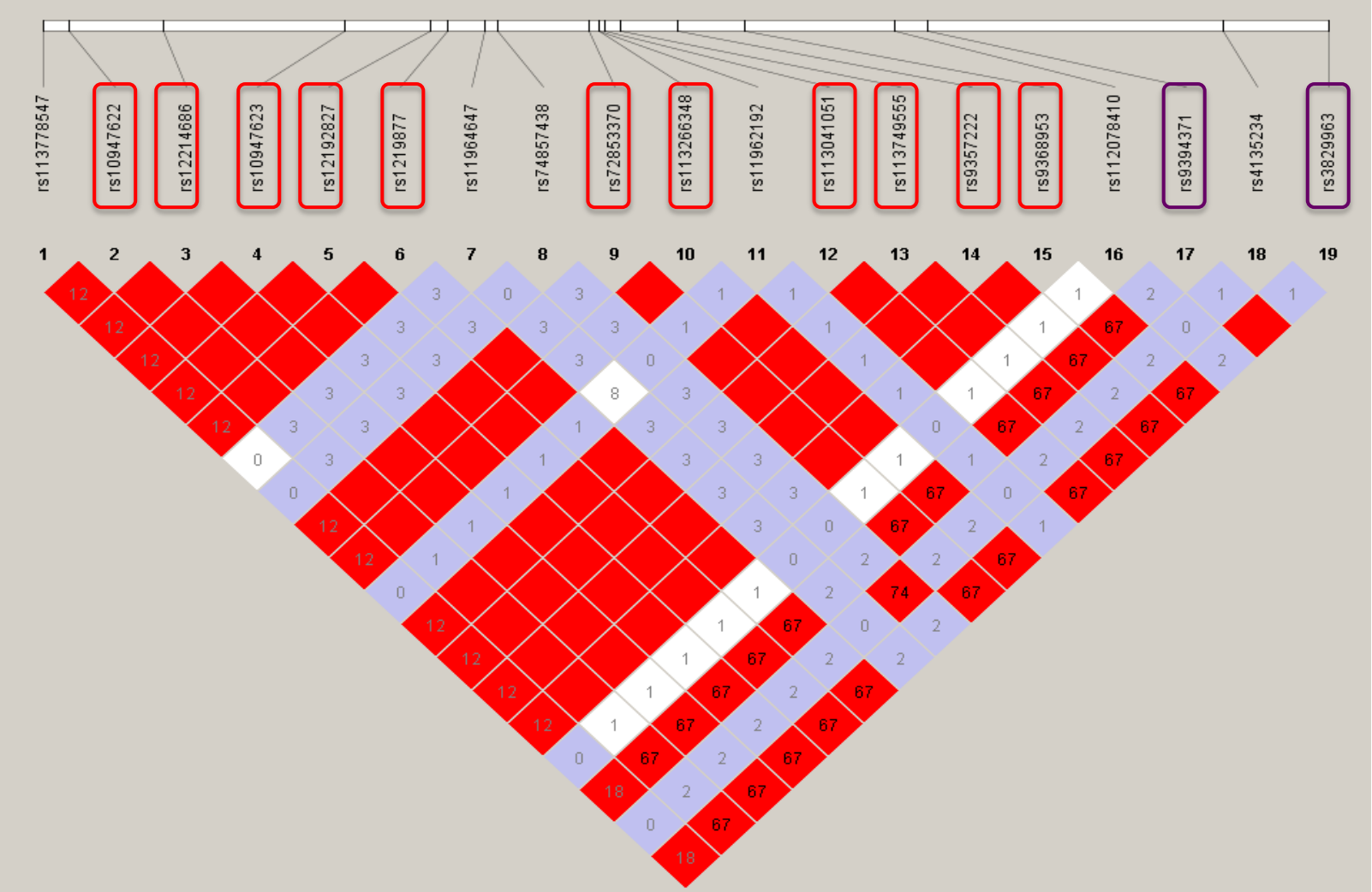


Figure 3.4 – Haploview LD Plot showing the r^2 values for all polymorphisms in the *p21* promoter region (n=96). The eleven-SNP Hap-p21-P1 (red blocks) has an r^2=1. The 2-SNP Hap-p21-P2 (purple blocks) has an r^2=1. The r^2 between Hap-p21-P1 and Hap-p21-p2 = 0.67. The numbers in each block show the r^2 value between SNPs.

Table 3.7 - Minor allele frequencies of polymorphisms in the *p21* promoter region for the HICs and progressor subgroups

SNP accession number	PROMOTER REGION		Minor Allele Frequency (sample number)	
	Position in gene	Base change (WT/Mt)	HICs	Progressors
rs113778547	-9518	A/G	11.54 (n=52)	6.82 (n=44)
rs10947622	-9477	C/T	39.42 (n=52)	40.91* (n=39)
rs12214686	-9328	A/G	38.46* (n=51)	48.86 (n=44)
rs10947623	-9043	G/A	39.42 (n=52)	48.86 (n=44)
rs12192827	-8908	C/T	39.42 (n=52)	48.86 (n=44)
rs12192877	-8880	C/A	39.42 (n=52)	48.86 (n=44)
rs11964647	-8821	T/A	5.77 (n=52)	3.41 (n=44)
rs74857438	-8802	A/G	5.77 (n=52)	2.27 (n=44)
rs72853370	-8657	A/G	38.46* (n=51)	48.86 (n=44)
rs113266348	-8642	C/T	39.42 (n=52)	48.86 (n=44)
rs11962192	-8641	G/A	0.96 (n=52)	3.41 (n=44)
rs113041051	-8633	A/G	39.42 (n=52)	48.86 (n=44)
rs113749555	-8609	G/-G	39.42 (n=52)	48.86 (n=44)
rs9357222	-8519	T/C	32.69* (n=48)	20.45* (n=24)
rs9368953	-8413	C/T	32.69* (n=48)	20.45* (n=24)
rs112078410	-8177	A/G	4.81 (n=52)	5.68 (n=44)
rs9394371	-8124	C/T	32.69 (n=52)	38.64 (n=44)
rs4135234	-7658	G/A	3.85 (n=52)	2.27 (n=44)
rs3829963	-7493	C/A	32.69 (n=52)	38.64 (n=44)

Polymorphisms involved in Hap-p21-P1 are highlighted in red. Polymorphisms involved in Hap-p21-P2 are highlighted in purple.

*Frequencies are lower than expected due to incomplete sequencing data. Number of individuals with complete data are shown in brackets next to the frequency.

Table 3.8 – Comparison of the genotypic frequencies of Hap-p21-P1 and Hap-p21-P2 in the HIC group and subgroups vs. the progressor group

Haplotype		**Genotypic frequency [Total number and frequency (%)]**							**Comparison between groups (p; CI; OR)**					
Hap-p21-P1	**Hap-p21-P2**	HICs (n=52)	Progressors (n=44)	HVL LTNPs (n=11)	VCs (n=30)	ECs (n=11)	<400s (n=20)	>400s (n=32)	HICs vs. progressors	HVL LTNPs vs. progessors	VCs vs. progressors	ECs vs. progressors	<400 HICs vs. progressors	>400 HICs vs. progressors
WT/WT	WT/WT	20 (38.46)	13 (29.55)	5 (45.45)	10 (33.33)	5 (45.45)	7 (35.00)	13 (40.63)	-	-	-	-	-	-
WT/Mt	WT/WT	4 (7.69)	4 (9.09)	0 (0.00)	2 (6.67)	2 (18.18)	3 (15.00)	1 (3.13)	0.70; 0.33-7.26; 1.54	0.54; NaN-∞; ∞	1.00; 0.23-10.15; 1.54	1.00; 0.11-5.61; 0.77	1.00; 0.12-4.16; 0.72	0.34; 0.39-40.80; 4.00
Mt/Mt	WT/WT	0 (0.00)	1 (2.27)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0.41; NaN-∞; ∞	1.00; NaN-∞; ∞	1.00; NaN-∞; ∞	1.00; NaN-∞; ∞	1.00; NaN-∞; ∞	1.00; NaN-∞; ∞
WT/Mt	WT/Mt	19 (36.54)	19 (43.18)	5 (45.45)	11 (36.67)	3 (27.27)	6 (30.00)	13 (40.63)	0.47; 0.60-3.96; 1.54	0.72; 0.35-6.08; 1.46	0.78; 0.44-4.03; 1.33	0.43; 0.49-12.01; 2.44	0.52; 0.47-6.25; 1.71	0.60; 0.51-4.15; 1.46
Mt/Mt	WT/Mt	3 (5.77)	3 (6.82)	0 (0.00)	2 (6.67)	1 (9.09)	1 (5.00)	2 (6.25)	0.67; 0.27-8.82; 1.54	0.55; NaN-∞; ∞	1.00; 0.16-8.27; 1.15	1.00; 0.10-13.88; 1.15	1.00; 0.14-18.52; 1.62	1.00; 0.21-10.52; 1.50
Mt/Mt	Mt/Mt	6 (11.54)	4 (9.09)	1 (9.09)	5 (16.67)	0 (0.00)	3 (15.00)	3 (9.38)	1.00; 0.24-4.35; 1.03	1.00; 0.14-17.34; 1.54	0.70; 0.13-2.90; 0.62	0.54; NaN-∞; ∞	1.00; 0.12-4.16; 0.72	1.00; 0.25-7.17; 1.33

Table 3.9 – Allelic frequencies of Hap-p21-P1 and Hap-p21-P2 in HICs, HIC subgroups and progressors

SNP position in *p21* gene													Haplotype Frequency (%)						
-9477	-9328	-9043	-8908	-8880	-8657	-8642	-8633	-8609	-8519	-8413	-8124	-7493	HICs (N=52)	ECs (N=11)	VCs (N=30)	HVL LTNPs (N=11)	<400 HICs (N=20)	>400 HICs (N=32)	Progressors (N=44)
Hap-p21-P1													39.4	31.8	45.0	31.8	42.5	37.5	42.9
											Hap-p21-P2		32.7	18.2	38.3	31.8	32.5	32.8	38.6

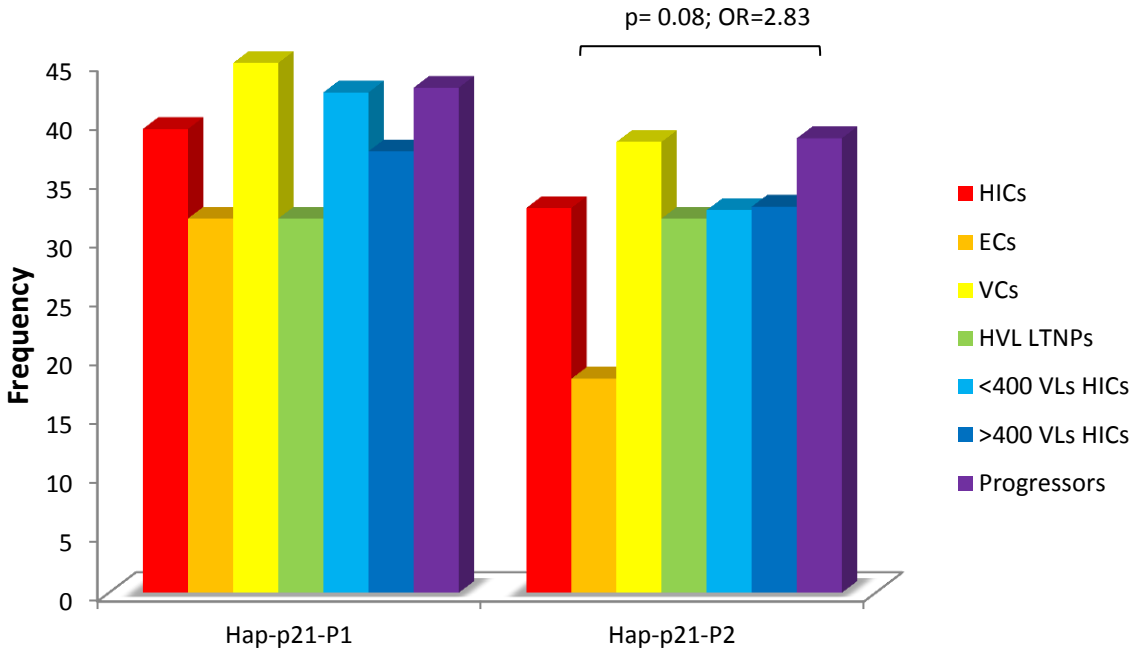


Figure 3.5 - Graphical representation of Hap-p21-P1 and Hap-p21-P2 allelic frequencies in the different groups studied. HICs: HIV-1 controllers, ECs: elite controllers, VCs: viraemic controllers, HVL LTNPs: high viral load long term non-progressors, <400 VLs HICs: HICs with <400 VL (RNA copies/ml), >400 VLs HICs: HICs with >400 VL (RNA copies/ml)

3.2.3 Variation in the *p21* 3'UTR

The *p21* 3'UTR was sequenced for 52 HICs and 44 progressors. Twelve polymorphisms were detected in the sequenced region, including two SNPs that have not previously been reported in the NCBI dbSNP database. The positions and nucleotide changes of these polymorphisms are shown in Table 3.10. Minor allele frequencies of polymorphisms found in the 3'UTR for the progressor and HIC cohorts and HIC subgroups are shown in Table 3.11. Genotypic frequencies are shown in Table 3.12.

Analysis revealed that the rs1059234 minor T allele had markedly lower representation in the ECs (p=0.09, CI=0.81-17.27, OR=3.75) and <400 HIC cohort (p=0.07, CI=0.92-7.49, OR=2.63) when compared to progressors. As the rs1059234 SNP was one of the three SNPs genotyped in the larger cohort in 3.1, the statistical analyses from that analysis was considered as more accurate.

In silico analysis of this SNP revealed that the binding sites for various miRNAs were either lost or gained, or the ddG score altered when the major allele or minor allele was present Table 3.13). ddG is a free energy score and thus, the more negative the value, the stronger the predicted binding will be (*http://genie.weizmann.ac.il/pubs/mir07/mir07_notes.html*). As there are high ddG scores present when either the major or minor allele of the rs1059234 SNP is present, this suggests that this SNP may have some functional impact on p21 expression. The number in the 'seed' column refers to an "X:Y:Z" notation for describing the seed, with the X value representing the size of the seed, the Y value representing the number of mismatches and the Z value the number of G:U wobble pairs (*http://genie.weizmann.ac.il/pubs/mir07/mir07_notes.html*). Analysis was performed using the Segal Lab of Computational Biology's microRNA prediction tool *(http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html*).

The unreported SNP found at position +2925 was only found in the progressor cohort, at a frequency of 2.27%, while the unreported SNP at position +3090 was only found in the HIC group, at a frequency of 0.96%. The rs111923164 and rs74801436 SNPs were only detected in the progressor cohort, while the rs186592256, rs114982296 and rs181350370 SNPs were only detected in the HIC group. LD analysis using Haploview software showed no LD between any polymorphisms in the 3'UTR (Figure 3.6).

No SNPs were significantly under- or over-represented in either of the two groups, indicating that the SNPs we detected in the 3'UTR may not be contributing to the differential expression of p21.

Table 3.10 – Base change of all polymorphisms found in the 3' UTR as well as position in the *p21* gene

SNP accession number	Position in gene	Base change (WT/Mt)
rs1059234	+1719	C/T
rs111926164	+1750	C/T
rs112675295	+1842	G/A
rs114982296	+1956	C/T
rs181350370	+2075	C/T
rs184742749	+2353	G/C
rs186592256	+2578	C/T
rs74801436	+2646	C/G
rs3176359	+2864	G/A
Unreported SNP	+2925	G/A
Unreported SNP	+3090	G/A
rs73730143	+3101	C/T

Table 3.11 – Minor allele frequencies of polymorphisms found in the 3'UTR for the progressor and HIC cohort and HIC subgroups

SNP	Minor allele frequency (Total number and frequency (%))						
	Progressors (n=44)	HICs (n=52)	HVL LTNPs (n=11)	VCs (n=30)	ECs (n=11)	<400 HICs (n=20)	>400 HICs (n=32)
rs1059234	24 (27.27)	20 (19.23)	4 (18.18)	14 (23.33)	2 (9.09)	5 (12.5)	15 (23.44)
rs111926164	2 (2.27)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs112675295	2 (2.27)	6 (5.77)	1 (4.55)	3 (5.00)	2 (9.09)	3 (7.50)	3 (4.69)
rs114982296	0 (0.00)	1 (0.96)	0 (0.00)	1 (1.66)	0 (0.00)	0 (0.00)	1 (1.56)
rs181350370	0 (0.00)	2 (1.92)	0 (0.00)	2 (3.33)	0 (0.00)	1 (2.50)	1 (1.56)
rs184742749	2 (2.27)	6 (5.77)	1 (4.55)	4 (6.67)	1 (4.55)	2 (5.00)	4 (6.25)
rs186592256	0 (0.00)	1 (0.96)	0 (0.00)	0 (0.00)	1 (4.55)	1 (2.50)	0 (0.00)
rs74801436	2 (2.27)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs3176359	7 (7.95)	10 (9.62)	4 (18.18)	3 (5.00)	3 (13.64)	3 (7.50)	7 (10.94)
NI SNP (+2925)	2 (2.27)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
NI SNP (+3090)	0 (0.00)	1 (0.96)	0 (0.00)	0 (0.00)	1 (4.55)	1 (2.50)	0 (0.00)
rs73730143	2 (2.27)	2 (1.92)	1 (4.55)	0 (0.00)	1 (4.55)	1 (2.50)	1 (1.56)

NI = newly identified

Table 3.12 - Genotypic frequencies of polymorphisms found in the 3'UTR for the progressor and HIC cohort and HIC subgroups

SNP	Genotype	Genotypic frequency (Total number and frequency (%))						
		Progressors (n=44)	HICs (n=52)	HVL LTNPs (n=11)	VCs (n=30)	ECs (n=11)	<400 HICs (n=20)	>400 HICs (n=32)
rs1059234	WT/WT	24 (54.55)	33 (63.46)	4 (36.36)	20 (66.67)	9 (81.82)	15 (75.00)	18 (56.25)
	WT/Mt	17 (38.64)	18 (34.62)	7 (63.64)	9 (30.00)	2 (18.18)	5 (25.00)	13 (40.63)
	Mt/Mt	3 (6.82)	1 (1.92)	0 (0.00)	1 (3.37)	0 (0.00)	0 (0.00)	1 (3.13)
rs111926164	WT/WT	42 (95.45)	52 (100.00)	11 (100.00)	30 (100.00)	11 (100.00)	20 (100.00)	32 (100.00)
	WT/Mt	2 (4.55)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs112675295	WT/WT	42 (95.45)	46 (88.46)	10 (90.91)	27 (90.00)	9 (81.82)	17 (85.00)	29 (90.63)
	WT/Mt	2 (4.55)	6 (11.54)	1 (9.09)	3 (10.00)	2 (18.18)	3 (15.00)	3 (9.37)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs114982296	WT/WT	44 (100.00)	51 (98.08)	11 (100.00)	29 (96.67)	11 (100.00)	20 (100.00)	31 (96.88)
	WT/Mt	0 (0.00)	1 (1.92)	0 (0.00)	1 (3.37)	0 (0.00)	0 (0.00)	1 (3.12)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs181350370	WT/WT	44 (100.00)	50 (96.15)	11 (100.00)	28 (93.33)	11 (100.00)	19 (95.00)	31 (96.88)
	WT/Mt	0 (0.00)	2 (3.85)	0 (0.00)	2 (6.67)	0 (0.00)	1 (5.00)	1 (3.12)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs184742749	WT/WT	42 (95.45)	46 (88.46)	10 (90.91)	26 (86.67)	10 (90.91)	18 (90.00)	28 (87.5)
	WT/Mt	2 (4.55)	6 (11.54)	1 (9.09)	4 (13.33)	1 (9.09)	2 (10.00)	4 (12.5)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs186592256	WT/WT	44 (100.00)	51 (98.08)	11 (100.00)	30 (100.00)	10 (90.91)	19 (95.00)	32 (100.00)
	WT/Mt	0 (0.00)	1 (1.92)	0 (0.00)	0 (0.00)	1 (9.09)	1 (5.00)	0 (0.00)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs74801436	WT/WT	42 (95.45)	52 (100.00)	11 (100.00)	30 (100.00)	11 (100.00)	20 (100.00)	32 (100.00)
	WT/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	Mt/Mt	2 (4.55)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs3176359	WT/WT	37 (84.09)	42 (80.77)	7 (63.64)	27 (90.00)	8 (72.73)	17 (85.00)	25 (78.13)
	WT/Mt	7 (15.91)	10 (19.23)	4 (36.36)	3 (10.00)	3 (27.27)	3 (15.00)	7 (21.87)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Unreported SNP (+2925)	WT/WT	42 (95.45)	52 (100.00)	11 (100.00)	30 (100.00)	11 (100.00)	20 (100.00)	32 (100.00)
	WT/Mt	2 (4.55)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Unreported SNP (+3090)	WT/WT	44 (100.00)	51 (98.08)	11 (100.00)	30 (100.00)	10 (90.91)	19 (95.00)	32 (100.00)
	WT/Mt	0 (0.00)	1 (1.92)	0 (0.00)	0 (0.00)	1 (9.09)	1 (5.00)	0 (0.00)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
rs73730143	WT/WT	42 (95.45)	50 (96.15)	10 (90.91)	30 (100.00)	10 (90.91)	19 (95.00)	31 (96.88)
	WT/Mt	2 (4.55)	2 (3.85)	1 (9.09)	0 (0.00)	1 (9.09)	1 (5.00)	1 (3.12)
	Mt/Mt	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

Table 3.13 – In silico analysis of the *p21* 3'UTR rs1059234 SNP

rs1059234 WT Allele (C)

Gene	microRNA	Position	Seed	ddG
Seq1	hsa-miR-509-3-5p	16	8:1:0	-6.68
Seq1	hsa-miR-363	16	8:1:1	-2.08
Seq1	hsa-miR-1288	18	8:1:0	-11.92
Seq1	hsa-miR-206	18	8:1:1	-10.12
Seq1	hsa-miR-647	18	8:1:1	-9.32
Seq1	hsa-miR-1	18	8:1:1	0.31
Seq1	hsa-miR-1278	19	8:1:1	0.67
Seq1	hsa-miR-1224-5p	21	8:1:1	-10.23
Seq1	hsa-miR-636	27	8:1:1	-4.63
Seq1	hsa-miR-520f	28	8:1:1	-9.87
Seq1	hsa-miR-517b	28	8:1:1	-5.01
Seq1	hsa-miR-1249	28	8:1:0	-3.81
Seq1	hsa-miR-517c	29	8:1:1	-3.41
Seq1	hsa-miR-517a	29	8:1:1	-2.71

rs1059234 Mt Allele (T)

Gene	microRNA	Position	Seed	ddG
Seq1	hsa-miR-509-5p	16	8:1:1	-5.3
Seq1	hsa-miR-509-3-5p	16	8:1:1	-5
Seq1	hsa-miR-140-3p	16	8:1:0	-1.1
Seq1	hsa-miR-1288	18	8:1:1	-10.2
Seq1	hsa-miR-1224-5p	21	8:1:1	-10.49
Seq1	hsa-miR-636	27	8:1:1	-3.35
Seq1	hsa-miR-520f	28	8:1:1	-10.59
Seq1	hsa-miR-517b	28	8:1:1	-4.79
Seq1	hsa-miR-1249	28	8:1:0	-3.99
Seq1	hsa-miR-517c	29	8:1:1	-2.82
Seq1	hsa-miR-517a	29	8:1:1	-2.12

Altered score
Large change in score
Loss of binding site for miRNA
Gain of binding site for miRNA

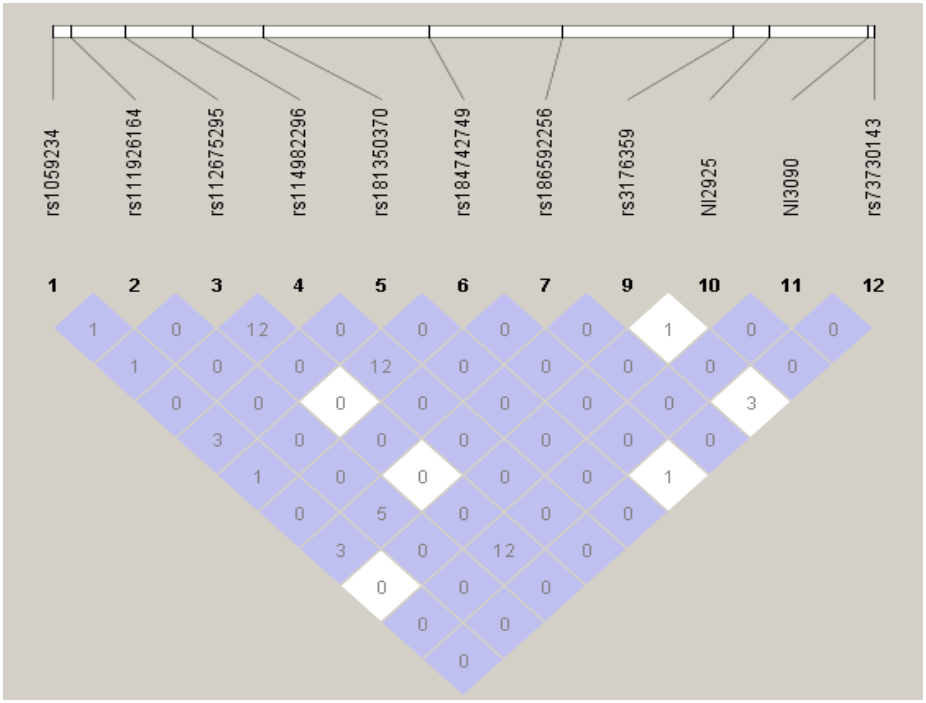


Figure 3.6 – Haploview LD plot for all polymorphisms found in the *p21* 3'UTR. NI=newly identified. The number after NI is the position of the newly identified SNP. The numbers in each block show the r^2 value between SNPs.

3.3 *p21* mRNA expression analysis

3.3.1 Flow cytometry showing CD4+ T cell purity

Flow cytometry was performed on isolated CD4+ T cells in order to determine the purity of the cells as described in section 2.3.1.3. Results showed a purity of 97.4% (Figure 3.7).

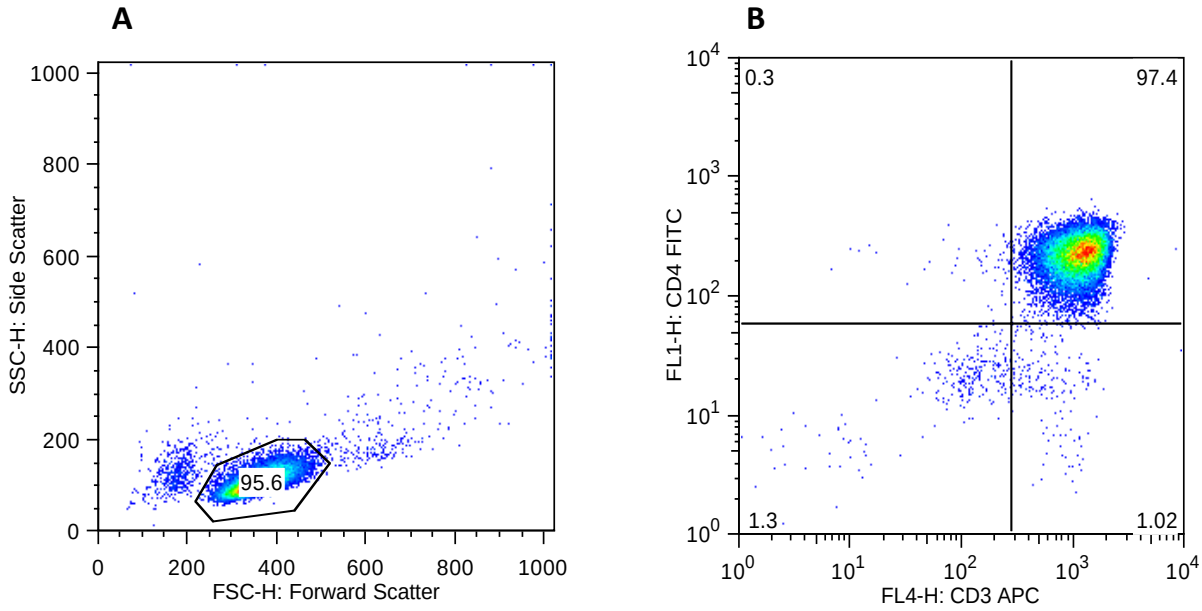


Figure 3.7 - Flow cytometry plots showing (A) gating of lymphocytes and (B) percentage of CD4+ T cells

3.3.2 Assessing the quality of extracted mRNA

All extracted RNA was analysed for integrity using Bioanalyzer software. An example of the gel and electropherogram resulting from RNA analysis is shown in Figure 3.8. The image shows excellent quality RNA with all RNA samples in this image having an RNA Integrity Number (RIN) of >7 as calculated by the Bioanalyzer system.

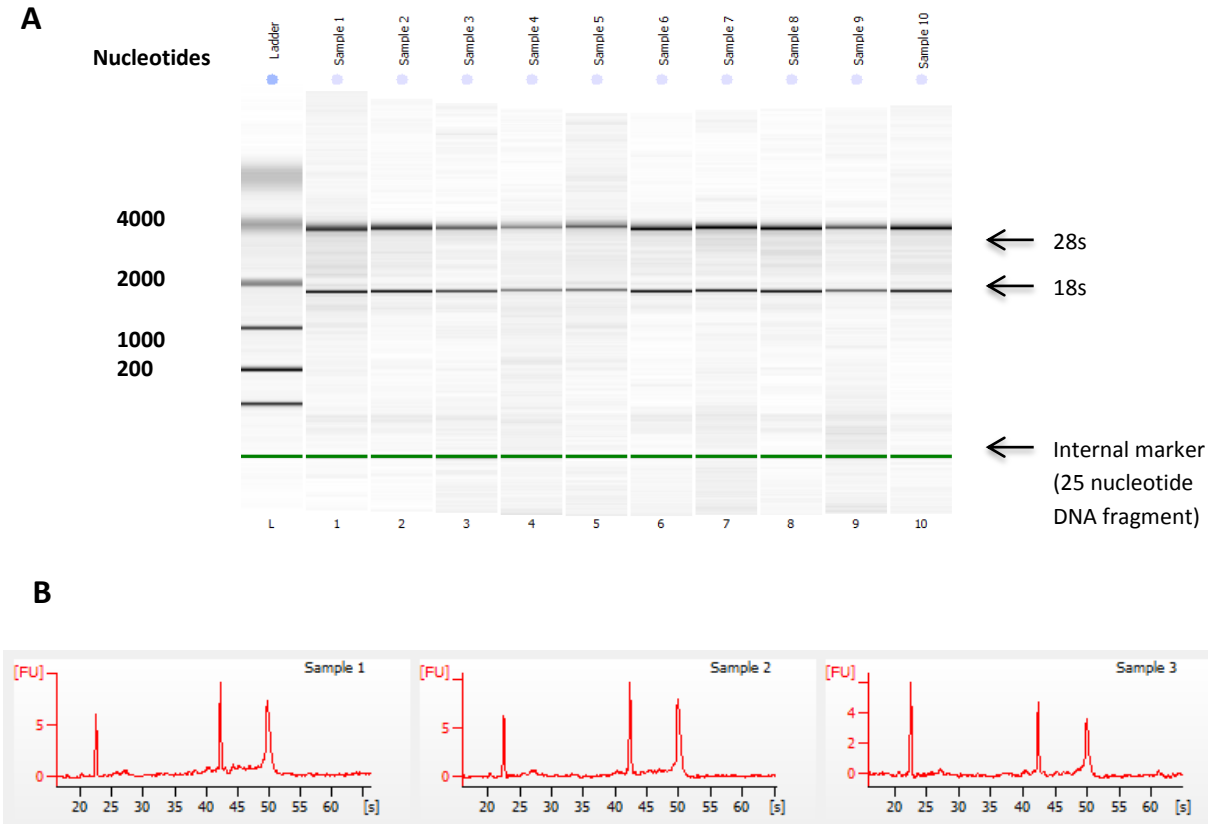


Figure 3.8 – Bioanalyzer image depicting (A) a gel image of the RNA samples against an RNA ladder and (B) the electropherogram results of the RNA samples. The green band in the gel image depicts the marker, with the first black band from the bottom depicting the 18s RNA, and the second black band depicting the 28s RNA. The first peak in the electropherogram is the marker, the second peak is the detection of the 18s ribosomal RNA and the third peak is the detection of the 28s ribosomal RNA

3.3.3 Real-Time relative expression of p21

Expression levels of *p21* mRNA were determined for a total of 50 individuals, grouped into 10 healthy controls, 12 progressors and 25 HICs (3 individuals were excluded from phenotypic groupings), as outlined in section 2.3.3.

3.3.3.1 CCR5 as a marker of HIV-1 activation

A CCR5 expression assay was included as a control since CCR5 expression is known to correlate with immune activation and hence viral load (Ostrowski et al., 1998). As expected, analysis revealed that *CCR5* mRNA expression was significantly correlated with VL (r=0.53, CI=0.25 to 0.72, p=0.0005). In addition, a significant inverse correlation was found between *CCR5* mRNA expression and CD4+ T cell count (r=-0.49, CI= -0.70 to -0.21, p=0.0011).

3.3.3.2 Relative p21 mRNA expression and HIV-1

When stratifying HIV-1 infected individuals (n=40) by CD4+ T cell counts (<200, 200-500, >500), there was a significant difference in *p21* mRNA expression between those with lower CD4+ T cell counts (<200) and higher CD4+ T cell counts (>500), with *p21* mRNA expression being higher in those with lower CD4+ T cell counts (p=0.004) (Figure 3.9A). When individuals were grouped by VL (<1000, >1000), *p21* mRNA expression was significantly higher in individuals in the >1000 VL group (p=0.0002) (Figure 3.9B). Results remained significant when the outlier in the group with <200 CD4 and >1000 VL was removed (p=0.0008 and p=0.0003 respectively).

A significant correlation was found between relative *p21* mRNA expression and VL, (r=0.68, CI=0.45 to 0.82, p<0.0001). CD4+ T cell count was also found to be significantly inversely correlated with *p21* mRNA expression (r=-0.53, CI=-0.73 to -0.25, p=0.0005), likely due to the strong inverse correlation found between VL and CD4 count (r=-0.75, CI=-0.86 to -0.57, p<0.0001). As expected, relative *CCR5* mRNA expression was significantly correlated with relative *p21* mRNA expression, most likely due to their respective relationships with VL (r=0.78, CI=0.61 to 0.88, p<0.0001) (Figure 3.10).

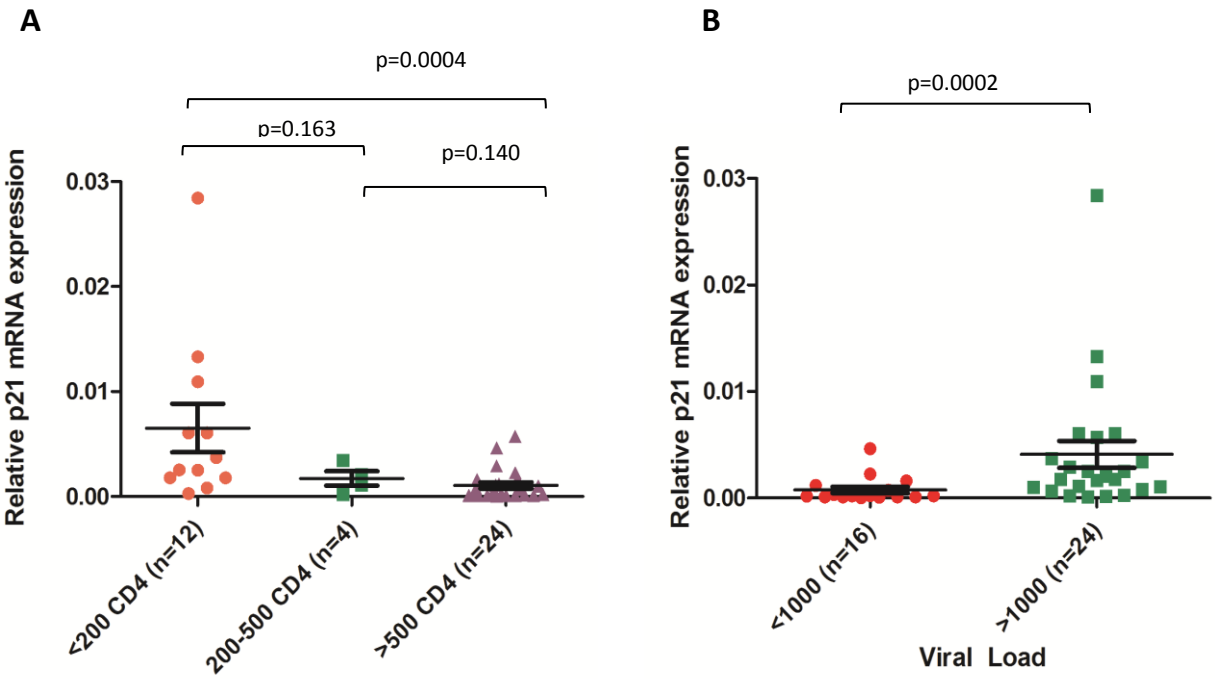


Figure 3.9 – Graphs showing relative *p21* mRNA expression when HIV-1 infected individuals were stratified by (A) CD4+ T cell count (cells/µl) and (B) viral load (RNA copies/ml). The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

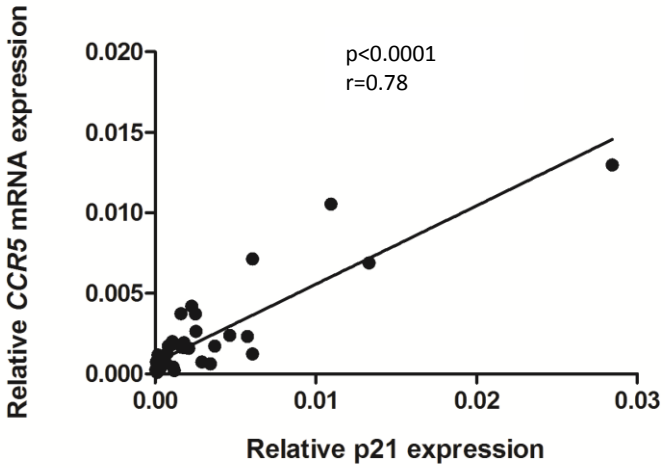


Figure 3.10 – Correlation between relative *CCR5* mRNA expression and relative *p21* mRNA expression

3.3.3.3 Relative p21 mRNA expression between HICs and progressors

When comparing relative *p21* mRNA expression between HICs (n=25) and progressors (n=12), a significant difference between the groups was found (Figure 3.11). However, instead of seeing p21 expression upregulated in the HIC group as previous work has shown, *p21* mRNA expression was significantly decreased (p=0.0004) when compared to progressors. When the outlier in the progressor group was removed from the analysis, significance remained (p=0.001). When the HIC group was divided into the EC, VC and HVL LTNP subgroups and compared to progressors, the ECs and VCs had significantly decreased *p21* mRNA expression when compared to the progressor group. The HC group had similar levels of *p21* mRNA expression when compared to the HICs. Interestingly, among the HIC subgroups, only the HVL LTNPs were not significantly different to the progressors with respect to *p21* mRNA expression, although this is a very small group (n=4).

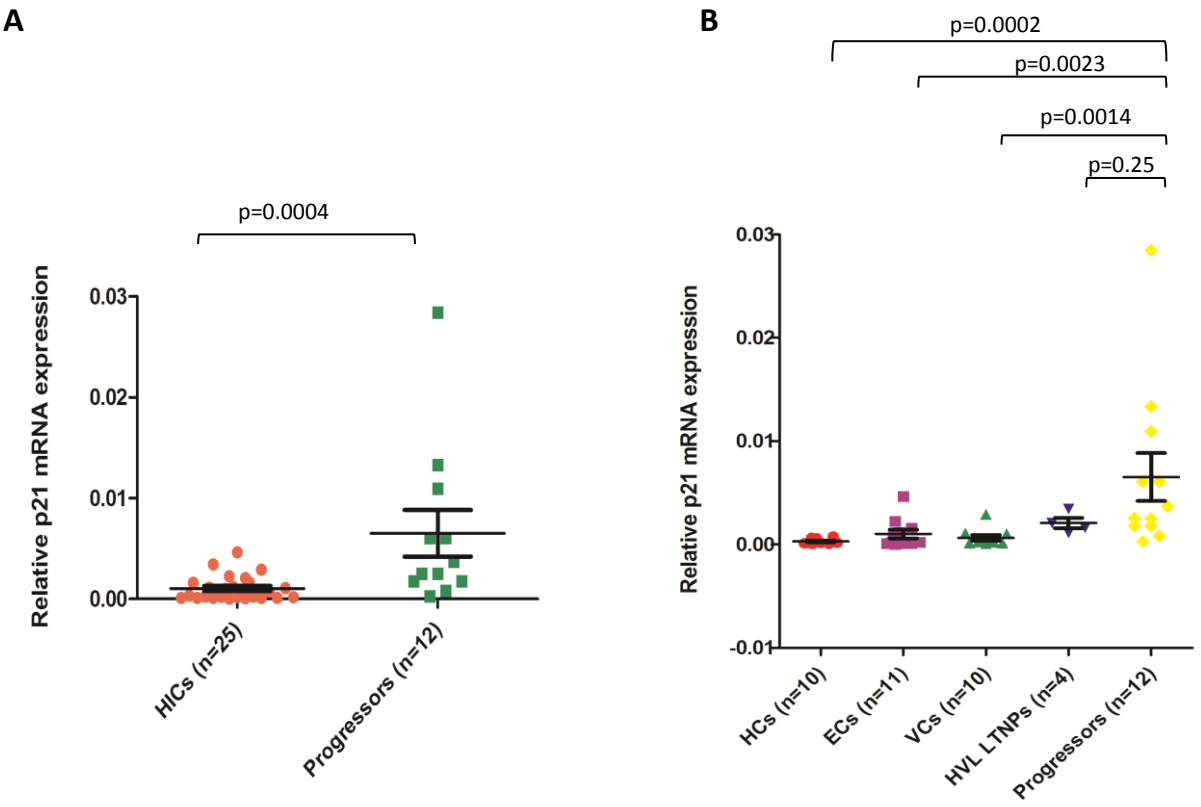


Figure 3.11 – Graphs showing relative *p21* mRNA expression between (A) HICs and progressor group and (B) the HIC subgroups and the progressor group. As this comparison was based on clinical phenotypic groupings, the three non-confirmed controllers were excluded from analyses. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

3.3.3.4 p21 rs1801270 and rs1059234 SNPs and miR-106b rs999885 SNP and HIV-1 markers of disease progression

Although we did not see a significant role in HIV-1 control for the two *p21* SNPs, all individuals genotypes including controls (n=50) were stratified according to their rs1801270 and rs1059234 genotypes in order to see if there was any difference in HIV-1 VL and CD4+ T cell counts with the presence or absence of the respective SNPs.

When stratifying the *p21* rs1801270 and rs1059234 SNP genotypes and comparing to viral load, no genotype of either of the two SNPS were significantly associated with viral load or CD4+ T cell count.

When stratifying the miR-106b rs999885 SNP genotypes against viral load, no significant association was found. In addition, no significant association was found when stratifying against CD4+ T cell counts.

3.3.3.5 p21 rs1801270 and rs1059234 SNPs and relative p21 mRNA expression

Although we did not see a significant role in HIV-1 control for the two *p21* SNPs, individuals were stratified according to their rs1801270 and rs1059234 genotypes in order to see if these SNPs were playing a role in differential *p21* mRNA expression.

Both the rs1801270 SNP genotype and the rs1059234 SNP genotype did not impact variation of *p21* mRNA expression (Figure 3.12A and Figure 3.13A). When both genotypes with the minor allele (WT/Mt + Mt/Mt) were grouped against the WT/WT genotype, no significance between the groups for either SNP was noted (Figure 3.12B and Figure 3.13B).

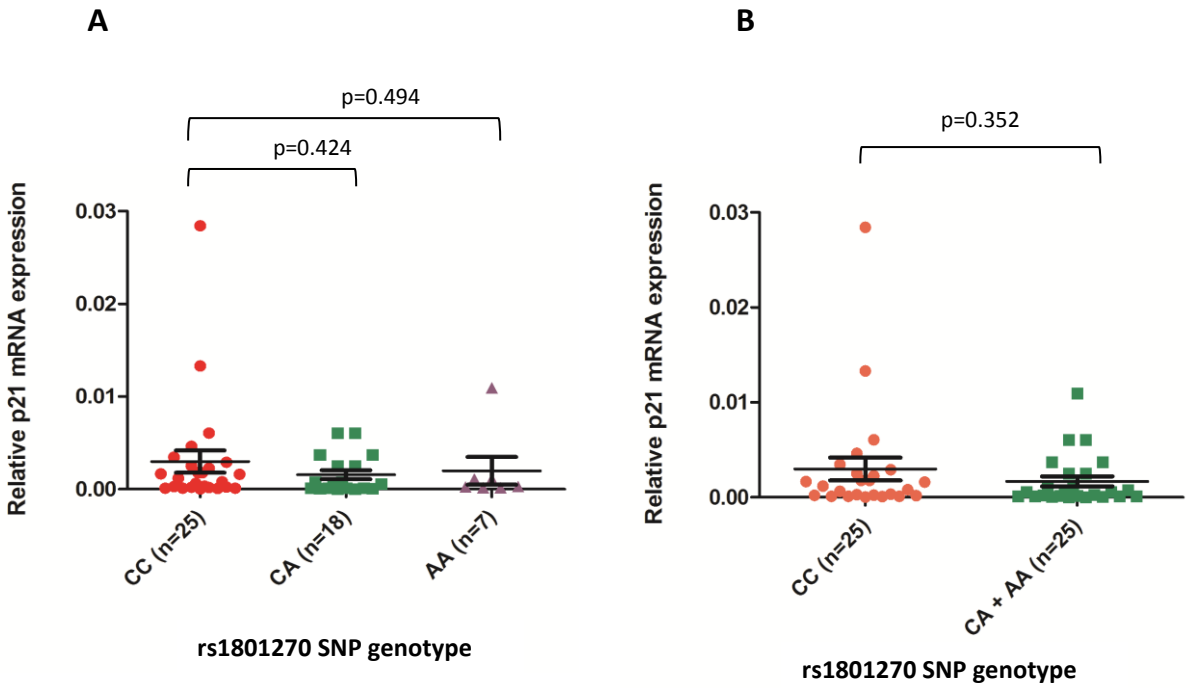


Figure 3.12 – Graphs comparing (A) relative *p21* mRNA expression for individuals with the CC, CA and AA genotype of the rs1801270 SNP as well as (B) the CC genotype and the CA + AA genotype. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

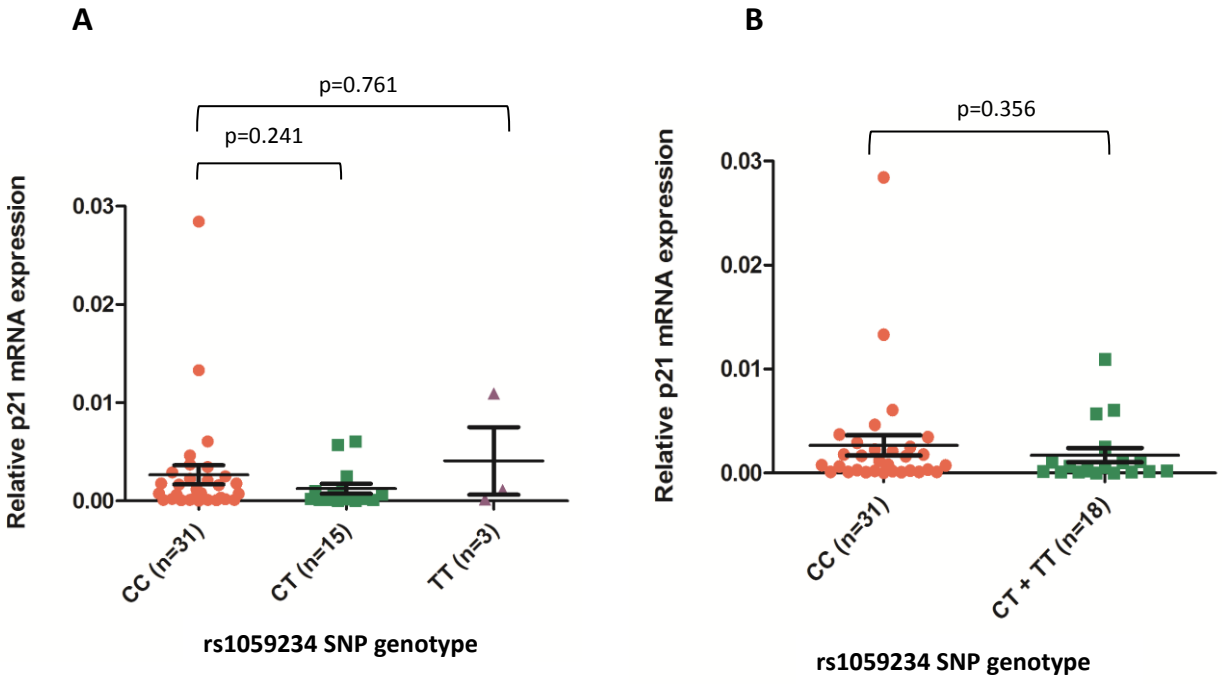


Figure 3.13 - Graphs comparing (A) relative *p21* mRNA expression for individuals with the CC, CT and TT genotype of the rs1059234 SNP as well as (B) the CC genotype and the CT + TT genotype. One individual was excluded due to failure to genotype. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

3.3.3.6 miR-106b rs999885 SNP and relative p21 expression

All individuals (n=50) were stratified according to their rs999885 genotype and *p21* mRNA expression was compared.

The GA genotype showed a strong trend of association with higher *p21* mRNA expression when compared to the GG genotype (p=0.066) (Figure 3.14A). The influence of the AA genotype could not be determined as only one individual possessed the AA genotype. When both genotypes with the minor allele (WT/Mt + Mt/Mt) were grouped against the WT/WT genotype, a trend to increased *p21* mRNA expression in the GA + AA group when compared to the GG genotype was maintained (p=0.08) (Figure 3.14B). When the outlier possessing the GA genotype was excluded from the analyses, however the trend towards significance was lost [(A) p=0.12; (B) p=0.14]. When comparing individuals with the GG genotype and the GA + AA genotype against VL, the GA + AA group had higher VLs when compared to the GG group (Figure 3.15), although this was not significant (p=0.155).

A bigger sample size will be necessary to determine if the mir-106b rs999885 SNP GA genotype truly has an effect on p21 expression.

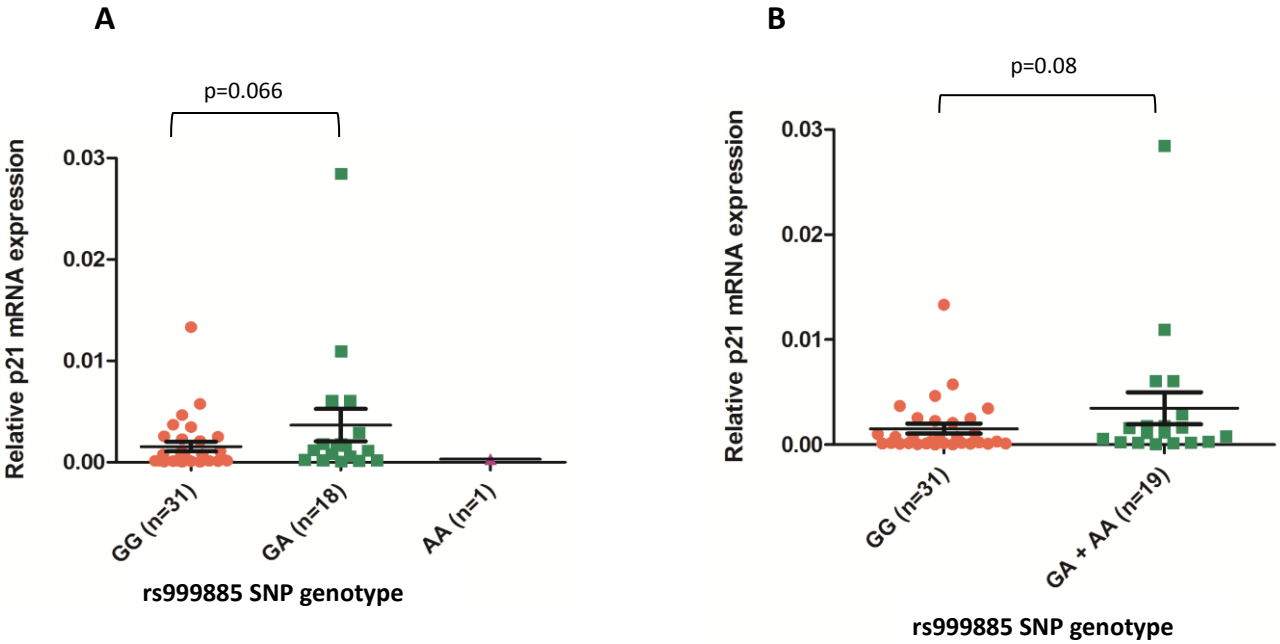


Figure 3.14 – Graphs comparing (A) relative *p21* mRNA expression for individuals with the GG, GA and AA genotype of the rs999885 SNP as well as (B) the GG genotype and the GA + AA genotype. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

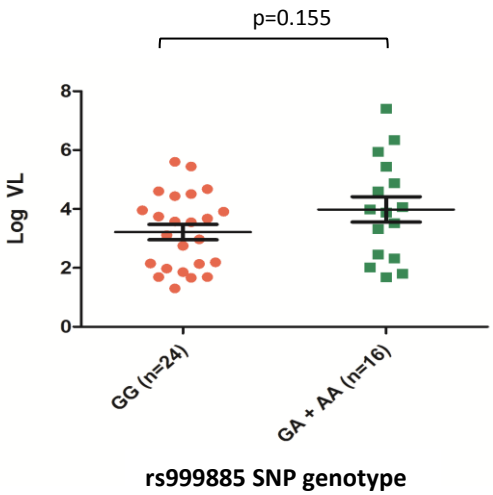


Figure 3.15 – Comparison of viral loads of individuals with the rs999885 SNP GG genotype and the GA + AA genotype. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

3.3.3.7 The p21 5'UTR rs733590 SNP and relative p21 mRNA expression

Given that a recent study showed the *p21* 5'UTR rs733590 SNP to be positively associated with increased relative *p21* mRNA expression in healthy individuals (de Pablo et al., 2015), we next genotyped the 50 individuals for the SNP. When stratifying p21 expression according to the various rs733590 SNP genotypes, no effect of any genotype or allele on *p21* mRNA expression was found in our population (Figure 3.16). When the CT outlier was removed, the CT group compared to the TT group resulted in p=0.241.

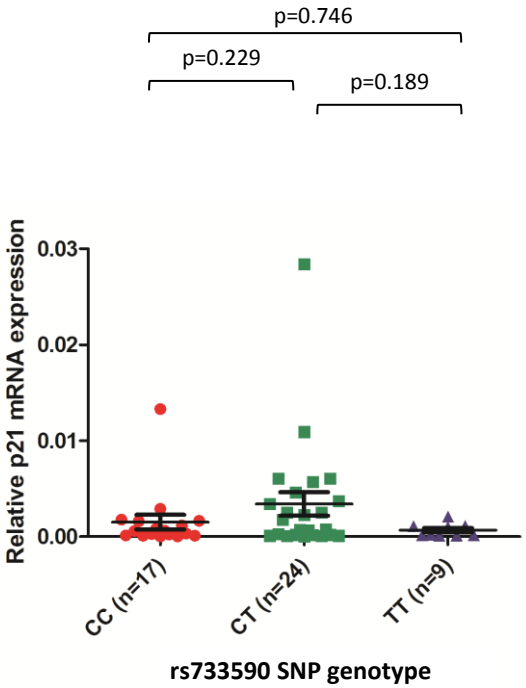


Figure 3.16 - Graph showing relative *p21* mRNA expression of individuals with the CC, CT and TT rs733590 SNP genotypes. The long solid line in the middle of the data points indicates the median, and the top and bottom lines are at the 75th and 25th percentiles respectively

4. DISCUSSION

Both the HIV-1 pathogen and the host have an inherent variability that results in differing levels of individual control with regards to the progression of HIV-1 (reviewed by Santa-Marta et al., 2013). There has been great interest into the concept of a functional cure for HIV-1. For an individual to be 'functionally cured' of HIV-1, they need to exhibit long-term control of viral replication to undetectable levels without the aid of ARVs (Katlama et al., 2013). While a 'sterilizing cure' rids the patient of all traces of HIV-1, functional cure can be achieved even if replication-competent viruses remain in the body (Katlama et al., 2013). Among HIV-1 infected individuals, there are unique individuals who are able to naturally suppress the virus and exhibit slow progression of the disease, without the use of ARVs. These rare individuals are termed elite controllers (ECs) and are found in less than 1% of HIV-1 infected individuals in studied HIV cohorts (reviewed by Okulicz, 2012). These individuals represent a model for the study of functional cure of HIV-1.

We do not yet fully understand the correlates of protective immunity with regards to HIV-1 infection. There is no one factor that seems to be responsible for natural control of HIV-1. Previous research has examined multiple mechanisms of control including chemokine receptor variation, HLA and KIR variation, and intrinsic host proteins. However, only a subset of ECs and HIV-1 controllers previously studied have an identified protective characteristic (viral and/or host) (reviewed by Okulicz, 2012), suggesting that there are other as yet unidentified factors involved that allow for the control of HIV-1 by these individuals.

In addition to different population groups exhibiting variation in the mechanisms that may underlie natural control of HIV-1, there is also variation among the differently defined controller groups. Even individuals within the strictly defined elite controller group display immense heterogeneity with regards to immune and genetic characteristics (Tiemessen and Martinson, 2012), emphasising the fact that multiple factors may play a role in natural HIV-1 control. In addition, different protective factors may be acting within different controller groups that have consistently low or high viraemia i.e. what may play a role in ECs who exhibit undetectable HIV-1 viraemia may not necessarily be shared by a group of controllers with higher viral loads who clearly do not control their viraemia but their CD4+ T cells are protected from the effects of high virus (unlike progressors).

As research regarding the development of an HIV-1 vaccine has been the focus of clinical trials in recent years, the need to identify a vaccine-induced immune response to HIV-1 is imperative (Qin et al., 2007). The development of an effective HIV vaccine is greatly

impacted by the lack of knowledge surrounding correlates of immune protection (Prado et al., 2011). Furthering knowledge of factors involved in natural control of HIV-1 and insight into both host and viral genetic variation will inform both therapeutic and protective vaccine design (Tomaras and Haynes, 2014). The access to cohorts exhibiting varying levels of natural HIV-1 control has given us the opportunity to investigate factors that could be having impacting HIV-1 natural control.

The cellular protein p21 (also known as CDKN1A/WAF-1/Cip-1) has been implicated in the natural control of HIV-1 infection. p21 is a *cip/kip* family inhibitor and acts as a regulator of the cell cycle by inhibiting cyclin-dependent kinases and regulating the transition of replicating cells from G_1 - S (Arias et al., 2007). p21 has been shown to inhibit both CDK9 (Salerno et al., 2007) and CDK2 (Brugarolas et al., 1998) disrupting HIV-1 replication through inhibiting HIV-1 reverse transcription and HIV-1 mRNA transcription from proviral DNA (Flores et al., 1999).

A key study regarding p21 and HIV-1 control, and one that has formed the basis for our study, has shown that both ECs and VCs have an upregulation of p21 in CD4+ T cells when compared to progressors and HCs (Sáez-Cirión et al., 2011, Chen et al., 2011). Chen et al. (2011) found both *p21* mRNA levels and p21 protein levels to be upregulated in CD4+ T cells of ECs and VCs, and found both to be correlated, showing that *p21* mRNA expression is a good indicator of p21 protein expression. Therefore in our study, we have used *p21* mRNA expression as a marker of p21 protein expression.

Subsequent to the work conducted by Chen et al. (2011), recent studies have introduced the idea that HIV-1 may be exploiting various cellular mechanisms in order to regulate expression of specific genes, including *p21* (Farberov et al., 2015, Guha et al., 2016). In macrophages, HIV-1-induced overexpression of p21 has been attributed to the HIV-1 Vpr protein (Amini et al., 2004a). In another study, Amini et al. (2004b) showed that Vpr (along with the cellular transcription factor Sp1 which binds to the promoter of *p21*) stimulated *p21* gene transcription. The role of p21 in HIV-1 control seems to be cell-dependent, as varying effects are shown in macrophages when compared to CD4+ T cells.

De Pablo et al. (2015) showed that regardless of the patients' phenotype, there was broad inter-individual variation in levels of p21 expression. However, they found a positive association with increased p21 expression and possession of the HLA-B*2705 allele and/or

the *p21* 5'UTR rs733590 SNP. When individuals possessed the HLA-B*2705 allele, significantly higher p21 expression was observed (p=0.007). This was also true when individuals possessed the TT genotype of the *p21* 5'UTR rs733590 SNP (p=0.039). When an individual possessed both the T allele of the *p21* 5'UTR rs733590 SNP and the HLA-B*2705 allele, p21 expression was found to be largely increased (p=0.004). De Pablo et al. (2015) did not find complete linkage disequilibrium between the HLA-B*2705 allele and the *p21* 5'UTR rs733590 SNP, however this SNP was highly enriched in individuals with the HLA-B*2705 allele. In our opinion, this suggests that there may be a certain level of LD between these variants, since they are found in close proximity on chromosome 6.

African populations in general exhibit large genetic variation when compared to Caucasian populations, with different polymorphisms, differential SNP or variant distributions, and differences in patterns of LD (Picton et al., 2010, Gentle et al., 2013). Given that most, if not all studies that have shown a role for p21 in the control of HIV-1 have been conducted in Caucasian populations infected with subtype B HIV-1, one cannot extrapolate a similar role for p21 in a sub-Saharan African population infected with a different subtype of HIV-1 (subtype C). Although not stated in the cohort description, it is assumed that Chen et al. (2011) saw this upregulation of p21 in ECs in a Caucasian population since the study was conducted in the United States of America (USA). In addition, the prevalent HIV-1 subtype in the USA is subtype B. The current study was thus set up to investigate the role of p21 in HIV-1 natural control in a subtype C HIV-1 infected South African Black population.

In this study, we undertook a thorough investigation into the variation in the two regulatory regions of the *p21* gene (5'UTR and 3'UTR), determined if these variants, in addition to two additional *p21* SNPs and a SNP in the miR-106b miRNA that has been shown to regulate p21, were involved in p21 HIV-1 natural control by looking at their representation in groups of controllers and progressors, and looked at relative *p21* mRNA expression in a subset of these individuals.

Gravina et al. (2009) investigated *p21* polymorphisms in the context of aging and showed that two *p21* SNPs, rs1801270 and rs1059234, were significantly under-represented among centenarians. They thus concluded that these two SNPs were potentially detrimental to longevity. The rs1801270 SNP has also been associated with increased risk of endometrial cancer in Korean populations (Roh et al., 2004), and Li et al. (2005) found the rs1801270 SNP, in addition to the rs1059234 SNP, to be associated with susceptibility to squamous cell

carcinoma of the head and neck. Since these two SNPs have not previously been studied in the context of p21 and natural HIV-1 control, individuals with varying levels of HIV-1 control were genotyped for the two SNPs to determine if the SNPs were, independently or together, assiciated with HIV-1 control.

Looking at data from the 1000 Genomes Project (1000genomesproject.org), as expected, the representation of these two SNPs varied greatly when comparing African populations to Caucasian populations. Our healthy Black South African population showed similar allelic and genotypic frequencies for both the rs1801270 and rs1059234 SNPs when compared to four other sub-Saharan African populations (with the Nigerian Yoruba population being the most similar to our population), in addition to the global population. This, together with no deviations from Hardy-Weinberg equilibrium, served to further confirm that our designed C_T shift assays to genotype these two SNPs performed well.

The *p21* exonic rs1801270 SNP was not found to be over- or under-represented in any group. In addition, when stratifying relative *p21* mRNA expression according to the possession of various rs1801270 SNP genotypes, no difference was shown in expression between the WT or Mt alleles. The *p21* 3'UTR rs1059234 SNP, however, showed a potential effect on control in that ECs had markedly less representation of the minor allele in addition to having a lower representation of the heterozygous genotype when compared to progressors, however this was not statistically significant. When stratifying relative *p21* mRNA expression to the various rs1059234 genotypes, however, no significant association of this SNP on mRNA expression was found. Future work with larger sample sizes is needed to determine if this SNP is contributing to HIV-1 elite control. Given that this SNP does not appear to influence p21 expression, the mode of influence on control in ECs, if any, may involve mechanisms that do not depend on p21 expression levels.

The miR-106b family of microRNAs has been found to strongly regulate p21 expression (Ivanovska et al., 2009). The miR-106b family is hosted in the *MCM7* gene, where a SNP (rs999885) has been reported in the promoter region. When we looked at this SNP in relation to *p21* mRNA expression, we saw a strong trend towards higher expression of *p21* mRNA in individuals possessing the GA genotype. Given that Liu et al. (2012) report that the AG/GG genotype is associated with increased expression of miR-106b, one might expect that this increase in miR-106b would be associated with lower p21 expression, as Ivanovska et al. (2008) showed miR-106b to downregulate p21. Thus our study showed a tendency to the

opposite scenario. When an outlier was removed from our analysis, the trend was lost, however we still did not find an association with decreased p21 expression.

There are two major differences between our study and the study by Liu et al. (2012). Firstly, Liu et al. (2012) studied the miR-106b rs999885 SNP in non-tumour liver cells whereas we studied the SNP in CD4+ T cells. Secondly, Liu et al. (2012) were studying an Asian population, and in sub-Saharan African populations, the major allele is the G allele, whereas in the Asian populations, the G allele is the minor allele (*1000genomes.org*). These differences may very likely contribute to the discordant results regarding this SNP's effect on miR-106b and consequent p21 expression. The significant under-representation of the GA genotype in controllers with higher viral loads is nonetheless interesting and worthy of future investigation. It is also important to note that miRNAs act on more than one gene and thus *p21* may not be the gene impacted by the SNP in miR-106b. As HIV-1 has recently been shown to dysregulate p21 through regulation of miR-106b *in vitro* (Guha et al., 2016), the presence of higher HIV-1 viraemia could directly impact on miR-106b. It is possible that the rs999885 SNP may be facilitating a more advantageous association with HIV-1, hence its significant underrepresentation in controllers in the presence of higher viraemia. There is therefore a need for further testing with larger sample sizes in order to evaluate the true significance of these findings, along with functional studies of the miR-106b miRNA with and without the rs999885 SNP and its consequent effect on p21 and its expression.

Since p21 expression can be affected by polymorphisms in the regulatory region of the gene, in addition to looking at representation of individual SNPs, a 2248 bp region from the 5'UTR and the entire 3'UTR were sequenced from a subset of the genotyped individuals to investigate any polymorphisms found and their association with HIV-1 natural control. In the *p21* 5'UTR region that was sequenced, 18 SNPs and one indel were detected. Eleven of these polymorphisms were found to always occur together, and LD analysis between these polymorphisms revealed an r^2 value of 1, indicating that these eleven polymorphisms were found in complete linkage.

This 11-SNP haplotype was named Hap-p21-P1 and to our knowledge is the first report of this haplotype in the *p21* 5'UTR. An additional 2-SNP haplotype (Hap-p21-P2) was also discovered in this region. Interestingly, Hap-p21-P1 and Hap-p21-P2 were also found to be in LD, with an r^2 value of 0.67. The linkage between these two haplotypes was found to be uni-directional i.e. Hap-p21-P1 was occasionally found without Hap-p21-P2 but Hap-p21-P2

was always found with Hap-p21-P1 (hence $r^2<1$). This suggests that evolutionarily, Hap-p21-P1 was likely to have formed first and later acquired the additional SNPs. It would be interesting to determine if the relationship between the two haplotypes represents a functional evolutionary relationship and if they impact on p21 expression.

Analysis of these haplotypes revealed that Hap-p21-P1 was not over- or under-represented in any group. When comparing the allelic frequency of Hap-p21-P2 between ECs and progressors, the ECs had markedly less representation of the Mt allele haplotype (p=0.08). Although we were unable to design C_T shift assays to detect a tag SNP in the two haplotypes, future work will involve looking at p21 expression relative to these two haplotypes.

Twelve SNPs were detected in the sequenced 3'UTR. No SNPs in the 3'UTR other than previously discussed rs1059234 were found to be significantly over- or under-represented in any group. In silico analysis of all the SNPs found in the 3'UTR revealed various miRNA binding sites that were either lost, gained or the ddG score altered when the major allele or minor allele was present. rs1059234 was found to be the most prevalent SNP in the 3'UTR and the one that showed a difference when comparing ECs with progressors. In silico analysis revealed large changes in the ddG scores of various miRNAs when comparing the major and minor allele of rs1059234, indicating that this SNP may be affecting p21 regulation although no impact on *p21* mRNA expression was seen. Interestingly, none of the SNPs detected in the 3'UTR were located in the predictive binding region of the miR-106b miRNA. The SNPs identified in the 3'UTR may very well be altering miRNA binding and affecting p21 expression as suggested by in silico analysis, however one needs to functionally assess this in *in vitro* assays, as in silico analysis is purely predictive and may not be functionally relevant.

Since, to our knowledge, nothing is known about the genetic variation in the *p21* gene regulatory regions in a Black South African population, these findings will help to further our knowledge on p21 and the effects of polymorphisms within the gene, not only in the context of HIV-1 control, but for cancers and other disease variants. Thus further work is required to determine the significance of these newly identified haplotypes and polymorphisms.

Our expression analysis revealed that *p21* mRNA expression was significantly correlated with both VL (r=0.68, p<0.0001) and CD4+ T cell count (r=-0.53, p=0.0005). HIV-1 controllers and the HIV-1 controller subgroups had significantly lower *p21* mRNA expression when compared to progressors, except for HVL LTNPs, supposedly due to the

impact of VL on p21 expression. CCR5 expression in CD4+ T cells has been shown to be upregulated in HIV-infected individuals when compared to uninfected controls (Ostrowski et al., 1998). Thus *CCR5* mRNA expression was measured along with *p21* mRNA expression as a marker of activation. *p21* mRNA expression was significantly correlated with *CCR5* mRNA expression, further emphasising the relationship between VL and *p21* mRNA expression.

Guha et al. (2016) also showed in an *in vitro* study that HIV-1 infected CD4+ T cells had an upregulation of p21 expression when compared to non-infected CD4+ T cells, indicating that the HIV-1 itself has some impact on p21 expression. HIV-1 was also shown to downregulate miR-106b in CD4+ T cells, thereby increasing p21 expression (Guha et al., 2016). This suggests that an upregulation of p21 in ECs, as found by Chen et. al. (2011) is likely due to a host factor in these individuals that is responsible for increased p21 expression, as the low viral load in ECs cannot alone explain the increased p21 expression seen.

As mentioned earlier, de Pablo et al. (2015) found a positive association between higher p21 expression and possession of the HLA-B*2705 allele and/or the *p21* 5'UTR rs733590 SNP. HLA-B*2705 is virtually absent in most of the sub-Saharan Africa populations (*http://www.allelefrequencies.net*), with Black South Africans not having any representation of the allele (Mijiyawa et al., 2000, Paximadis et al., 2012). As HLA-B*2705 has been significantly associated with an increased p21 expression, the Black South African population would then be expected to have an overall lower level of p21, and would therefore be unlikely to show an effect on HIV-1 control as has been shown in Caucasian populations. It is interesting to speculate whether the protective effect displayed by HLA-B*2705 against HIV-1 could thus partially be due to the action of p21 in Caucasian populations.

De Pablo et al. (2015) also showed that the TT genotype of the *p21* 5'UTR rs733590 SNP was associated with higher relative *p21* mRNA expression when compared to the CC genotype. This was also shown previously in a Caucasian population by Korthagen et al. (2012). Representation data available on the 1000 Genomes Project (*1000genomes.org*) showed that the rs733590 SNP T allele has a much higher representation in the Caucasian populations (43.5%) versus the African populations (18%), again indicating that African populations may inherently have a lower level of p21 expression when compared to Caucasians, if this SNP is actually impacting on p21 expression and is not just in LD with another variant that be the true functional variant. We thus decided to genotype the

individuals with expression data for the *p21* 5'UTR rs733590 SNP and interestingly, found no effect of this SNP on *p21* mRNA expression in our population group.

This result, together with the absence of HLA-B*2705 in our population group, may be the reason that we do not see upregulation of *p21* mRNA in our ECs and VCs as seen in Caucasian cohorts. This study thus serves to highlight the dangers of extrapolating identified correlates of protection from one population group to another without consideration of the host genetic background.

It will be interesting to look at the rs733590 SNP in HIV-1 negative South African Caucasians with and without the HLA-B*2705 allele to see if this SNP is associated with p21 expression in this population, in addition to comparing Caucasians to the Black South Africans that were genotyped in this study to see if the Caucasian population has inherently higher levels of p21 expression.

In the Black South African population, the presence of HIV-1 itself correlates positively with *p21* mRNA expression i.e. p21 looks to be serving as a marker of immune activation as is seen with CCR5 (Ostrowski et al., 1998). It has previously been shown that individuals infected with HIV-1 had a 2.35-fold increase in *p21* mRNA expression relative to healthy individuals (Serrao et al., 2014). In addition, Guha et al. (2016) determined that CD4+ T cells that were productively infected with HIV-1 had higher amounts of p21 expression when compared to bystander cells. It has also been proposed that HIV-1 can affect the host miRNAs thereby blocking innate inhibitory mechanisms in the cell, and so promoting the spread of HIV-1 (Farberov et al., 2015). As there is still some confusion with regards to whether the host cell and/or HIV-1 is modulating p21, further studies quantitating baseline levels of p21 and effect on HIV-1 control in varied population groups will be highly informative.

An intrinsic factor, SAMHD1, has also been shown to play a role in HIV-1 control and has been associated with p21. SAMHD1 is a cellular enzyme that halts replication of the HIV-1 virus in dendritic cells, macrophages and monocytes (Laguette et al., 2011, Hrecka et al., 2011) due to SAMHD1 depleting the amount of deoxynucleotide triphosphates (dNTPs) in the cell. Similarly, Allouch et al. (2013) showed that one of the protective mechanisms of p21 against HIV-1 is its ability to block dNTP biosynthesis, through downregulating the RNR2 subunit of ribonucleotide reductase. While Allouch et al. (2013) concluded that HIV-1

restriction by p21 was in a SAMHD1 independent manner, Pauls et al. (2014) suggested that the effect of p21 was dependent on SAMHD1 expression. More studies will need to be done to determine the relationship, if any, between p21 and SAMHD1 and the effect on HIV-1 control, especially in a South African Black population.

This study emphasises the importance of studying disease in the context of the affected population. We found major variations in Black South Africans when compared to Caucasians. Variation in protective mechanisms in different population groups can be attributed to host differences as well as to HIV-1 subtype/strain differences.

In conclusion, genotypic data analysis revealed a potential role for the *p21* 3'UTR rs1059234 SNP in HIV-1 control, while the *p21* exonic rs1801270 SNP did not seem to be playing a role. Possessing the miR-160b rs999885 SNP GA genotype was found to be associated with higher p21 expression, although further work is needed to understand this relationship. Sequencing of the *p21* 5'UTR revealed two interesting haplotypes that need to be further studied and compared to p21 expression. Sequencing of the 3'UTR did not reveal any additional SNPs, other than the rs1059234 SNP, that seemed to participate in HIV-1 control, however functional experiments will need to be performed to truly understand the roles of these SNPs and the consequent interaction of various miRNAs. Expression analysis in our population group showed a significant correlation between HIV-1 VL and *p21* mRNA expression. Future work with larger sample sizes will be beneficial in uncovering more information regarding the role of p21 and natural HIV-1 control in a South African Black population.

5. REFERENCES

ABBAS, T. & DUTTA, A. 2009. p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer,* 9**,** 400-14.

ABRAM, M. E., FERRIS, A. L., SHAO, W., ALVORD, W. G. & HUGHES, S. H. 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol,* 84**,** 9864-78.

ALLOUCH, A., DAVID, A., AMIE, S. M., LAHOUASSA, H., CHARTIER, L., MARGOTTIN-GOGUET, F., BARRE-SINOUSSI, F., KIM, B., SAEZ-CIRION, A. & PANCINO, G. 2013. p21-mediated RNR2 repression restricts HIV-1 replication in macrophages by inhibiting dNTP biosynthesis pathway. *Proc Natl Acad Sci U S A,* 110**,** E3997-4006.

ALTFELD, M., ADDO, M. M., ROSENBERG, E. S., HECHT, F. M., LEE, P. K., VOGEL, M., YU, X. G., DRAENERT, R., JOHNSTON, M. N., STRICK, D., ALLEN, T. M., FEENEY, M. E., KAHN, J. O., SEKALY, R. P., LEVY, J. A., ROCKSTROH, J. K., GOULDER, P. J. & WALKER, B. D. 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS,* 17**,** 2581-91.

AMINI, S., KHALILI, K. & SAWAYA, B. E. 2004a. Effect of HIV-1 Vpr on cell cycle regulators. *DNA Cell Biol,* 23**,** 249-60.

AMINI, S., SAUNDERS, M., KELLEY, K., KHALILI, K. & SAWAYA, B. E. 2004b. Interplay between HIV-1 Vpr and Sp1 modulates p21(WAF1) gene expression in human astrocytes. *J Biol Chem,* 279**,** 46046-56.

ARCHIN, N. M., SUNG, J. M., GARRIDO, C., SORIANO-SARABIA, N. & MARGOLIS, D. M. 2014. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat Rev Micro,* 12**,** 750-764.

ARIAS, C. F., BALLESTEROS-TATO, A., GARCIA, M. I., MARTIN-CABALLERO, J., FLORES, J. M., MARTINEZ, A. C. & BALOMENOS, D. 2007. p21CIP1/WAF1 controls proliferation of activated/memory T cells and affects homeostasis and memory T cell responses. *J Immunol,* 178**,** 2296-306.

BARRE-SINOUSSI, F., ROSS, A. L. & DELFRAISSY, J.-F. 2013. Past, present and future: 30 years of HIV research. *Nat Rev Micro,* 11**,** 877-883.

BASHIROVA, A. A., THOMAS, R. & CARRINGTON, M. 2011. HLA/KIR Restraint of HIV: Surviving the Fittest. *Annual review of immunology,* 29**,** 295-317.

BERGAMASCHI, A., DAVID, A., LE ROUZIC, E., NISOLE, S., BARRE-SINOUSSI, F. & PANCINO, G. 2009. The CDK inhibitor p21Cip1/WAF1 is induced by FcgammaR activation and restricts the replication of human immunodeficiency virus type 1 and related primate lentiviruses in human macrophages. *J Virol,* 83**,** 12253-65.

BOYER, J. C., BEBENEK, K. & KUNKEL, T. A. 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 6919-6923.

BRAUNSTEIN, S. L., VAN DE WIJGERT, J. H. & NASH, D. 2009. HIV incidence in sub-Saharan Africa: a review of available data with implications for surveillance and prevention planning. *AIDS Rev,* 11**,** 140-56.

BRUGAROLAS, J., BRONSON, R. T. & JACKS, T. 1998. p21 Is a Critical CDK2 Regulator Essential for Proliferation Control in Rb-deficient Cells. *The Journal of Cell Biology,* 141**,** 503-514.

BURKHART, B. A., ALCORTA, D. A., CHIAO, C., ISAACS, J. S. & BARRETT, J. C. 1999. Two posttranscriptional pathways that regulate p21(Cip1/Waf1/Sdi1) are identified by HPV16-E6 interaction and correlate with life span and cellular senescence. *Exp Cell Res,* 247**,** 168-75.

BUTLER, J. E. & KADONAGA, J. T. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev,* 16**,** 2583-92.

CELLA, M., LONGO, A., FERRARA, G. B., STROMINGER, J. L. & COLONNA, M. 1994. NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. *The Journal of Experimental Medicine,* 180**,** 1235-1242.

CHEN, H., LI, C., HUANG, J., CUNG, T., SEISS, K., BEAMON, J., CARRINGTON, M. F., PORTER, L. C., BURKE, P. S., YANG, Y., RYAN, B. J., LIU, R., WEISS, R. H., PEREYRA, F., CRESS, W. D., BRASS, A. L., ROSENBERG, E. S., WALKER, B. D., YU, X. G. & LICHTERFELD, M. 2011. CD4+ T cells from elite controllers resist HIV-1 infection by selective upregulation of p21. *J Clin Invest,* 121**,** 1549-60.

CHEN, W. C., WU, H. C., HSU, C. D., CHEN, H. Y. & TSAI, F. J. 2002. p21 gene codon 31 polymorphism is associated with bladder cancer. *Urol Oncol,* 7**,** 63-6.

CHENG, T., RODRIGUES, N., SHEN, H., YANG, Y., DOMBKOWSKI, D., SYKES, M. & SCADDEN, D. T. 2000. Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science,* 287**,** 1804-8.

CHIU, Y.-L., SOROS, V. B., KREISBERG, J. F., STOPAK, K., YONEMOTO, W. & GREENE, W. C. 2005. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. *Nature,* 435**,** 108-114.

DE PABLO, A., BOGOI, R., BEJARANO, I., TORO, C., VALENCIA, E., MORENO, V., MARTIN-CARBONERO, L., GOMEZ-HERNANDO, C. & RODES, B. 2015. p21/CDKN1A Expression Shows Broad Interindividual Diversity in a Subset of HIV-1 Elite Controllers. *AIDS Res Hum Retroviruses.*

EL-DEIRY, W. S., TOKINO, T., VELCULESCU, V. E., LEVY, D. B., PARSONS, R., TRENT, J. M., LIN, D., MERCER, W. E., KINZLER, K. W. & VOGELSTEIN, B. 1993. WAF1, a potential mediator of p53 tumor suppression. *Cell,* 75**,** 817-25.

FAN, X., LIU, Y. & CHEN, J. J. 2005. Down-regulation of p21 contributes to apoptosis induced by HPV E6 in human mammary epithelial cells. *Apoptosis,* 10**,** 63-73.

FARBEROV, L., HERZIG, E., MODAI, S., ISAKOV, O., HIZI, A. & SHOMRON, N. 2015. MicroRNA-mediated regulation of p21 and TASK1 cellular restriction factors enhances HIV-1 infection. *Journal of Cell Science,* 128**,** 1607-1616.

FLORES, O., LEE, G., KESSLER, J., MILLER, M., SCHLIEF, W., TOMASSINI, J. & HAZUDA, D. 1999. Host-cell positive transcription elongation factor b kinase activity is essential and limiting for HIV type 1 replication. *Proc Natl Acad Sci U S A,* 96**,** 7208-13.

FUJITA, M., OTSUKA, M., MIYOSHI, M., KHAMSRI, B., NOMAGUCHI, M. & ADACHI, A. 2008. Vpx Is Critical for Reverse Transcription of the Human Immunodeficiency Virus Type 2 Genome in Macrophages. *Journal of Virology,* 82**,** 7752-7756.

GARTEL, A. L. & TYNER, A. L. 2002. The role of the cyclin-dependent kinase inhibitor p21 in apoptosis. *Mol Cancer Ther,* 1**,** 639-49.

GENTLE, N. L., PAXIMADIS, M., PUREN, A. & TIEMESSEN, C. T. 2013. Genetic variability in markers of HLA-C expression in two diverse South African populations. *PLoS One,* 8**,** e67780.

GILLESPIE, S., KADIYALA, S. & GREENER, R. 2007. Is poverty or wealth driving HIV transmission? *AIDS,* 21**,** S5-S16.

GONZALEZ, E., BAMSHAD, M., SATO, N., MUMMIDI, S., DHANDA, R., CATANO, G., CABRERA, S., MCBRIDE, M., CAO, X.-H., MERRILL, G., O'CONNELL, P., BOWDEN, D. W., FREEDMAN, B. I., ANDERSON, S. A., WALTER, E. A., EVANS, J. S., STEPHAN, K. T., CLARK, R. A., TYAGI, S., AHUJA, S. S., DOLAN, M. J. & AHUJA, S. K. 1999. Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. *Proceedings of the National Academy of Sciences,* 96**,** 12004-12009.

GOULDER, PHILIP J. R. & WALKER, BRUCE D. 2012. HIV and HLA Class I: An Evolving Relationship. *Immunity,* 37**,** 426-440.

GRAVINA, S., LESCAI, F., HURTEAU, G., BROCK, G. J., SARAMAKI, A., SALVIOLI, S., FRANCESCHI, C. & RONINSON, I. B. 2009. Identification of single nucleotide polymorphisms in the p21 (CDKN1A) gene and correlations with longevity in the Italian population. *Aging (Albany NY),* 1**,** 470-80.

GUHA, D., MANCINI, A., SPARKS, J. & AYYAVOO, V. 2016. HIV-1 Infection Dysregulates Cell Cycle Regulatory Protein p21 in CD4+ T Cells through miR-20a and miR-106b Regulation. *Journal of Cellular Biochemistry***,** n/a-n/a.

GUMPERZ, J. E., LITWIN, V., PHILLIPS, J. H., LANIER, L. L. & PARHAM, P. 1995. The Bw4 public epitope of HLA-B molecules confers reactivity with natural killer cell clones that express NKB1, a putative HLA receptor. *The Journal of Experimental Medicine,* 181**,** 1133-44.

HANSSON, G. K., LIBBY, P., SCHONBECK, U. & YAN, Z. Q. 2002. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ Res,* 91**,** 281-91.

HEINRICHS, H. & ORR, H. T. 1990. HLA non-A,B,C class I genes: Their structure and expression. *Immunologic Research,* 9**,** 265-274.

HERBECK, J., GHORAI, S., CHEN, L., RINALDO, C. R., MARGOLICK, J. B., DETELS, R., JACOBSON, L., WOLINSKY, S. & MULLINS, J. I. 2015. p21(WAF1/CIP1) RNA Expression in Highly HIV-1 Exposed, Uninfected Individuals. *PLoS ONE,* 10**,** e0119218.

HINZ, A., MIGUET, N., NATRAJAN, G., USAMI, Y., YAMANAKA, H., RENESTO, P., HARTLIEB, B., MCCARTHY, A. A., SIMORRE, J.-P., GOTTLINGER, H. & WEISSENHORN, W. 2010. Structural basis of HIV-1 tethering to membranes by the Bst2/tetherin ectodomain. *Cell host & microbe,* 7**,** 314-323.

HRECKA, K., HAO, C., GIERSZEWSKA, M., SWANSON, S. K., KESIK-BRODACKA, M., SRIVASTAVA, S., FLORENS, L., WASHBURN, M. P. & SKOWRONSKI, J. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature,* 474**,** 658-661.

HUNT, P. W., BRENCHLEY, J., SINCLAIR, E., MCCUNE, J. M., ROLAND, M., PAGE-SHAFER, K., HSUE, P., EMU, B., KRONE, M., LAMPIRIS, H., DOUEK, D., MARTIN, J. N. & DEEKS, S. G. 2008. Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy. *J Infect Dis,* 197**,** 126-33.

IVANOVSKA, I., BALL, A. S., DIAZ, R. L., MAGNUS, J. F., KIBUKAWA, M., SCHELTER, J. M., KOBAYASHI, S. V., LIM, L., BURCHARD, J., JACKSON, A. L., LINSLEY, P. S. & CLEARY, M. A. 2008. MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol Cell Biol,* 28**,** 2167-74.

JASCUR, T., BRICKNER, H., SALLES-PASSADOR, I., BARBIER, V., EL KHISSIIN, A., SMITH, B., FOTEDAR, R. & FOTEDAR, A. 2005. Regulation of p21(WAF1/CIP1) stability by WISp39, a Hsp90 binding TPR protein. *Mol Cell,* 17**,** 237-49.

JUVEN-GERSHON, T., HSU, J.-Y., THEISEN, J. W. M. & KADONAGA, J. T. 2008. The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology,* 20**,** 253-259.

KATLAMA, C., DEEKS, S. G., AUTRAN, B., MARTINEZ-PICADO, J., VAN LUNZEN, J., ROUZIOUX, C., MILLER, M., VELLA, S., SCHMITZ, J. E., AHLERS, J., RICHMAN, D. D. & SEKALY, R. P. 2013. Barriers to a Cure: New concepts in targeting and eradicating HIV-1 reservoirs. *Lancet,* 381**,** 10.1016/S0140-6736(13)60104-X.

KAWAMURA, T., SUZUKI, J., WANG, Y. V., MENENDEZ, S., MORERA, L. B., RAYA, A., WAHL, G. M. & IZPISUA BELMONTE, J. C. 2009. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature,* 460**,** 1140-4.

KELLEY, C. F., BARBOUR, J. D. & HECHT, F. M. 2007. The relation between symptoms, viral load, and viral load set point in primary HIV infection. *J Acquir Immune Defic Syndr,* 45**,** 445-8.

KIEPIELA, P., LESLIE, A. J., HONEYBORNE, I., RAMDUTH, D., THOBAKGALE, C., CHETTY, S., RATHNAVALU, P., MOORE, C., PFAFFEROTT, K. J., HILTON, L., ZIMBWA, P., MOORE, S., ALLEN, T., BRANDER, C., ADDO, M. M., ALTFELD, M., JAMES, I., MALLAL, S., BUNCE, M., BARBER, L. D., SZINGER, J., DAY, C., KLENERMAN, P., MULLINS, J., KORBER, B., COOVADIA, H. M., WALKER, B. D. & GOULDER, P. J. 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature,* 432**,** 769-75.

KIRCHHOFF, F. 2010. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host Microbe,* 8**,** 55-67.

KOEHLER, R. N., SAATHOFF, E., PRITSCH, M., GELDMACHER, C., HOFFMANN, O., MABOKO, L., MAGANGA, L., GEIS, S., MCCUTCHAN, F. E., KIJAK, G. H., KIM, J. H., ARROYO, M. A., GERHARDT, M., TOVANABUTRA, S., ROBB, M. L., WILLIAMSON, C., MICHAEL, N. L. & HOELSCHER, M. 2010. Viral and host factors associated with the HIV-1 viral load setpoint in adults from Mbeya Region, Tanzania. *J Acquir Immune Defic Syndr,* 54**,** 324-30.

KORTHAGEN, N. M., VAN MOORSEL, C. H., BARLO, N. P., KAZEMIER, K. M., RUVEN, H. J. & GRUTTERS, J. C. 2012. Association between variations in cell cycle genes and idiopathic pulmonary fibrosis. *PLoS One,* 7**,** e30442.

KUHL, B., SLOAN, R., DONAHUE, D., BAR-MAGEN, T., LIANG, C. & WAINBERG, M. 2010. Tetherin restricts direct cell-to-cell infection of HIV-1. *Retrovirology,* 7**,** 115.

LAGUETTE, N., RAHM, N., SOBHIAN, B., CHABLE-BESSIA, C., MÜNCH, J., SNOECK, J., SAUTER, D., SWITZER, W. M., HENEINE, W., KIRCHHOFF, F., DELSUC, F., TELENTI, A. & BENKIRANE, M. 2012. Evolutionary and Functional Analyses of the Interaction between the Myeloid Restriction Factor SAMHD1 and the Lentiviral Vpx Protein. *Cell host & microbe,* 11**,** 205-217.

LAGUETTE, N., SOBHIAN, B., CASARTELLI, N., RINGEARD, M., CHABLE-BESSIA, C., SÉGÉRAL, E., YATIM, A., EMILIANI, S., SCHWARTZ, O. & BENKIRANE, M. 2011. SAMHD1 is the dendritic– and myeloid–cell–specific HIV–1 restriction factor counteracted by Vpx. *Nature,* 474**,** 654-657.

LAHOUASSA, H., DADDACHA, W., HOFMANN, H., AYINDE, D., LOGUE, E. C., DRAGIN, L., BLOCH, N., MAUDET, C., BERTRAND, M., GRAMBERG, T., PANCINO, G., PRIET, S., CANARD, B., LAGUETTE, N., BENKIRANE, M., TRANSY, C., LANDAU, N. R., KIM, B. & MARGOTTIN-GOGUET, F. 2012. SAMHD1 restricts HIV-1 by reducing the intracellular pool of deoxynucleotide triphosphates. *Nature immunology,* 13**,** 223-228.

LAVREYS, L., BAETEN, J. M., CHOHAN, V., MCCLELLAND, R. S., HASSAN, W. M., RICHARDSON, B. A., MANDALIYA, K., NDINYA-ACHOLA, J. O. & OVERBAUGH, J. 2006. Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. *Clin Infect Dis,* 42**,** 1333-9.

LI, G., LIU, Z., STURGIS, E. M., SHI, Q., CHAMBERLAIN, R. M., SPITZ, M. R. & WEI, Q. 2005. Genetic polymorphisms of p21 are associated with risk of squamous cell carcinoma of the head and neck. *Carcinogenesis,* 26**,** 1596-602.

LI, R., WAGA, S., HANNON, G. J., BEACH, D. & STILLMAN, B. 1994. Differential effects by the p21 CDK inhibitor on PCNA-dependent DNA replication and repair. *Nature,* 371**,** 534-7.

LIU, F., LI, B., WEI, Y., CHEN, X., MA, Y., YAN, L. & WEN, T. 2011. P21 codon 31 polymorphism associated with cancer among white people: evidence from a meta-analysis involving 78,074 subjects. *Mutagenesis,* 26**,** 513-21.

LU, J., GETZ, G., MISKA, E. A., ALVAREZ-SAAVEDRA, E., LAMB, J., PECK, D., SWEET-CORDERO, A., EBERT, B. L., MAK, R. H., FERRANDO, A. A., DOWNING, J. R., JACKS, T., HORVITZ, H. R. & GOLUB, T. R. 2005. MicroRNA expression profiles classify human cancers. *Nature,* 435**,** 834-8.

MA, Y., ZHANG, Y., LIN, L., GUO, X., WU, Y., WEN, W., LI, F. & LIU, P. 2013. Quantitative assessment of the relationship between p21 Ser31Arg polymorphism and cervical cancer. *Tumour Biol,* 34**,** 3887-92.

MABALA, R. 2006. From HIV prevention to HIV protection: addressing the vulnerability of girls and young women in urban areas. *Environment and Urbanization,* 18**,** 407-432.

MACLEOD, K. F., SHERRY, N., HANNON, G., BEACH, D., TOKINO, T., KINZLER, K., VOGELSTEIN, B. & JACKS, T. 1995. p53-dependent and independent expression of p21 during cell growth, differentiation, and DNA damage. *Genes Dev,* 9**,** 935-44.

MARTIN, M. P., GAO, X., LEE, J. H., NELSON, G. W., DETELS, R., GOEDERT, J. J., BUCHBINDER, S., HOOTS, K., VLAHOV, D., TROWSDALE, J., WILSON, M., O'BRIEN, S. J. & CARRINGTON, M. 2002. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet,* 31**,** 429-34.

MEDZHITOV, R. & JANEWAY, C., JR. 2000. Innate immune recognition: mechanisms and pathways. *Immunol Rev,* 173**,** 89-97.

MELLORS, J. W., MUNOZ, A., GIORGI, J. V., MARGOLICK, J. B., TASSONI, C. J., GUPTA, P., KINGSLEY, L. A., TODD, J. A., SAAH, A. J., DETELS, R., PHAIR, J. P. & RINALDO, C. R., JR. 1997. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med,* 126**,** 946-54.

MERRITT, C., RASOLOSON, D., KO, D. & SEYDOUX, G. 2008. 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Current biology : CB,* 18**,** 1476-1482.

MIDDLETON, D., CURRAN, M. & MAXWELL, L. 2002. Natural killer cells and their receptors. *Transpl Immunol,* 10**,** 147-64.

MIJIYAWA, M., ONIANKITAN, O. & KHAN, M. A. 2000. Spondyloarthropathies in sub-Saharan Africa. *Curr Opin Rheumatol,* 12**,** 281-6.

MULHERIN, S. A., O'BRIEN, T. R., IOANNIDIS, J. P., GOEDERT, J. J., BUCHBINDER, S. P., COUTINHO, R. A., JAMIESON, B. D., MEYER, L., MICHAEL, N. L., PANTALEO, G., RIZZARDI, G. P., SCHUITEMAKER, H., SHEPPARD, H. W., THEODOROU, I. D., VLAHOV, D., ROSENBERG, P. S. & INTERNATIONAL META-ANALYSIS OF, H. I. V. H. G. 2003. Effects of CCR5-Delta32 and CCR2-64I alleles on HIV-1 disease progression: the protection varies with duration of infection. *AIDS,* 17**,** 377-87.

MUMMIDI, S., AHUJA, S. S., GONZALEZ, E., ANDERSON, S. A., SANTIAGO, E. N., STEPHAN, K. T., CRAIG, F. E., O'CONNELL, P., TRYON, V., CLARK, R. A., DOLAN, M. J. & AHUJA, S. K. 1998. Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. *Nat Med,* 4**,** 786-93.

NISOLE, S. & SAIB, A. 2004. Early steps of retrovirus replicative cycle. *Retrovirology,* 1**,** 9.

NYAMWEYA, S., HEGEDUS, A., JAYE, A., ROWLAND-JONES, S., FLANAGAN, K. L. & MACALLAN, D. C. 2013. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Reviews in Medical Virology,* 23**,** 221-240.

OKULICZ, J. F. 2012. Elite Controllers and Long-term Nonprogressors: Models for HIV Vaccine Development?. *J AIDS Clinic Res,* 3**,** 1-4.

OKULICZ, J. F., MARCONI, V. C., LANDRUM, M. L., WEGNER, S., WEINTROB, A., GANESAN, A., HALE, B., CRUM-CIANFLONE, N., DELMAR, J., BARTHEL, V., QUINNAN, G., AGAN, B. K., DOLAN, M. J. & INFECTIOUS DISEASE CLINICAL RESEARCH PROGRAM, H. I. V. W. G. 2009. Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study. *J Infect Dis,* 200**,** 1714-23.

OSTROWSKI, M. A., JUSTEMENT, S. J., CATANZARO, A., HALLAHAN, C. A., EHLER, L. A., MIZELL, S. B., KUMAR, P. N., MICAN, J. A., CHUN, T.-W. & FAUCI, A. S. 1998. Expression of Chemokine Receptors CXCR4 and CCR5 in HIV-1-Infected and Uninfected Individuals. *The Journal of Immunology,* 161**,** 3195-3201.

PAULS, E., RUIZ, A., RIVEIRA-MUÑOZ, E., PERMANYER, M., BADIA, R., CLOTET, B., KEPPLER, O. T., BALLANA, E. & ESTE, J. A. 2014. p21 regulates the HIV-1 restriction factor SAMHD1. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** E1322-E1324.

PAXIMADIS, M., MATHEBULA, T. Y., GENTLE, N. L., VARDAS, E., COLVIN, M., GRAY, C. M., TIEMESSEN, C. T. & PUREN, A. 2012. Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. *Hum Immunol,* 73**,** 80-92.

PELAK, K., NEED, A. C., FELLAY, J., SHIANNA, K. V., FENG, S., URBAN, T. J., GE, D., DE LUCA, A., MARTINEZ-PICADO, J., WOLINSKY, S. M., MARTINSON, J. J., JAMIESON, B. D., BREAM, J. H., MARTIN, M. P., BORROW, P., LETVIN, N. L., MCMICHAEL, A. J., HAYNES, B. F., TELENTI, A., CARRINGTON, M., GOLDSTEIN, D. B., ALTER, G. & IMMUNOLOGY, N. C. F. H. A. V. 2011. Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol,* 9**,** e1001208.

PEREYRA, F., ADDO, M. M., KAUFMANN, D. E., LIU, Y., MIURA, T., RATHOD, A., BAKER, B., TROCHA, A., ROSENBERG, R., MACKEY, E., UEDA, P., LU, Z., COHEN, D., WRIN, T., PETROPOULOS, C. J., ROSENBERG, E. S. & WALKER, B. D. 2008. Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis,* 197**,** 563-71.

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

POOLE, A. J., HEAP, D., CARROLL, R. E. & TYNER, A. L. 2004. Tumor suppressor functions for the Cdk inhibitor p21 in the mouse colon. *Oncogene,* 23**,** 8128-34.

PRADO, J. G., CARRILLO, J., BLANCO-HEREDIA, J. & BRANDER, C. 2011. Immune correlates of HIV control. *Curr Med Chem,* 18**,** 3963-70.

QI, F., HUANG, M., PAN, Y., LIU, Y., LIU, J., WEN, J., XIE, K., SHEN, H., MA, H., MIAO, Y. & HU, Z. 2014. A genetic variant in the promoter region of miR-106b-25 cluster predict clinical outcome of HBV-related hepatocellular carcinoma in Chinese. *PLoS One,* 9**,** e85394.

QI, Y., MARTIN, M. P., GAO, X., JACOBSON, L., GOEDERT, J. J., BUCHBINDER, S., KIRK, G. D., O'BRIEN, S. J., TROWSDALE, J. & CARRINGTON, M. 2006. KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathog,* 2**,** e79.

QIN, L., GILBERT, P. B., COREY, L., MCELRATH, M. J. & SELF, S. G. 2007. A Framework for Assessing Immunological Correlates of Protection in Vaccine Trials. *Journal of Infectious Diseases,* 196**,** 1304-1312.

RICHARDSON, B. A., MBORI-NGACHA, D., LAVREYS, L., JOHN-STEWART, G. C., NDUATI, R., PANTELEEFF, D. D., EMERY, S., KREISS, J. K. & OVERBAUGH, J. 2003. Comparison of human immunodeficiency virus type 1 viral loads in Kenyan women, men, and infants during primary and early infection. *J Virol,* 77**,** 7120-3.

RIVEIRA-MUNOZ, E., RUIZ, A., PAULS, E., PERMANYER, M., BADIA, R., MOTHE, B., CRESPO, M., CLOTET, B., BRANDER, C., BALLANA, E. & ESTE, J. A. 2014. Increased expression of SAMHD1 in a subset of HIV-1 elite controllers. *J Antimicrob Chemother,* 69**,** 3057-60.

ROH, J. W., KIM, J. W., PARK, N. H., SONG, Y. S., PARK, I. A., PARK, S. Y., KANG, S. B. & LEE, H. P. 2004. p53 and p21 genetic polymorphisms and susceptibility to endometrial cancer. *Gynecol Oncol,* 93**,** 499-505.

RONINSON, I. B. 2002. Oncogenic functions of tumour suppressor p21(Waf1/Cip1/Sdi1): association with cell senescence and tumour-promoting activities of stromal fibroblasts. *Cancer Lett,* 179**,** 1-14.

RUDORF, D. C. & KRIKORIAN, S. A. 2005. Adverse Effects Associated With Antiretroviral Therapy and Potential Management Strategies. *Journal of Pharm Prac,* 18**,** 258-277.

RYOO, J., CHOI, J., OH, C., KIM, S., SEO, M., KIM, S.-Y., SEO, D., KIM, J., WHITE, T. E., BRANDARIZ-NUNEZ, A., DIAZ-GRIFFERO, F., YUN, C.-H., HOLLENBAUGH, J. A., KIM, B., BAEK, D. & AHN, K. 2014. The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat Med,* 20**,** 936-941.

SABETI, P. C., WALSH, E., SCHAFFNER, S. F., VARILLY, P., FRY, B., HUTCHESON, H. B., CULLEN, M., MIKKELSEN, T. S., ROY, J., PATTERSON, N., COOPER, R., REICH, D., ALTSHULER, D., O'BRIEN, S. & LANDER, E. S. 2005. The Case for Selection at CCR5-Δ32. *PLoS Biology,* 3**,** e378.

SÁEZ-CIRIÓN, A., HAMIMI, C., BERGAMASCHI, A., DAVID, A., VERSMISSE, P., MÉLARD, A., BOUFASSA, F., BARRÉ-SINOUSSI, F., LAMBOTTE, O., ROUZIOUX, C., PANCINO, G. & FOR THE, A. C. O. C. 2011. Restriction of HIV-1 replication in macrophages and CD4(+) T cells from HIV controllers. *Blood,* 118**,** 955-964.

SAEZ-CIRION, A., PANCINO, G., SINET, M., VENET, A., LAMBOTTE, O. & GROUP, A. E. H. C. S. 2007. HIV controllers: how do they tame the virus? *Trends Immunol,* 28**,** 532-40.

SALERNO, D., HASHAM, M. G., MARSHALL, R., GARRIGA, J., TSYGANKOV, A. Y. & GRAÑA, X. 2007. Direct inhibition of CDK9 blocks HIV-1 replication without preventing T cell activation in primary human peripheral blood lymphocytes. *Gene,* 405**,** 65-78.

SANTA-MARTA, M., DE BRITO, P. M., GODINHO-SANTOS, A. & GONCALVES, J. 2013. Host Factors and HIV-1 Replication: Clinical Evidence and Potential Therapeutic Approaches. *Frontiers in Immunology,* 4**,** 343.

SERRAO, E., WANG, C.-H., FREDERICK, T., LEE, C.-L., ANTHONY, P., ARRIBAS-LAYTON, D., BAKER, K., MILLSTEIN, J., KOVACS, A. & NEAMATI, N. 2014. Alteration of select gene expression patterns in individuals infected with HIV-1. *Journal of Medical Virology,* 86**,** 678-686.

SIVONOVA, M. K., VILCKOVA, M., JURECEKOVA, J., HATOK, J., DOBROTA, D., DUSENKA, R. & KLIMENT, J. 2013. The role of p21 3'UTR gene polymorphism in the risk of prostate cancer: a pilot study. *Mol Med Rep,* 7**,** 986-90.

STIER, S., CHENG, T., FORKERT, R., LUTZ, C., DOMBKOWSKI, D. M., ZHANG, J. L. & SCADDEN, D. T. 2003. Ex vivo targeting of p21Cip1/Waf1 permits relative expansion of human hematopoietic stem cells. *Blood,* 102**,** 1260-6.

TIEMESSEN, C. T. & MARTINSON, N. 2012. *Elite controllers: Understanding natural suppressive control of HIV-1 infection*.

TOMARAS, G. D. & HAYNES, B. F. 2014. Advancing Toward HIV-1 Vaccine Efficacy through the Intersections of Immune Correlates. *Vaccines (Basel),* 2**,** 15-35.

VENKATESH, S. & BIENIASZ, P. D. 2013. Mechanism of HIV-1 Virion Entrapment by Tetherin. *PLoS Pathog,* 9**,** e1003483.

WALKER, B. D. & YU, X. G. 2013. Unravelling the mechanisms of durable control of HIV-1. *Nat Rev Immunol,* 13**,** 487-498.

WANG, N., WANG, S., ZHANG, Q., LU, Y., WEI, H., LI, W., ZHANG, S., YIN, D. & OU, Y. 2012a. Association of p21 SNPs and risk of cervical cancer among Chinese women. *BMC Cancer,* 12**,** 589.

WANG, X., AO, Z., CHEN, L., KOBINGER, G., PENG, J. & YAO, X. 2012b. The Cellular Antiviral Protein APOBEC3G Interacts with HIV-1 Reverse Transcriptase and Inhibits Its Function during Viral Replication. *Journal of Virology,* 86**,** 3777-3786.

WU, X., CONWAY, J. A., KIM, J. & KAPPES, J. C. 1994. Localization of the Vpx packaging signal within the C terminus of the human immunodeficiency virus type 2 Gag precursor protein. *Journal of Virology,* 68**,** 6161-6169.

YAMASHITA, M. & EMERMAN, M. 2006. Retroviral infection of non-dividing cells: Old and new perspectives. *Virology,* 344**,** 88-93.

YAN, N. & CHEN, Z. J. 2012. Intrinsic antiviral immunity. *Nat Immunol,* 13**,** 214-222.

YAWATA, M., YAWATA, N., ABI-RACHED, L. & PARHAM, P. 2002. Variation within the human killer cell immunoglobulin-like receptor (KIR) gene family. *Crit Rev Immunol,* 22**,** 463-82.

YOSHIDA, I., OKA, K., HIDAJAT, R., NAGANO-FUJII, M., ISHIDO, S. & HOTTA, H. 2001. Inhibition of p21/Waf1/Cip1/Sdi1 expression by hepatitis C virus core protein. *Microbiol Immunol,* 45**,** 689-97.

ZHANG, J., SCADDEN, D. T. & CRUMPACKER, C. S. 2007. Primitive hematopoietic cells resist HIV-1 infection via p21. *J Clin Invest,* 117**,** 473-81.

APPENDIX A: ETHICS CLEARANCE FORM

R14/49 Ms Gemma Koor and Dr Maria Paximadis

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M140996

NAME: (Principal Investigator)	Ms Gemma Koor and Dr Maria Paximadis
DEPARTMENT:	Virology Faculty of Health Sciences National Institute for Communicable Diseases National health Laboratory Services
PROJECT TITLE:	Genetic Variation and Differential Expression of p21 (WAF1/cip 1) in the Context of HIV-1 Control
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Caroline Tiemessen
APPROVED BY:	Professor Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	29/10/2014

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

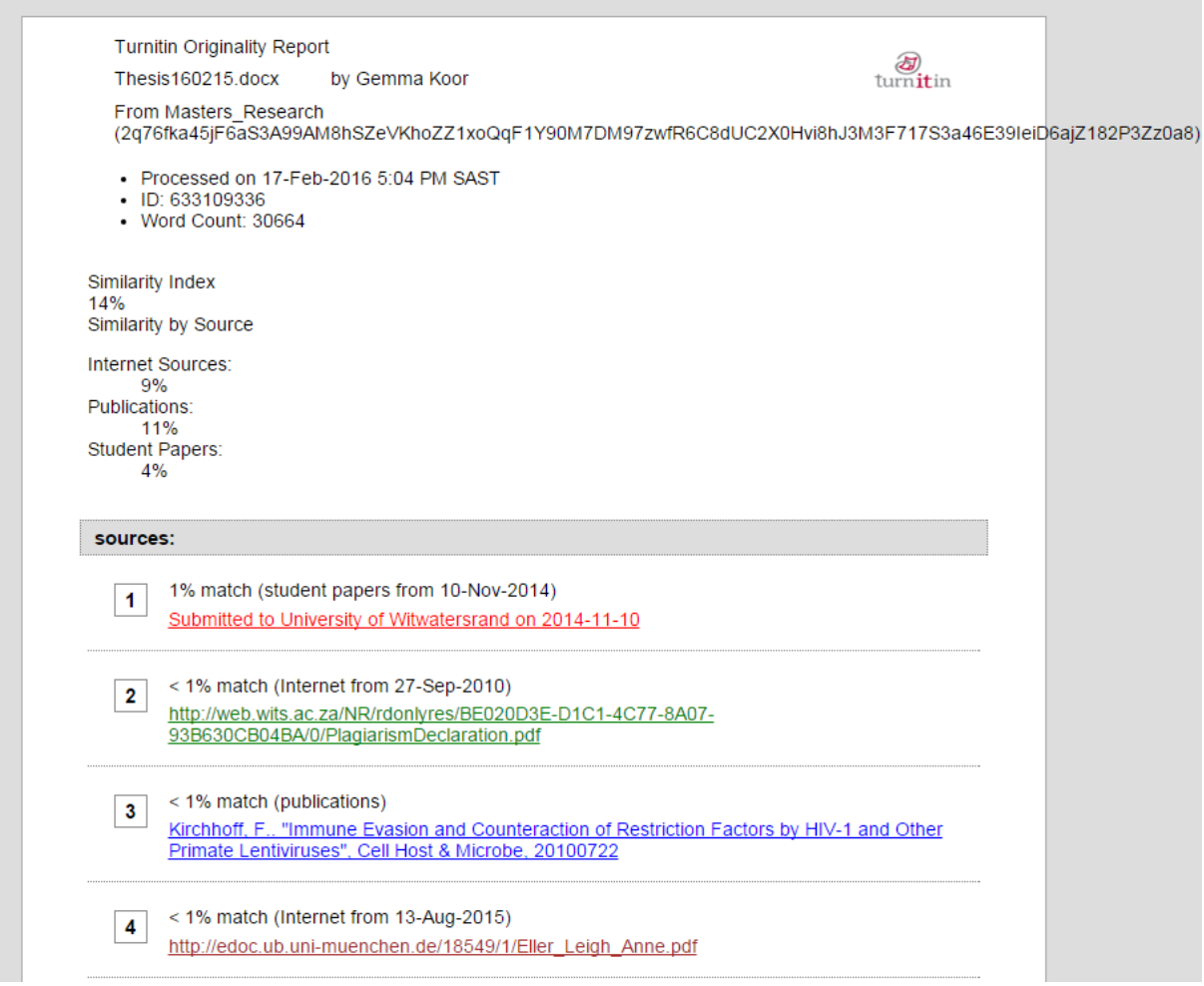
To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

APPENDIX B: TURNITIN REPORT



Turnitin Originality Report

Thesis160215.docx by Gemma Koor

From Masters_Research
(2q76fka45jF6aS3A99AM8hSZeVKhoZZ1xoQqF1Y90M7DM97zwfR6C8dUC2X0Hvi8hJ3M3F717S3a46E39leiD6ajZ182P3Zz0a8)

- Processed on 17-Feb-2016 5:04 PM SAST
- ID: 633109336
- Word Count: 30664

Similarity Index
14%
Similarity by Source

Internet Sources:
9%
Publications:
11%
Student Papers:
4%

sources:

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

2 < 1% match (Internet from 27-Sep-2010)
http://web.wits.ac.za/NR/rdonlyres/BE020D3E-D1C1-4C77-8A07-93B630CB04BA/0/PlagiarismDeclaration.pdf

3 < 1% match (publications)
Kirchhoff, F., "Immune Evasion and Counteraction of Restriction Factors by HIV-1 and Other Primate Lentiviruses", Cell Host & Microbe, 20100722

4 < 1% match (Internet from 13-Aug-2015)
http://edoc.ub.uni-muenchen.de/18549/1/Eller_Leigh_Anne.pdf