Functional characterization of human immunodeficiency virus type 1 (HIV-1) subtype C transmitted/founder (T/F) viruses long terminal repeat (LTR) variants and association with disease outcome

Shamara Naicker

Supervisor: Dr Paradise Madlala

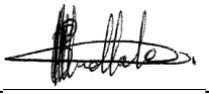
Submitted in fulfilment of the requirements for the degree of Master of Medical Science in the Department of Molecular Virology, School of Laboratory Medicine and Medical Sciences, Faculty of Health Science, Nelson R. Mandela School of Medicine at the University of KwaZulu-Natal.

2020

PREFACE

The experimental procedures presented in this thesis was performed at the Hasso Plattner Research Laboratory, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa. This study was performed under the supervision of Dr. Paradise Z. Madlala. This dissertation represents the original work of the author and has not been submitted for any degree or examination at any other university. Where the work of others has been used, the authors have been duly acknowledged.

Shamara Naicker (Student)



Dr. Paradise Z. Madlala (Supervisor)

PLAGIARISM DECLARATION

I, Shamara Naicker declare that:

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

(ii) This thesis has not been submitted for any degree or examination at any other university.

(iii) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

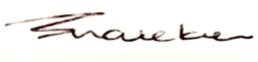
(iv) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

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

b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

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

(vi) This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.



06/03/2020

Signed Date

ETHICS APPROVAL

The ethics approval for this study was obtained from the Biomedical Research Committee of the Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa. The ethics approval reference number is as follows: BREC/00001051/2020.

POSTER PRESENTATIONS

1. Shamara Naicker, Paradise Madlala. Functional characterization of human immunodeficiency virus type 1 (HIV-1) subtype C transmitted/founder (T/F) viruses long terminal repeat (LTR) variants and association with disease outcome. College of Health Science Annual Research Symposium, 1st November 2019, Nelson R. Mandela School of Medicine, Durban (South Africa).

DEDICATION

To Jesus Christ, my saviour, through whom all things are made possible. This thesis is dedicated to my loving parents, Suraya Naicker and Alex Naicker. Thank you for your strong faith, unconditional love and constant support.

ACKNOWLEDGEMENTS

Sincere thanks to my supervisor, Dr. Paradise Madlala (HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, South Africa) for his dedication, encouragement and support towards this project. I express my sincere appreciation for having been given this opportunity to complete this degree at the HIV Pathogenesis Programme under the supervision of a committed and motivated supervisor.

Special thanks to Ms. Zakithi Mkhize (PhD Candidate) for providing me with the recombinant autologous Tat plasmids and autologous Tat mutation data used in this experimental study. I am also grateful for her guidance and training in the lab.

The staff and students of the HIV Pathogenesis Programme for their assistance and training.

Participants of the HIV Pathogenesis Acute Infection Cohort and Females Rising through Education, Support and Health (FRESH) cohort for their samples.

Special thanks to my parents, Suraya Naicker and Alex Naicker for their unconditional love and support that has given me the will to persevere. Sincere gratitude to my sister Ranaicia Nundlall, thank you for your faith in me, your dedication, motivation and support both financially and emotionally. To my best friend, Craig Singh, you are highly appreciated for having always believed in me since day one.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS: Acquired immune deficiency syndrome

ART: Antiretroviral therapy

cART: Combination antiretroviral therapy

CCD: Catalytic core domain

cDNA: complementary Deoxyribonucleic Acid

CRF: Circulating recombinant form

CTD: C-terminal domain

DMT: DNA methyltransferase

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphate

E-box: Enhancer box

EDTA: Ethylenediaminetetraacetic acid

FRESH: Females Rising Through Education, Support and Health

H3/ H4: Histone 3/ Histone 4

HAT: Histone acetyltransferase

HDAC: Histone deacetylase

HIV-1: Human immunodeficiency virus type 1

HIV-2: Human immunodeficiency virus type 2

HPP: HIV Pathogenesis Programme

HTLV-III: Human T-cell lymphotropic virus

IgM: Immunoglobulin M

IN: Integrase

IST: Inducer of short transcripts

LAV: Lymph Adenopathy virusLTR: Long terminal repeat

MBD2: Methyl Binding CpG Domain-2

mRNA: Messenger Ribonucleic Acid

NaOAc: Sodium Acetate

NELF: Negative elongation factor

NFAT: Nuclear factor of Transcription

ng: Nanogram

NPC: Nuclear Pore Complex

NRE: Negative regulatory element

Nuc-0: Nucleosome-0

Nuc-1: Nucleosome-1

PBS: Primer binding site

PCR: Polymerase chain reaction

PIC: Pre-integration Complex

PLHIV: People living with HIV

PMA: Phorbol 12-myristate 13-acetate

PPT: Polypurine tract

P-TEFb: Positive transcription elongation factor-b

R: Repeat region

RLU: Relative light unit

RNA: Ribonucleic acid

RNA pol II: RNA Polymerase II

RRE: Rev response element

RT: Reverse Transcriptase

RTC: Reverse transcription complex

SAHA: Suberoylanilide hydroxamic acid

SIV: Simian immunodeficiency virus

Tat: Transactivator of Transcription

TBP: Tata Binding Protein

T/F: Transmitted/Founder

TFs: Transcription factors

TFBS: Transcription factor binding site

TNF-α: Tumour necrosis factor-α

U3 region: Unique 3 region

U5: Unique 5 region

U3R- Unique 3 Repeat region

µL: Microlitre

UNAIDS: The Joint United Nations Programme on HIV/AIDS

ABSTRACT

Background: The persistence of latent viral reservoirs is a major roadblock to human immunodeficiency virus type 1 (HIV-1) cure development. Latent reservoirs harbour transcriptionally silent yet replication competent proviruses. However, the molecular mechanisms that govern HIV-1 latency at the transcriptional level is unknown. Therefore, we hypothesize that HIV-1 subtype C (HIV-1C) transmitted/founder (T/F) 5' long terminal repeat (LTR) genetic variation may affect disease outcome.

Methods: To address this, viral RNA was extracted from plasma samples obtained from 25 HIV-1 infected patients from the HPP and FRESH acute infection cohorts (QIAamp® Viral RNA Mini Kit, Qiagen, Hilden, Germany). Viral RNA was reverse transcribed to DNA using SuperScript™ III One Step RT-PCR System with Platinum™ Taq DNA Polymerase (Invitrogen, Massachusetts, United States). Nested PCR was performed (Platinum® Taq DNA Polymerase High Fidelity PCR Kit (Invitrogen, Massachusetts, United States) and PCR products cloned into the pGL3 Basic plasmid. LTR/pGL3 recombinant plasmids were sequenced using BigDye Terminator v3.1 Sequencing Kit (Invitrogen, Massachusetts, United States) to confirm correct sequences. The LTR-pGl3 recombinant plasmids were transfected into Jurkat cells alone or co-transfected with either consensus (wild type) subtype C Tat (conTat) or autologous *tat* (autoTat) to determine the effect of LTR genetic variation on expression of a luciferase reporter gene.

Results: Interestingly, our data demonstrate that basal transcription activity significantly differs between LTR variants. Specifically, patients harbouring the Sp1 III: G2A mutation demonstrated significantly lower transcription compared to the wild type LTR. Although conTat co-transfection increased the LTR activity for most of the LTR variants, the T/F virus LTR containing the TATA box mutation (TATAA → TA**A**AA) in combination with other LTR mutations was not induced. Interestingly, the transactivation activity of the autologous Tat was variable among patients. Specifically, the TATA box variant was marginally induced. Lastly, we observed that the majority of LTR variants were more responsive to stimulation by PMA as compared to TNF-α, SAHA and prostratin. Interestingly, our data demonstrate that autologous *tat* induced transcription positively correlated with viral load at transmission ($p=0.0134$, $r=0.66$) and at one-year post infection but was not significant ($p=0.3905$, $r=0.26$).

Conclusion: These data suggest that the TATAA → TA**A**AA mutation in combination with other LTR mutations may reduce transcription activity. Taken together our data suggest that HIV-1 subtype C T/F viruses LTR genetic variation may modulate viral gene transcription and impact disease outcome.

CHAPTER 1: INTRODUCTION

Since its discovery as the causative agent of acquired immunodeficiency syndrome (AIDS) in 1983 (Barré-Sinoussi et al., 1983), human immunodeficiency virus (HIV) has become a global health pandemic. The recently published Joint United Nations Programme on HIV/AIDS (UNAIDS) 2019 report documented that there were approximately 37.9 million individuals living with HIV, 1.7 million new HIV infections, and 770 000 AIDS-related deaths globally in 2018 (UNAIDS, 2019).

The introduction of combination antiretroviral therapy (cART) has significantly reduced morbidity and mortality associated with HIV-1 infection (Collaboration, 2010, Perelson et al., 1997, Trickey et al., 2017, UNAIDS, 2019). Specifically, cART successfully suppresses viremia to even below clinically detectable levels (20-50 copies HIV-1 RNA/mL plasma) in people living with HIV (PLHIV) who have access and are adherent to treatment (Gulick et al., 1997, Hammer et al., 1997, Perelson et al., 1997). Despite the ability of cART to suppress HIV replication, treatment interruption results in viral rebound, thus cART is unable to eradicate or cure HIV-1 infection (Chun et al., 2010, Davey et al., 1999). The failure of cART to cure HIV-1 infection is due to the establishment of stable HIV-1 latent reservoirs early in acute infection (Chun et al., 1998, Chun et al., 1997, Finzi et al., 1997, Siliciano et al., 2003b). HIV-1 latent reservoirs comprise a subset of resting memory $CD4^+$ T cells that are infected with replication competent and yet transcriptionally silent HIV-1 proviruses (Chun et al., 1997, Finzi et al., 1999, Persaud et al., 2000, Siliciano et al., 2003a). Therefore, latent reservoir remains the primary obstacle in the development of an effective HIV-1 cure.

However, the mechanisms that govern HIV-1 latency have not yet been fully elucidated. Interestingly, several mechanisms of latency development have been proposed. Host chromatin compaction around the integrated HIV-1 proviral genome in non-dividing (inactive) cells has been associated with a state of viral latency. The host chromatin is associated with histone proteins that together form a functional unit called nucleosome. The chromatin configuration within the nucleosomes may either promote a state of viral transcription or viral latency. Actively transcribed genes are located within an open or relaxed state of chromatin called euchromatin. However, the heterochromatin configuration (tightly compacted chromatin configuration) is associated with viral latency (Jordan et al., 2003). The heterochromatin configuration prevents the transcription factors (TFs) from accessing their respective transcription factor binding sites (TFBS). Therefore, this imposes a block on active gene transcription. However, this form of viral latency is reversed when infected cells become activated by

varying stimuli such as phorbol esters (eg. Phorbol myristate acetate) and tumor necrosis-alpha (TNF-α) (Verdin et al., 1993). Viral latency is further enforced by two nucleosomes, nuc-1 and nuc-0 that are deposited on the integrated proviral LTR (Rafati et al., 2011, Van Lint et al., 1996, Verdin et al., 1993). Particularly, nuc-1 is positioned near the transcription start site and is located such that it impairs RNA polymerase II (RNA pol II) elongation (Verdin et al., 1993). The positioning of nuc-1 exerts a highly repressive effect on LTR transcription but this nucleosome is particularly disrupted during proviral LTR activation (Verdin et al., 1993).

Epigenetic modification such as histone acetylation and methylation significantly alter the chromatin and transcriptional state of the proviral genome. Hypoacetylation of the histone tails are associated with the latent proviral LTR (Coull et al., 2000, Jiang et al., 2007, Tyagi and Karn, 2007, Williams et al., 2006). The recruitment of histone acetyltransferases (HATs) results in acetylation of the lysine residues of histone 3 (H3) and histone 4 (H4). The significance of the hyperacetylation of the lysine residues of H3 and H4 is noted by the inhibition of histone deacetyltransferases [(HDACs) enzymes that remove acetyl groups from the histone tails] that significantly increases the reactivation potential of the latent proviral LTR (Archin et al., 2014, Archin et al., 2009, Shirakawa et al., 2013). In addition, methylation of the lysine residues of H3 are associated with repression of the proviral LTR. Specifically, the trimethylation of H3 lysine residue 9 and lysine residue 27 induces proviral latency (du Chéné et al., 2007, Friedman et al., 2011, Kim et al., 2011, Mateescu et al., 2008).

A third mechanism that promotes proviral latency includes DNA methylation. The HIV-1 proviral transcription start site is flanked by two CpG islands that when methylated, induce transcriptional silencing of the proviral 5'LTR (Blazkova et al., 2009, Kauder et al., 2009). Methylated CpG islands prevents the interaction between TFs such as NF-κB and Sp1 and their respective binding sites (Bednarik et al., 1991). One of the two CpG islands further contain a methyl binding domain protein-2 (MBD2) that further represses transcription by the proviral 5'-LTR (Kauder et al., 2009). A fourth mechanism of viral latency is transcriptional interference. Transcriptional interference may occur via two mechanisms based on the polarity of the integrated viral genome. When the polarity of the integrated proviral genome is the same as the host genome then interference occurs via an upstream RNA pol II that causes viral promoter occlusion and displaces the transcription factors at the viral promoter during RNA Pol II read-through (Greger et al., 1998, Lenasi et al., 2008). However, when the proviral genome is integrated in the opposite direction to the host genome then RNA pol II transcription complexes originating at different regions within the genome may collide and result in premature transcription termination (Han et al., 2008). This further promotes a state of viral latency. However, the proposed mechanisms of viral latency explain the development of HIV-1 latency in part meaning,

the precise mechanism that modulates HIV-1 latency at the transcriptional level has not yet been fully elucidated. Therefore, the current study focuses on the impact of HIV-1 subtype C (HIV-1C) transmitted/founder (T/F) viruses LTR variation on disease outcome.

The HIV-1 LTR is divided into three regions, the U3, R and U5 regions (reviewed in (Groen and Morris, 2013). The U3 region is composed of the core promoter, enhancer and modulatory regions that play a central role in LTR's ability to drive viral gene transcription (reviewed in (Groen and Morris, 2013). The structure and function of the 5'-LTR will be discussed in detail in the subsequent literature review (Chapter 2). Inter- and intra-subtype 5'-LTR genetic variation has been reported only during chronic infection where the virus has already diversified (Baar et al., 2000, Bachu et al., 2012, Boullosa et al., 2014, Hunt et al., 2001, Jeeninga et al., 2000, Obasa et al., 2019, Scriba et al., 2002). Baar et al., (2000) demonstrated subtype specific differences in the LTRs of subtypes A through G. they demonstrated that subtypes B, D and F exhibited a second CTG motif located downstream of the TCF1-α which was not seen among the other subtypes (Baar et al., 2000). In addition, subtype C specifically exhibited the 3rd NF-κB binding site while subtype E specifically replaces one of the conserved canonical NF-κB binding site with a GABP binding site (Baar et al., 2000). This emphasizes the inter-subtype LTR genetic variation.

Early studies (Hunt et al., 2001, Scriba et al., 2002) had reported that the HIV-1 subtype C circulating within South Africa exhibited the 4th NF-κB at a low frequency. Interestingly, a later study reported that subtype C viruses circulating within India contained the 4th NF-κB binding site and were inferred to be taking over the epidemic within India (Bachu et al., 2012). Furthermore, these viruses with the 4th-NF-κB binding site were shown to be associated with a stronger transcriptional activity, enhanced replication fitness and higher viral loads when compared to the subtype C viruses with 3 NF-κB binding sites (Bachu et al., 2012). Specifically, Jeeninga et al., (2000) reported that the subtype E core promoter is more potent than the subtype B core promoter. Thus, demonstrating that inter-subtype genetic variation of the core promoter translates to differential transcription and replication activity. Taken together, these data demonstrate that intra-subtype 5'-LTR genetic variation may have implications for the transcription and replication capacity of the virus. The effect of genetic variation within the HIV-1 LTR is discussed in detail in the subsequent literature review (Chapter 2).

As aforementioned, previous studies investigated the functional variation of the HIV-1 LTR in chronic infection cohorts. However, in chronic infection most proviruses are defective (Bruner et al., 2016). On the other hand, the transmitter/founder (T/F) virus is the virus that is transmitted from a donor and

establishes infection in the new host (Haaland et al., 2009, Keele et al., 2008). Immediately following transmission, the pool of HIV-1 viruses circulating in the blood are homogenous (Keele et al., 2008). The homogenous character of the circulating viruses during acute infection is reported to be as a result of selective pressures that occur during transmission and establishment of the infection in the new host (Learn et al., 2002).

Interestingly, the data from our group and others (Carlson et al., 2014, Naidoo et al., 2017) show that the consensus like viruses for *gag*, *nef* and *env* sequences are selected for at transmission. Therefore, the data demonstrate that the viruses containing consensus like sequences for *gag*, *nef and env* are selected for at transmission and are transmitted to establish an infection in the next host (Carlson et al., 2014, Naidoo et al., 2017). However, it is not known whether this is true for other parts of the HIV-1 genome, more especially, the HIV-1 subtype C LTR that regulates viral gene transcription. Interestingly, the preliminary data from our group show intra-subtype HIV-1C T/F viruses LTR variation exists at transmission with viruses obtained from 2/25 (8%) of the patients exhibiting 4 NF-κB binding sites instead of the usual 3 NF-κB binding sites (Shamara Naicker *et al*., unpublished). In addition, preliminary LTR sequences obtained from 14 out of 25 patients exhibited the A5T mutation within the Sp1 III binding site. This LTR genetic variation associated with a lower viral load. However, the effect of this genetic variation on the transcription activity of the T/F viruses 5' LTR is not known. Understanding the impact of the 5' LTR genetic variation on its transcription activity and disease outcome may guide the development of HIV cure or eradication strategies.

1.1. Hypothesis

It is hypothesized that HIV-1 subtype C T/F viruses LTR genetic variation may have an impact on the viruses' transcriptional activity and may influence disease outcome in a cohort from South Africa.

1.2. Aim

The main aim of this project was to characterize the functional activity of the different HIV-1 subtype C T/F viruses LTR variants and to correlate this with markers of disease progression in an acute infection cohort of South Africans.

1.3. Objectives

1.3.1 Clone and sequence the patient derived HIV-1C transmitted/founder (T/F) viruses LTR sequences from FRESH and HPP cohorts - to study functional differences between LTR variants.

1.3.2 Determine the transcriptional activity of different HIV-1C T/F viruses LTR variants using a transcriptional assay.

1.3.3 Determine the cell activation potential of the HIV-1 subtype C T/F viruses LTR variants using a transcriptional assay and stimulation with tumor necrosis factor-α (TNF-α), phorbol myristate acetate (PMA), suberoylanilide hydroxamic acid (SAHA) and prostratin.

1.3.4 Correlate the HIV-1 LTR variants with markers of disease progression such as viral load and CD4 count.

CHAPTER 2: LITERATURE REVIEW

2.1 Background

The origin of human immunodeficiency syndrome (HIV) has been traced back to Africa from the early 1920's (Faria et al., 2014). The analyses of archived HIV positive blood samples provided evidence that the spread of HIV initiated through the transmission of the genetically similar simian immunodeficiency virus (SIV) (Faria et al., 2014). This transmission was from chimpanzees to the human population in Kinshasa, Congo (Faria et al., 2014). The widely accepted mechanism of transmission of SIV to the human population is believed to be as a result of Congolese communities that may have hunted, consumed or contacted SIV contaminated blood from the chimpanzees (Sharp and Hahn, 2011). The earliest confirmed case of HIV was traced to 1959 in Kinshasa from a blood sample obtained from an apparently healthy male (Science, 1998). The spread of HIV from Kinshasa to other geographical locations may be attributed to the expanding sex trade and migration dynamics within the Congo during that time (Faria et al., 2014, Wise, 2014).

Although HIV had originated from Africa, it was assumed that HIV had begun in the United States of America (USA) as the first cases of AIDS were only documented in October 1980 to May 1981 in the USA. The cases of rare cancer diseases in the USA was critical to the progressive discovery of HIV as discussed henceforth. In June 1981, the Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report (CDC's MMWR) reported the occurrence of a rare lung condition known as *Pneumocystis carinii pneumonia* in 5 previously healthy young homosexual males. These 5 males were hospitalized in Los Angeles from October 1980 – May 1981 (CDC, 1981b). However, within months of diagnosis 2 out of the 5 homosexual males passed away. Pneumocystis *carinii* pneumonia is a disease that was known to be associated with severe immunosuppression, which occurred mainly in elderly individuals (Walzer et al., 1974). However, the occurrence of this disease in young individuals without the predisposition to immunosuppression was a rare incidence (CDC, 1981b).

A month later, in July 1981, medical doctors in California and New York City reported 26 cases of a rare cancer disease known as Kaposi's Sarcoma in homosexual men (CDC, 1981a). Notably, in July 1981 an additional 10 cases of *Pneumocystis carini pneumonia* were identified in another group of homosexual males from Los Angeles (4 cases) and San Francisco (6 cases) (CDC, 1981a). Between May 1981 and July 1981, the number of rare cancer disease cases in homosexual males had reached a total of 41 reported cases (Altman, 1981). This disease was therefore associated with gay population. In the beginning of June 1982, the disease became known as gay-related immune deficiency (GRID)

because it had only been documented among the homosexual male population (CDC, 1982a). However, by the end of September 1982, the disease was also reported in 3 individuals with hemophilia, 36 Haitians and 77 intravenous drug users (CDC, 1982b, CDC, 1982d). At this time, it became clear that the disease could become established in other population groups (CDC, 1982b, CDC, 1982d). In addition, it became clear that the disease could be acquired by coming into direct contact with contaminated body fluids. In September 1982, the CDC had termed the disease acquired immunodeficiency syndrome (AIDS) (CDC, 1982c).

However, the viral agent responsible for the AIDS outbreak was unknown until May 1983 when scientists at the Pasteur Institute reported the discovery of the retrovirus, Lymphadenopathy-Associated Virus (LAV), as the agent responsible for AIDS (Barré-Sinoussi et al., 1983). Subsequently, in 1984, scientist Luc Montagnier and colleagues at the USA National Cancer Institute (NCI) identified this retrovirus but called it the Human T-Lymphotropic virus II (HTLV-II) (Marx, 1984). In addition, they had confirmed that HTLV-II may likely be the cause of AIDS (Marx, 1984). In March 1986 these viruses were renamed human immunodeficiency virus (HIV), the known causative agent of AIDS (Case, 1986).

2.2. Epidemiology

2.2.1. The Role of Circulating Non-Human Primate Simian Immunodeficiency Virus (SIV)

The global HIV-1 epidemic is due to the zoonotic transmission of SIV from chimpanzees. The diversfication of HIV-1 into four groups (Group M, N, O and P) is shown to be as a result of acrossspecies transmission event from different species of primates infected with their own species-specific SIV (keele et al., 2006, Neel et al., 2010). The Phylogenetic analysis has demonstrated that the zoonotic transmission of non-human primate SIV to humans has occurred during at least 8 transmission events (Hahn et al., 2000, Sharp et al., 2001). Currrently, serological evidence shows that over 40 species of non-human primate species are infected with a species specific virus. Approximately 90% of the non-human primate species have tested postive for a species specific SIV (Peeters et al., 2010). While the origin of HIV in the human population is known to be derived from the zoonotic transmision of SIV from chimpanzees, gorrilas and sooty mangabeys, other non-human primates (Colobus sp., mandrills, etc) may also be a source of transmitting their species specific SIV to the human population (Peeters et al., 2002). The human population in Sub-Saharan Africa is exposed to primates other than chimpanzees due to the hunting of bushmeat (Hahn et al., 2000).

The transmission of non-human primate SIVs to the human population may most likely be the result of human exposure to infected non-human primate blood (Hahn et al., 2000). Since the most likely trasnmisssion route is blood, hunters of primates are at increased risk for zoonotic transmission into the human population. The risk of zoonotic transmission to humans has increased with changes in the socioeconomic status of the Sub-Saharan population as well as with the increase in prevalence of non-human primate SIVs. The increased market for bushmeat as a source of food and income supports the possibility of zoonotic transmission events into the human population (Asibey., 1974, Geist, 1988) . In additon, in vitro assays have shown that non-human primate SIVs other than chimpanzees can replicate efficiently in primary human lymphocytes (Georges-Courbot et al., 1998). This further supports that possibility that zoonotic transmission to the human population may result in productive infection in humans.

Another factor to consider is whether any zoonotic transmission of non-human primate SIV from primates other than chimpanzees has been occuring (Peeters et al., 2010). A major concern is that a number of SIV strains may not be recognized by the standard HIV assays reslulting in a prolonged incbation period before the virus can be isolated or identified. This comes with the risk that newly circulating strains of SIV in humans may recombine with the circulating HIV viruses already established in the human population possibly leading to a new epidemic (Bailes et al., 2003).

2.2.2. HIV Epidemiology

Human immunodeficiency virus was discovered as the causative agent of Aids in 1983 (Barré-Sinoussi et al., 1983) and has since then become a global health concern. The Joint United Nations Programme on HIV/AIDS (UNAIDS) 2019 report estimated that there were approximately 37.9 million individuals living with HIV, 1.7 million new HIV infections, and 770 000 AIDS-related deaths globally in 2017 (UNAIDS, 2019).

While Eastern and Southern Africa account for only 12% of the world's population (Kharsany and Karim, 2016), they constitute the majority (approximately 54%) of global HIV-1 infections (UNAIDS, 2019). Moreover, South Africa has a rapidly expanding epidemic (Karim and Karim, 2010) and constitutes 30% of the new HIV infections in Southern Africa (**Figure 1**) (UNAIDS, 2019). Thus, further highlighting the severity of the HIV pandemic in this region.

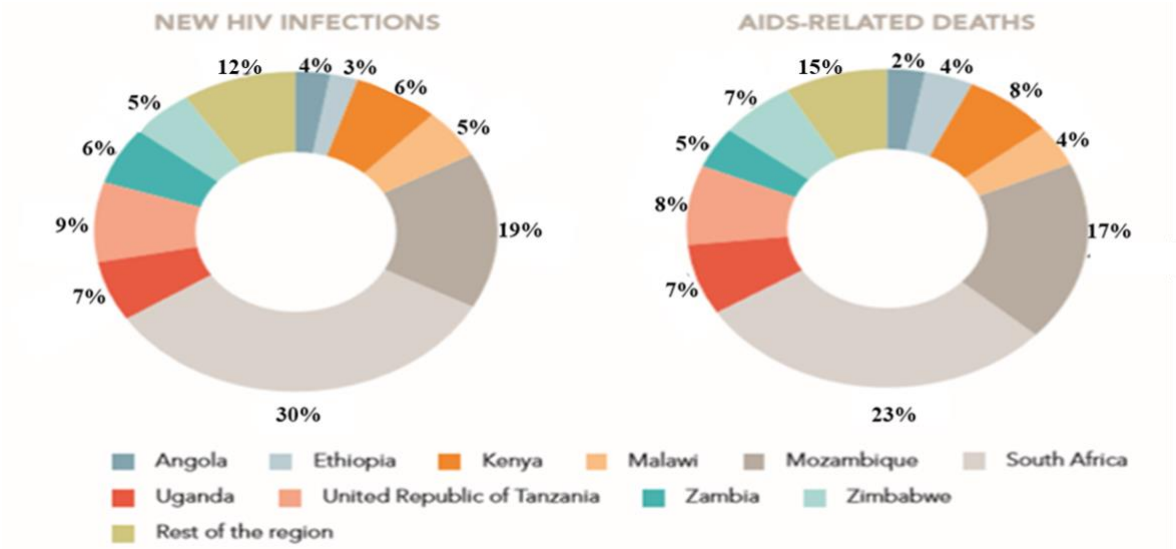


Figure 1. **The distribution of HIV infections in Eastern and Southern Africa**. The chart records the percentage contribution of each country within the Eastern and Southern regions. The percentage contribution of each country is marked by a specific colour that correlates with the respective country in the key presented below the chart. Specifically, the circles highlight that South Africa, marked by a silver colour, has the largest number of HIV infections (30%) in Eastern and Southern Africa. Reproduced with modifications from the UNAIDS 2018 report.

Thus far HIV has been characterised into two distinguished types, human immunodeficiency virus type 1 (HIV-1) and human immunodeficiency virus type 2 (HIV-2) (Clavel et al., 1986). The HIV-1 and HIV-2 exhibit similar genome organisation with specific genetic variation that sets them apart (Fletcher et al., 1996, Schulz et al., 1990, Zagury et al., 1988) and these differences will be highlighted as the review continues. In addition, they are distinguished by their geographical spread. The HIV-2 is largely restricted to West Africa (De Cock et al., 1993), while HIV-1 is globally distributed and accounts for 90% of the global HIV infections (Hemelaar, 2012). HIV-1 is diversified into four distinct groups namely, group M (Main), N (New), O (Outlier) and P (Ayouba et al., 2000, De Leys et al., 1990, Gürtler et al., 1994, Simon et al., 1998). Group M predominantly accounts for the majority of global HIV-1 infections and is further diversified into nine genetically distinct subtypes, subtypes A-D, F-H, J and K (Hemelaar et al., 2011). Recombination events between pure subtypes results in the formation of new strains of HIV-1 classified as circulating recombinant forms (CRFs) (Eliuk et al., 2011). A recombinant strain is defined as a CRF based upon its appearance in at least 3 infected subjects in the absence of an epidemiological link (Robertson et al., 2000). Thus far, at least 96 CRF's have been documented in the Los Alamos HIV database (www.hiv.lanl.gov).

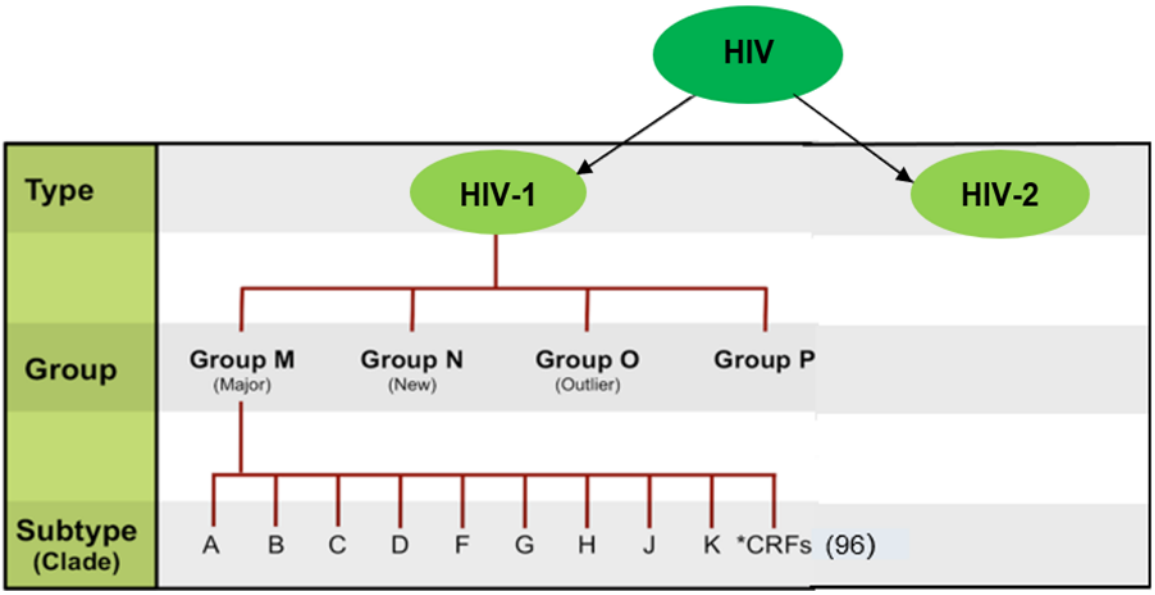


Figure 2**. HIV diversification**. The left side of the diagram characterizes HIV into two types, and further classifies HIV-1 into groups and subtypes. The HIV is divided into two different types, HIV-1 and HIV-2. The HIV-1 is divided into 4 groups, Group M (Major), Group N (New), Group O (Outlier) and Group P. Group M is further divided into 9 subtypes A-D, F-H, J and K and 96 circulating recombinant forms (CRFs) of HIV (www.hiv.lanl.gov). Reproduced with modifications from https://www.hiv.uw.edu/pdf/screening-diagnosis/epidemiology/core-concept/all

The HIV-1 subtypes are unevenly distributed globally (**Figure 3**). Subtype A is predominant within Central and East Africa (Buonaguro et al., 2007), while subtype B predominates North and South America, Australia and West and Central Europe (Hemelaar et al., 2011). Subtype D dominates North Africa and the Middle East whereas subtype F is established in South and Southeast Asia (Hardy, 2017). Subtype G is predominantly found in West and Central Africa (Hardy, 2017). The least common HIV-1 subtypes, subtype H, J and K persist stably at low levels in the population (Castley et al., 2017) and is geographically dominant in Africa and the Middle East (Hardy, 2017). Although the CRF's are distributed worldwide, three CRF's contribute significantly to the HIV-1 epidemic, CRF01_AE, CRF_02AG and CRF07_BC (Lau and Wong, 2013). The CRF02_AG is predominantly located in West and Central Africa while CRF01_AE is dominant within Southeast Asia (Bbosa et al., 2019, Hemelaar et al., 2011). In addition, CRF01_AE and CRF07_BC are predominate in china (Bbosa et al., 2019).

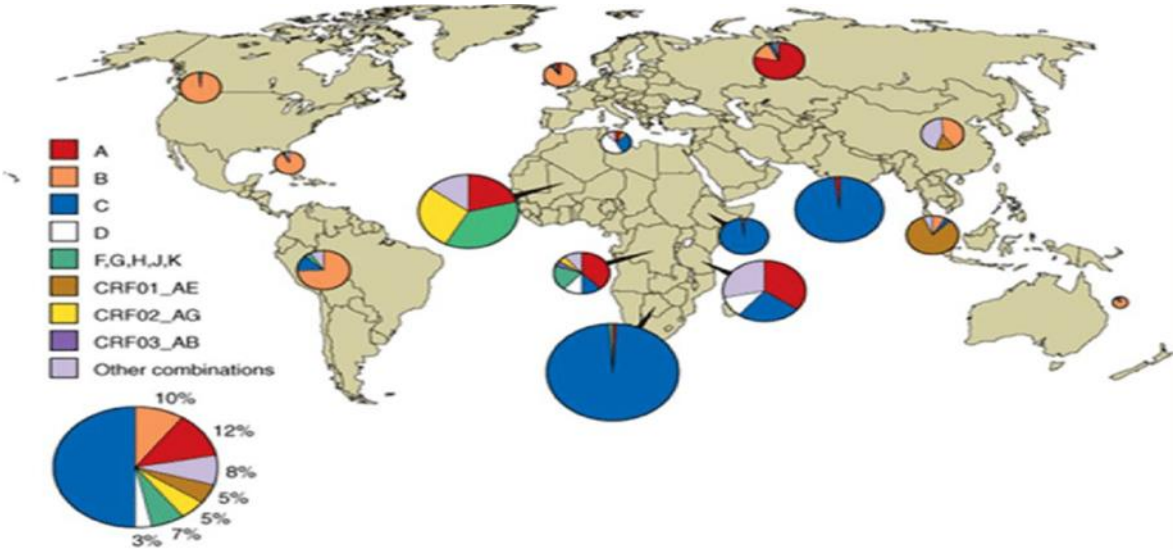


Figure 3. **The global distribution of HIV-1 subtypes**. Each subtype is represented by a different colour as demonstrated by the key in diagram. Subtype B represented by the orange colour dominates in North and South America. However, subtype C represented by the blue colour is seen to dominate Eastern, Southern Africa and India. Furthermore, the pie chart in the figure indicates that subtype C contributes to approximately 50% of the global HIV-1 pandemic while the other 50% of the HIV-1 pandemic is divided among the remaining HIV-1 subtypes (Reproduced from (Hemelaar et al., 2011).

On the other hand, HIV-1 subtype C is predominant in Sub-Saharan Africa and India (Hemelaar et al., 2011, Venner et al., 2016). The HIV-1 subtype C accounts for approximately 50% of the current worldwide infections (**Figure 3**) (Bbosa et al., 2019, Buonaguro et al., 2007, Castley et al., 2017, Hemelaar et al., 2011, Venner et al., 2016). Therefore HIV-1 subtype C is the focus of this study. As aforementioned, HIV-1 and HIV-2 are similar with a few genetic traits that set them apart. Therefore, the rest of this project proposal will refer to HIV-1 and only highlights differences between HIV-1 and HIV-2.

2.3 The Virology of HIV-1

Human immunodeficiency virus type 1 is classified into the genus lentivirus belonging to the family Retroviridae (Coffin, 1992). The HIV-1 genome is approximately 9.7 Kb long (Hahn et al., 1984) consisting of two genetically identical single stranded RNA genomes, each bordered by an incomplete long terminal repeat (LTR) on both the 5'- and 3' end of the genome (Coffin, 1992, Montagnier, 1999). Briefly, the 5' end LTR of the RNA genome lacks the U3 region while the 3' end LTR of the RNA

genome lacks the U5 region. The structure and the function of the full LTR element is discussed in detail in the subsequent sections since it is the focus of this study.

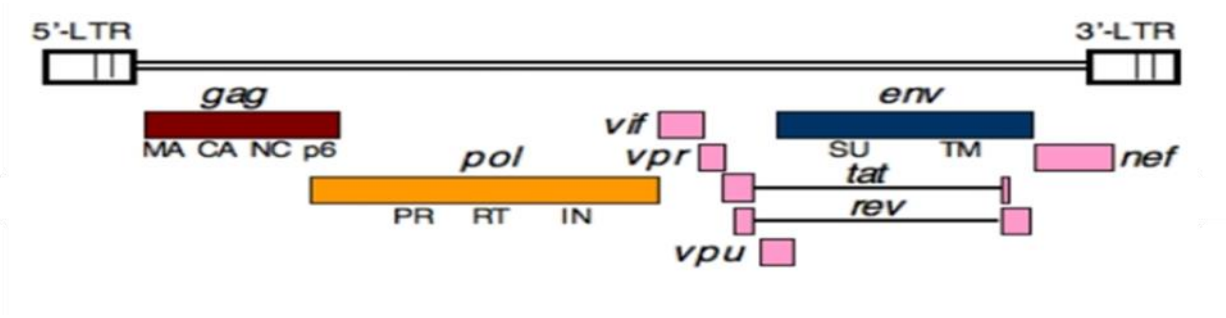


Figure 4. **The HIV-1 genome.** The HIV-1 genome is approximately 9.7 Kb long. The 5'and 3'-LTR flanks the entire HIV-1 genome. The rectangles indicate the reading frames of the HIV-1 genes. The structural genes include *gag* (brown), *pol* (yellow) and *env* (blue). The HIV-1 accessory and regulatory genes *vif, vpr, vpu* and *nef* are represented by pink rectangles. The two HIV-1 regulatory genes *tat* and *rev* also represented by pink colour. Reproduced from (Waheed and Freed, 2010).

The HIV-1 genome consists of 9 genes, which are categorized into 3 structural genes, 4 accessory genes and 2 regulatory genes (**Figure 4**) (Gallo et al., 1988, Stevens and Miller, 2016). The structural genes consist of: *gag, pol and env*, the accessory genes comprise the: *vif, vpr* (*vpx* in the case of HIV-2)*; vpu* and *nef* while the regulatory genes comprise of the *tat* and *rev* (Gallo et al., 1988, Stevens and Miller, 2016).

2.4 The HIV-1 Proteome

Although the HIV-1 genome has a small genome size with only 9 genes, it codes for a total of 16 viral proteins. The *gag* gene encodes for the p55 precursor protein that is cleaved to form four main structural proteins (Fields et al., 1996, Stevens and Miller, 2016). These include the symmetrical p17 matrix protein (MA) of the outer core membrane, the canonical p24 capsid protein (CA protein), the p7 nucleocapsid protein and a p6 protein (Fields et al., 1996, Henderson et al., 1992, Mervis et al., 1988). The *pol* gene encodes a precursor protein that is cleaved to produce 4 viral enzymes, which include Protease (PR), Reverse transcriptase (RT), RNAse H and Integrase (IN) (Ellegård, 2018, Fields et al., 1996, Stevens and Miller, 2016). The *env* gene encodes the gp160 precursor protein that is cleaved to form glycoprotein gp120 and glycoprotein gp41 (Fields et al., 1996, Stevens and Miller, 2016).

In addition to the structural proteins, HIV-1 encodes two important regulatory proteins which include the Transactivator of transcription (Tat) and the regulator of expression of virion (Rev) from the *tat* and *rev* genes, respectively (Gallo et al., 1988). Lastly, the accessory genes *nef, vpu, vpr* and *vif* codes for the negative regulating factor (Nef), viral infectivity factor (Vif), virus protein unique (Vpu) (or VpX in terms of HIV-2) and virus protein r (Vpr) (Gallo et al., 1988).

2.5 The HIV-1 Replication Cycle

The HIV-1 replication cycle will be briefly discussed with the aim of highlighting the role of the viral promoter, 5' LTR. The HIV-1 replication cycle is characterised into two phases, the early phase and the late phase (Tozser, 2003). The early phase of HIV-1 replication encompasses the attachment, entry, fusion and uncoating, reverse transcription, nuclear import and integration (Tozser, 2003). The late phase of HIV-1 replication is composed of the final steps which include viral gene transcription, assembly and budding (Tozser, 2003).

2.5.1 Attachment and Entry

The HIV-1 replication cycle begins with the binding of the mature virions to the target host cell. The mature virions successfully binds to the CD4 receptor of the host susceptible cell via its gp120 that project from the surface of the virions (Dalgleish et al., 1984, Klatzmann et al., 1984). The cellular attachment of gp120 to the host CD4 receptor induces a conformational change in the gp120 protein. The co-receptor binding sites of gp120 bind the co-receptors CXCR4 or CCR5, resulting in a more stable two-pronged attachment (Choe et al., 1996, Dragic et al., 1996, Friedman et al., 2007, Wu et al., 1996). This further induces a conformational change in the gp41 protein resulting in the exposure of the fusion peptide (Chan et al., 1997). The gp41 fusion peptide penetrates the host cell membrane thus tethering the viral envelop close to the cell membrane (Gallaher, 1987, Gallo et al., 2003, Liu et al., 2010). Consequentially, the gp41 fusion peptide creates a membrane fusion pore in the host cell membrane that facilitates the entry of the viral core into the host cell (Chan et al., 1997).

2.5.2 Uncoating

Although uncoating is an important step in HIV-1 replication cycle since it makes viral RNA genome available for viral DNA synthesis, it is poorly understood. One theory proposes that viral core dis-assembles immediately after cell entry making viral RNA available for reverse transcription to take place in the cytoplasm within the reverse transcription complex (RTC) (Fassati and Goff, 2001, Karageorgos et al., 1993, Miller et al., 1997). While the other theory proposes that uncoating occurs during viral core trafficking through the cytoplasm or once it has reached the nuclear pore complex (NPC), suggesting that reverse transcription occurs within the viral core (Arhel et al., 2007, Hulme et al., 2011, Lukic et al., 2014, Pawlica and Berthoux, 2014, Xu et al., 2013, Zhou et al., 2011). The subsequent steps that are linked to the first theory of transcription will be discussed in this dissertation.

2.5.3 Reverse Transcription

As aforementioned, following uncoating the reverse transcription of the viral RNA takes place immediately within the reverse transcriptase complex (RTC). The reverse transcriptase complex (RTC) contains the enzyme reverse transcriptase and the two single stranded HIV-1 RNA genome, IN, Vpr, MA and CA and cellular components such as dNTPs (Fassati and Goff, 2001, Iordanskiy et al., 2006, Nermut and Fassati, 2003). During the microtubule-based transport to the nucleus, the viral single-stranded RNA genome is reverse transcribed into a double helix RNA/DNA hybrid complementary (c)DNA within the RTC (Bukrinskaya et al., 1998, McDonald et al., 2002). Simultaneously, the viral RNase H degrades the viral RNA while the host DNA polymerase synthesizes a complementary DNA strand to complete the synthesis of double stranded HIV-1 DNA. Following the synthesis of the complete HIV-1 double stranded DNA molecule, the RTC matures into the pre-integration complex (PIC) (Iordanskiy et al., 2006). The PIC comprises the HIV-1 DNA, IN associated viral proteins such as IN, NC and Vpr (Bukrinsky et al., 1992, KARAGEORGOS et al., 1993, Miller et al., 1997) and cellular proteins such as lens epithelium derived growth factor p75 (LEDGF/p75) (Cherepanov et al., 2003) and Transportin SR2 (TRN-SR2) (Christ et al., 2008).

2.5.3.1. Synthesis of the Full Length LTR

The HIV-1 LTR is located at both the 5'and 3' end of the HIV-1 genome. The LTR is divided into three regions namely the Unique 3 (U3), Repeat (R) and Unique 5 (U5) regions. The U3 region consists of the enhancer and promoter region while the R region consists of the initiation and termination sites for reverse transcription. As aforementioned, the 5' LTR of HIV-1 RNA genome lacks the U3 region while

the 3’ LTR lacks the U5 regions (Delviks-Frankenberry et al., 2011). The synthesis of the full length LTR occurs via a strand transfer mechanism in sequential steps (Coffin et al., 1997) (Figure 5).

The plus strand single stranded RNA genome serves as a template for the synthesis of a complementary minus strand of DNA. Reverse transcription initiates near the 3’end of the 5’LTR when the tRNA base pairs to the complementary Primer Binding Site (PBS) (Step 1). The enzyme RT then synthesizes the minus strand DNA by transcribing the U5 and R regions (Step 2). The newly synthesized minus strand DNA can then initiate the first jump and transfer to the 3’end of the plus strand RNA (Step 3). During the first strand transfer, the R region of the minus strand DNA base-pairs to the R-region of the plus strand RNA. Minus strand DNA synthesis then continues in the 3’to 5’direction when RT transcribes the U3 region adjacent to the R region and, continues transcribing the minus strand DNA until the 5’U5 and R region are transcribed for the second time (Step 4).

Upon synthesis of the minus DNA strand, RNase H subsequently degrades the plus strand template RNA except for the fragment of RNA that is annealed to the polypurine tract (PPT) at the 5’end of the 3’-U3 region (Step 5). This RNA fragment acts as a primer for RT of the second strand of DNA which is the plus strand of DNA. Using the minus strand of DNA as a template, RT extends the plus strand DNA by transcribing the 3’Region and U5 region (Step 6). Following this, the second jump of DNA occurs in which the newly synthesized plus strand of DNA transfers to the 5’end of the minus strand DNA by base pairing with the R and U5 region (Step 7). DNA synthesis then proceeds in both directions to synthesize two complete and complementary double strands of DNA with the full length 5’and 3’LTR (Step 8).

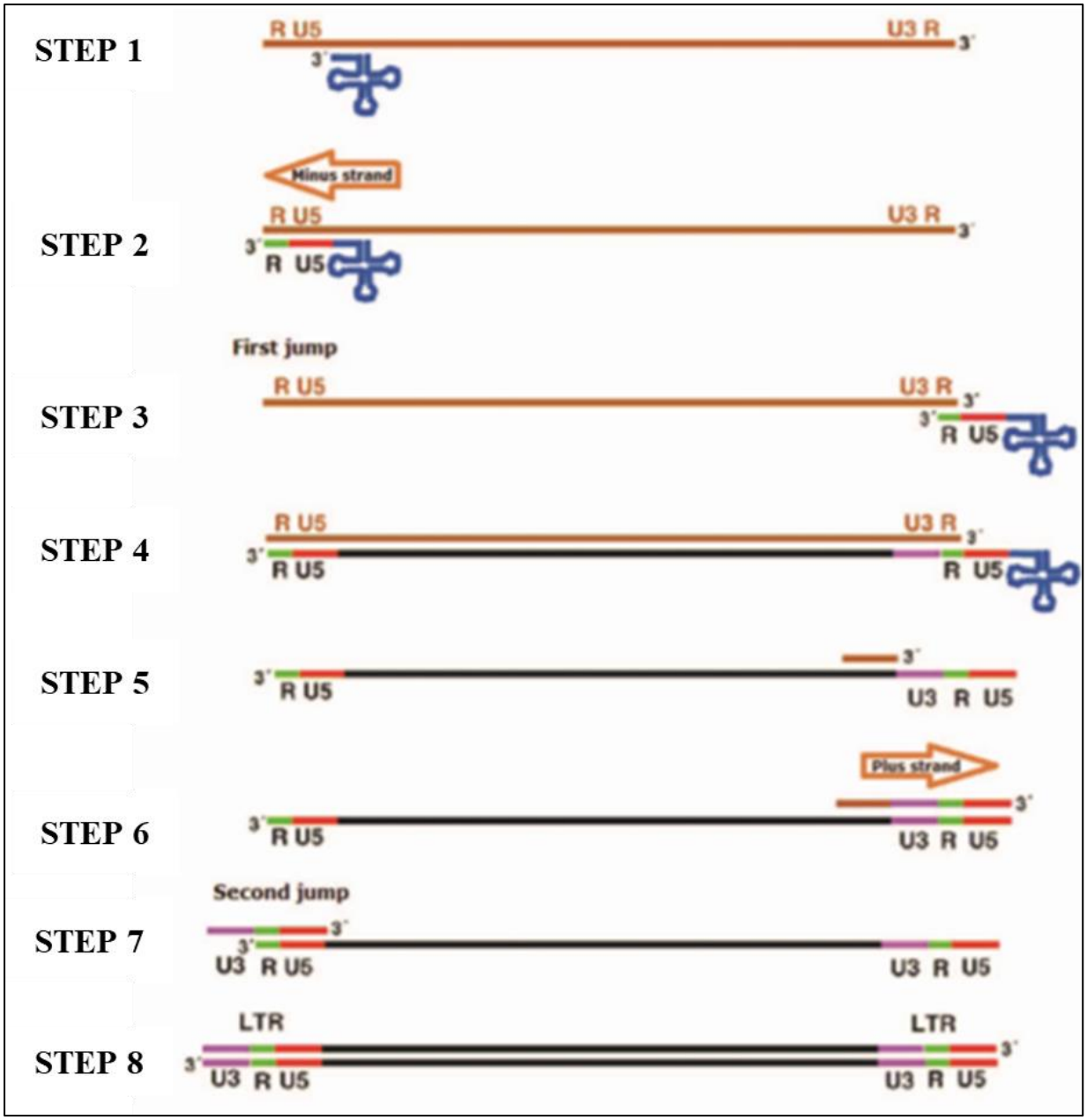


Figure 5. **Synthesis of the full length LTR**. Step 1: tRNA binds to the PBS on the 5'-U3 region. Step 2: RT transcribes the 5'U3 and R region to produce a minus strand DNA. Step 3: Minus strand DNA makes its first jump and binds to the 3'R region. Step 4: RT then extends the minus strand DNA in the 3'→ 5'direection. Step 5: RNase H degrades the plus strand RNA template while leaving a short fragment of the plus strand RNA that remains bound to the PPT of the 5' end of the 3'LTR. Step 6: RT then uses the minus strand DNA as a template to extend the fragment of RNA and synthesize the plus strand DNA in the 5' → 3'direction. Step 7: Plus strand DNA then makes the second jump and binds to the 5'U3R region. Step 8: RT then extends the both the minus and plus strand of DNA to synthesize the full length 5'and 3'LTR. Reproduced with modifications from (Zhang et al., 2014).

2.5.4 Nuclear Import and Integration

The TRN-SR2 protein mediates the transport of the PIC from the cytoplasm into the nucleus (Christ et al., 2008). The TRN-SR2 interacts directly with the viral IN through the IN catalytic core domain (CCD) and the IN C-terminal domain (CTD) (De Houwer et al., 2012). The IN CCD dimerizes with the IN CTD and the CCD-CTD dimers binds to the TRN-SR2 (De Houwer et al., 2012, Tsirkone et al., 2017). Integrase further interacts with the TRN-SR2 by binding to the HEAT proteins of TRN-SR2 (Tsirkone et al., 2017). This ultimately facilitates the migration of the PIC through the nuclear pore into the nucleus. Upon arrival of the PIC complex at the nucleus, TRN-SR2 binds to the Ras-related nuclear protein (Ran) to facilitate the release of the PIC into the nucleus (Taltynov et al., 2013).

Once inside the nucleus, LEDGF/p75 tethers the PIC to the host (Cherepanov et al., 2003). The viral IN catalyses the integration of the viral double stranded DNA into the host genome through two processes (Engelman et al., 1991, Pauza, 1990, Vink et al., 1991). The integration process begins in the cytoplasm, where the IN catalytically removes the di-nucleotides (GT) from the 3' end of both ends of the LTR to produce active 3'-OH groups at the 3'-ends of the LTR (Engelman et al., 1991, Fujiwara and Mizuuchi, 1988, Pauza, 1990, Roth et al., 1989, Vink et al., 1991). Once inside the nucleus, the 3'-OH groups of the HIV-1 DNA initiate a nucleophilic attack on the host DNA to facilitate the stable insertion of the viral DNA into the host chromatin (Brown et al., 1989, Engelman et al., 1991, Fujiwara and Mizuuchi, 1988, Pauza, 1990, Vink et al., 1991). The gaps are then repaired by the host enzyme, ligase to form a stable integrated viral DNA, referred to as provirus.(Guth and Sodroski, 2014). This step occurs in the nucleus of the infected host cell. Once the HIV-1 genome is integrated into the host chromosome then it is referred to as provirus (proviral DNA).

2.5.5 Viral Gene Expression

The proviral DNA serves as a template for the transcription of HIV-1 genes (Guth and Sodroski, 2014). Viral gene transcription is driven by the LTR, which will now be discussed in detail since it is the focus of this study.

2.5.5.1 The Structure and Function of the HIV-1 LTR

The HIV-1 DNA genome is flanked at both the 5' and 3' end by the LTR, which is approximately 640 bp in length (reviewed in (Groen and Morris, 2013, Mbonye and Karn, 2014). The LTR of the DNA genome is divided into 3 distinguished regions, the U3, R and U5 regions. The U3 region is further divided into 3 domains that promote positive sense transcription, namely the core promoter, core enhancer and modulatory domain (Gaynor, 1992, Pereira et al., 2000) (**Figure 6**). The core promoter consists of various transcription factor binding sites including three tandem Sp1 binding sites, TATA box and an enhancer box (E-box) (Jones et al., 1986, Jones and Peterlin, 1994). Interestingly, the TATA box in subtype E viruses is mutated to contain a TA**A**AA sequence instead of the TA**T**AA sequence described for subtype B and C (Bachu et al., 2012, Montano et al., 1998) In addition, the initiator site at the transcription start site is located within the core promoter (Groen and Morris, 2013, Mbonye and Karn, 2014, Qu et al., 2016).

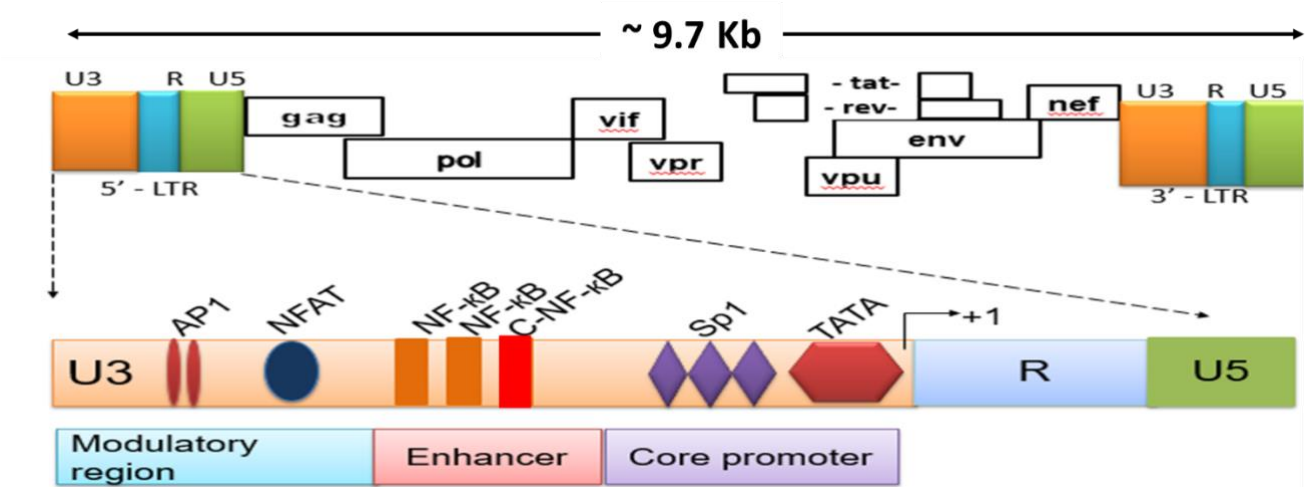


Figure 6. **Structure of the full length 5'-LTR**. The HIV-1 genome is ~~9.7 Kb long and consists of 9 genes (*gag, pol, env, vif, vpr, vpu, env, tat, rev and nef*) that are flanked on both sides by the LTR. The 3'and 5'LTR are identical at the structural and nucleotide level. The 5'LTR is divided into the U3, R and U5 region. The U3 region is divided into the modulatory domain, core enhancer domain and core promoter domain. The modulatory domain includes the TFBS such as AP1 (shown by the maroon oval), NFAT (shown by a blue circle) while the core enhancer consists of the NF-κB binding sites that vary in copy number depending on the subtype. Almost all subtypes consist of two conserved canonical NF-κB binding sites (shown by the orange rectangles) however subtype E which consists of only one. Subtype C on the other hand consists of an additional 3rd NF-κB binding site (designated C-NF-κB and shown by a red rectangle). The core promoter domain consists of three tandem sp1 binding sites (shown by the purple diamonds) and a TATA box (shown by the maroon hexagon). The imitator site located in the core promoter is indicated by the arrow and +1

The core enhancer has been understood to typically contain 2 highly conserved NF-κB binding sites. However, NF-κB binding sites vary in number depending on the subtype being examined (Baar et al., 2000, Bachu et al., 2012, Verhoef et al., 1999). A majority of subtypes consists of 2 canonical and conserved NF-κB binding sites whereas subtype E consists of only one NF-κ binding site (Baar et al., 2000, Jeeninga et al., 2000). Notably, subtype C exclusively consists of an additional third NF- κB binding site and (Bachu et al., 2012, Montano et al., 1997), subtype C viruses with an insertion of a 4th NF-κB binding site has also been described (Bachu et al., 2012) (**Figure 6**). The modulatory region consists of transcription factor binding sites located upstream of the enhancer element. These transcription factor binding consist of the CCAAT/enhancer binding protein (C/EBP), lymphocyte enhancer factor (LEF-1), nuclear factor of activated T cells (NFAT) and negative regulatory element (NRE) (Henderson et al., 1995, Shaw et al., 1988, Tesmer et al., 1993).

2.5.5.2 The Function of Host Cellular Factors in Basal Transcription

Following integration of the HIV-1 DNA, viral gene expression is mediated via the action of cellular and viral proteins that bind to the transcription factor binding sites (TFBS) on the core promoter, enhancer domains and modulatory region (Fields et al., 2007, Hunt et al., 2001). The HIV-1 gene expression consists of two phases, the early phase and late phase of transcription. During the early phase of transcription, the host cellular factors inclusive of NF-κB and Sp1 and Tata Binding Protein (TBP), bind to the core modulatory region, core enhancer and promoter region to initiate gene expression (Kao et al., 1987, Yedavalli et al., 2003). The core promoter specifically initiates viral gene transcription and is sufficient for low levels of gene expression, However, the interaction between the cellular factors and the core modulatory and enhancer region is required to further stimulate the productive levels of basal gene transcription.

In the modulatory region, the cellular transcription factor, nuclear factor of activated T cells (NFAT), plays a significant role in regulating HIV-1 transcription. In the inactive form, NFAT transcription factors remain sequestered and phosphorylated within the cell cytoplasm. However, upon T-cell activation, the calcium-calmodulin-dependent phosphatase calcineurin dephosphorylates and activates NFAT (Pessler and Cron, 2004). This subsequently leads to the translocation of the NFAT factors from the cytoplasm to the nucleus as well as the activation of genes containing the NFAT binding sites (Pessler and Cron, 2004). Although NFAT can bind to the NFAT binding site in the modulatory region, this is a low affinity interaction. Interestingly, NFAT factors bind with high affinity to the NF-kB binding sites in the enhancer region and is one of the significant binding sites for NFAT mediated HIV-1 gene transcription (Cron et al., 2000).

In the activated form, NF-kB is a dimer composed of p65/p50 subunits (Baeuerle and Baltimore, 1989, Ballard et al., 1992, Ghosh et al., 1990, Kieran et al., 1990, Nolan et al., 1991, Ruben et al., 1992). However, when a cell is inactivated (in a resting state) the p65 NF-kB subunit is repressed by IkB while the p50 subunit circulates freely in the cell cytoplasm (Baeuerle and Baltimore, 1988, Ganchi et al., 1992, Inoue et al., 1992, Nolan et al., 1991, Sen and Baltimore, 1986, Zabel and Baeuerle, 1990). Activated HIV-1 infected host cells relieve the restriction on p65 thus allowing p65 to dimerize with p50 and form NF-kB complexes (Nabel and Baltimore, 1987). The NF-kB is translocated to cell nucleus from the cytoplasm and, binds to the NF-kB TFBS on the core-enhancer domain of the 5'-LTR (Lenardo and Baltimore, 1989, Sen and Baltimore, 1986).

One of the mechanisms via which the NF-kb and NFAT transcription factors modulates HIV-1 LTR driven transcription is through host chromatin modification. Host activating factors such as NF-kB and NFAT function to recruit histone modifiers such as histone acetyltransferases that modify lysine residues on histone 3 and histone 4 (Cron et al., 2000, Okamura et al., 2000, Zhong et al., 2002). This further modifies chromatin to an open configuration that exposes the transcription binding sites to additional host and viral transcription factors (reviewed in (Mbonye and Karn, 2014, Schiralli Lester and Henderson, 2012).

The core promoter specifically initiates viral transcription. As aforementioned, the core promoter consists of the TATAA element and three tandem Sp1 binding sites, Sp1 I, Sp1 II and Sp1 III. Together, the TATAA element and the Sp1 binding sites function to initiate viral gene transcription (Roebuck and Saifuddin, 1999). The TATAA element binds the protein complex known as the transcription factor (TF) IID that is composed of the TATA binding protein (TBP) and multiple TBP associated factors (TAF) (Majello et al., 1998). The TBP plays a critical role in initiating the formation of the preinitiation complex for viral gene transcription (Pendergrast et al., 1996). The binding of cellular Sp1 transcription factors to the three tandem Sp1 sites on the HIV-1 core promoter activates the Sp1 sites. The activated Sp1 sites enhances the assembly of the preinitiation complex (Harrich et al., 1989). The preinitiation complex is stabilized by transcription factor (TF) IIA and TFIIB and is further completed by the recruitment of RNA Polymerase II-TFIIF, TFIIE and TFIIH (Thomas and Chiang, 2006). This promotes the initiation of transcription by the viral promoter, the 5'-LTR

Following the initiation of transcription, the RNA polymerase II becomes regulated by the negative elongation factor (NELF) that inhibits the synthesis of complete (long) gene transcripts (Ping and Rana, 2001). Therefore, although the LTR activity is efficiently initiated, the LTR is inhibited from producing

full length functional gene transcripts by NELF (Kao et al., 1987). The production of short abortive gene transcripts is mediated by the Inducer of Short Transcripts (IST) region located downstream from the transcription start site (Pessler and Hernandez, 1998, Sheldon et al., 1993). The newly synthesized abortive mRNA transcripts undergo splicing to remove the intronic regions of the early mRNA (Daelemans et al., 1999, Purcell and Martin, 1993). These multiple spliced mRNA transcripts encode the two regulatory genes, *tat* and *rev*. The multiple spliced mRNA transcripts are then exported to the cytoplasm for translation into the Tat and Rev proteins and, are then shuttled back to the nucleus.

2.5.5.3 Tat Induced Viral Gene Transcription

Although transcription has been initiated at this point, RNA Pol II is unable to elongate efficiently resulting in a paused transcription that produces short gene transcripts. However, Tat is synthesized in the cytoplasm and localizes to the nucleus where it functions as a strong activator of transcription (Raha et al., 2005). The Tat binds to the TAR RNA located in the R-region of 5'-LTR (Raha et al., 2005). The interaction between Tat and TAR enables Tat to recruit the co-factor positive transcription elongation factor b (P-TEFb) (Wei et al., 1998, Zhu et al., 1997). The P-TEFb is composed of two subunits namely, cyclin T1 and CDK9 (Wei et al., 1998, Zhu et al., 1997). The CDK9 subunit of P-TEFb phosphorylates the c-terminal domain of RNA Pol II and activates RNA Pol II to synthesize functional gene transcripts (Kim et al., 2002). The synthesis of Tat activates the Tat-dependent mechanism of LTR and increases the efficiency of gene expression resulting the formation of primary full length viral gene transcripts called pre-mRNAs (Karn, 2000).

The primary pre-mRNA transcript can then either undergo single splicing or remain unspliced (Saltarelli et al., 1996). The single spliced mRNAs are translated into the Env, Vpu, Vpr and Vif viral proteins. While the unspliced viral transcripts are destined to be translated into the Gag or Gag-Pol polyprotein, or alternatively are packaged into the newly synthesized virions as the viral genome. The next step in viral gene expression is the export of the single spliced and unspliced viral gene transcripts that is mediated by the Rev regulatory protein that was synthesized during early phase of transcription. The single spliced and unspliced mRNAs are intron containing transcripts that consist of a Rev Response Element (REV) within the intronic region (Cullen, 2000, Cullen, 2003, Malim et al., 1989). Rev binds to the RRE and multimerizes resulting in the recruitment of nuclear factors such as CRM1 and Ran-GTP. These nuclear factors then associate with the Rev-RRE complex to from a Rev-RRE-CRM1-RanGTP RNA complex that aids the transport of the mRNAs to the cytoplasm through the nuclear pore complexes (Suhasini and Reddy, 2009). The mRNAs are released into the cytoplasm through the conversion of RanGTP to RanGDP that results in dissociation of the complex.

2.5.6 Assembly and Budding

In the cytoplasm, the translation of the single spliced or unspliced transcripts results in the production of Gag, Gag-Pol, Vpr and Vif viral proteins. Upon the completion of viral proteins translation, viral proteins and the unspliced viral RNA that functions as the viral genome, then assemble into immature virions at the cell membrane (Guth and Sodroski, 2014). During assembly, host proteins such as tumor suppression gene 101 assist the assembly process to produce the immature virions. Viral budding occurs at the cholesterol rich plasma membrane. The process of budding is aided by the viral protein Vpu and the immature virions then bud through the host cell thus acquiring a viral envelope and then undergo maturation process to form matured virions (Fanales-Belasio et al., 2010).

2.6. The Impact of Genetic Variation in the 5'-LTR

Genetic variation within the LTR may largely be due to the error prone reverse transcriptase that introduces nucleotide changes during reverse transcription (Shah et al., 2014). Genetic variation within the transcription binding sites alters the interaction of the transcription factors with their binding sites in the LTR (Jeeninga et al., 2000). This may consequentially alter the LTR mediated transcription.

Previously, studies have investigated the effect of inter- and intra-subtype variation on disease outcome. Subtype B consists of 2 identical and conserved NF-κB binding sites (Bachu et al., 2012, Qu et al., 2016). Although most viral subtypes share this characteristic, two exceptions include subtype AE and subtype C (Bachu et al., 2012). While subtype AE has only one NF-κB binding site (Jeeninga et al., 2000, Montano et al., 1998), subtype C contains at least three NF-κB binding sites (Bachu et al., 2012). A majority of subtype C viruses exhibit three NF-κB binding sites while a small proportion of subtype C viruses contain a fourth NF-κB binding site (Bachu et al., 2012, Hunt and Tiemessen, 2000). Moreover, the occurrence of subtype C viruses with only two NF-κB binding sites have also been described (Bachu et al., 2012, Hunt et al., 2001, Scriba et al., 2001). The subtype C viruses consisting of three NF-κB binding sites have been correlated with higher transcriptional activity as compared to subtypes with one or two NF-κB binding sites (Bachu et al., 2012, Montano et al., 1998, Qu et al., 2016).

Montano *et al*., (1998) had shown that the HIV-1 subtype AE LTR, which contains one NF-κB binding site was less inducible upon TNF-α stimulation compared to the prototype subtype B. Therefore, the inter-subtype variability in the number of NF-κB binding sites may impact the replication rate of the

virus (Montano et al., 1998). In another study, Jeeninga et al., (2000) showed a correlation between the number of NF-κB binding sites and the degree of LTR activity. Consistent with Montano et al., the results reported by Jeeninga et al., indicated that the subtype AE which, contains one NF-κB binding site was the least activated with TNF-α as compared to subtype B, which contains two NF-κB binding sites. In addition, subtype C exhibited a greater degree of activation as compared to subtype B and AE (Jeeninga et al., 2000). Therefore, the presence of additional NF-κB binding site may confer an increased transcriptional and replicative ability to the virus.

As aforementioned, subsequent studies have shown intra-subtype variability. The occurrence of insertions in the Mozambican and Brazilian HIV-1 subtype C LTR generated an additional NF-κB binding site (Boullosa et al., 2014). Interestingly, insertions found within the Brazilian HIV-1 subtype C LTR was linked to the presence of an additional RBE III binding site (Boullosa et al., 2014). These findings were demonstrated in isolates obtained from patients that failed to respond to therapeutic intervention (Boullosa et al., 2014).

In addition to the variation observed within the NF-κB binding sites, studies have also investigated variation within the Sp1 binding sites. Qu *et al*. has demonstrated that the Chinese subtype B 5'-LTR exhibits variation in the Sp1 binding site and NF-κB binding site is linked to a variation in the viruses' ability to drive viral gene transcription. Furthermore, in another study aimed at establishing a link between LTR sequence variation and disease progression, it was found that a mutation in the Sp1 site III that produces a C to T alteration at position 5 results in an increase in disease progression (Nonnemacher et al., 2004).

2.7 The Bottleneck Theory

A majority of global HIV-1 infections are as a result of heterosexual transmission (Joseph et al., 2015). Approximately 80% of new heterosexual HIV-1 infections are initiated by a single infecting viral variant termed the transmitted/founder (T/F) virus (Joseph et al., 2015). The heterosexual transmission of HIV-1 is distinguished by a stringent genetic bottleneck (Carlson et al., 2014).

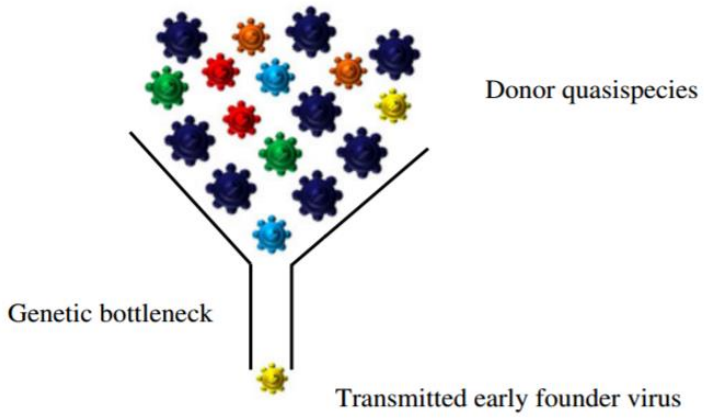


Figure 7**. A bottleneck of viral quasispecies exists at transmission.** A chronically infected HIV-1 individual contains a quasispecies of HIV-1 variants. The quasispecies is composed of major viral variants represented by the large dark blue virions and minor viral variants represented by the smaller green, red, yellow and orange virions. During transmission, there is a narrowing of the genetic bottleneck at the point of transmission resulting in the transmission of a minor viral variant (yellow). The minor variant (yellow) becomes known as the transmitted/founder virus that establishes infection in the newly infected individual.

The bottleneck theory states that a chronically infected and transmitting partner has a diverse viral quasispecies consisting of a major viral variant and multiple minor variants (Braibant and Barin, 2013). However, at the point of transmission, there is a sharp narrowing of the viral quasispecies towards selection for a single viral variant that will be transmitted to the recipient host **(Figure 7)** (Kariuki et al., 2017). Following transmission, the T/F virus remains homogenous in the newly infected recipient mainly during Fiebig stage I and Fiebig stage II (the Fiebig classification system is used to identify the stages of HIV-1 infection. The Fiebig stage I characterize the period during which HIV-1 RNA becomes detectable while Fiebig stage II characterizes the detection of the HIV-1 p24 antigen. This will be discussed in detail in the subsequent paragraphs.) while continuing to replicate at an exponential rate. The re-emergence of the viral quasispecies occurs during replication and is chiefly found in the chronic phase of infection (Keele et al., 2008).

It remains unclear whether HIV-1 transmission is chiefly a stochastic event or a selective event. However, in the experimental design of our study, the HIV-1 LTR element sequences were analysed using the plasma samples from the recipients of the HIV-1 infection. Therefore, we did not examine the HIV-1 LTR element sequences from other bodily fluids such as genital fluids neither did we examine the HIV-1 LTR sequences in both the transmission donor and recipient. Thus, this study was unable to provide a distinction of a stochastic event from a rare variant or dominant variant. Therefore, the discussion presented here is to briefly understand the events surrounding HIV-1 transmission.

Previously, it was suggested that HIV-1 transmission may be a stochastic event in which each viral variant present in the quasispecies has an equal probability of transmission (Kariuki et al., 2017). Although some stages of transmission may be as a result of stochasticity, recently it has been suggested that a selection bias may exist at transmission (Song et al., 2016). Studies in favour of selection bias undertake that specific genotypic and phenotypic characteristics may confer increased transmissibility (Joseph et al., 2015).

Interestingly, a previous study showed that T/F virus is not the most dominant variant present in the plasma or genital tract of the transmitting donor, suggesting that transmission could not be completely stochastic (Deymier et al., 2015). Other studies demonstrated that the T/F viruses exhibit consensus like sequence particularly when examining the viral *gag*, *pol, nef env* genes (*env*) (Carlson et al., 2014, Deymier et al., 2015). While selection bias of the *env* has been well described, selection bias has also been demonstrated outside of the *env* (Deymier et al., 2015). Therefore, the T/F virus that exhibits a greater degree of genetic similarity to the consensus sequence is preferentially selected for at transmission.

2.8 The Course of HIV Infection

A typical HIV-1 infection is characterized by three different phases including the acute phase of infection, the chronic phase of infection and the final end-stage AIDS phase (Cohen et al., 2016). The three phases of infection remain well identifiable (Cohen et al., 2016). A detailed discussion on the latter 2 phases of infection are beyond the scope of this study, therefore a brief definition is given. Briefly, the chronic phase of infection is defined by the detection of HIV specific antibodies, which usually develops after 28 days of infection. This phase of infection is highly variable between patients and may persist for 7 to 10 years (Hernandez-Vargas and Middleton, 2013). The chronic stage is followed by the end-stage AIDS phase. The end-stage AIDS phase is defined by a $CD4^+$ T-cell count that is below 200 cells/mm^3 and can be complicated by opportunistic infections thus increasing the risk of mortality (Hernandez-Vargas and Middleton, 2013, Rutstein et al., 2017)

2.9. Acute Infection

The course of a HIV-1 infection starts with an acute phase of infection, defined as the detection of HIV-1 RNA and p24 antigens in patient blood prior to the detection of HIV-1 specific antibodies in the blood

(Cohen et al., 2016). The acute phase of infection is categorized into 6 different stages known as Fiebig stages (Fiebig et al., 2003). The Fiebig stage I is characterized by the detection of only HIV-1 RNA in the patients' blood with Fiebig stage II occurring approximately 7 days thereafter and is defined by the detection of HIV-1 RNA and the p24 antigen in the patients' blood. The Fiebig stage III corresponds to a period of 5 days after a positive p24 antigen result and is defined by the detection of HIV-1 immunoglobulin M (IgM) antibodies by enzyme immunoassays. Fiebig stage IV occurs approximately 3 days after positive enzyme immunoassay and defined by indeterminate Western blot test and Fiebig stage V, which occurs 7 days after Fiebig stage IV or approximately 1 month after infection and is defined by clearly positive Western blot test.

The HIV-1 acute infection cohort provides several key characteristics that can be exploited by studies aimed at therapeutic interventions. Investigations based on acutely infected individuals could generate data that can develop our understanding of HIV-1 transmission pathways further. And, consequentially, therapeutic intervention at the acute infection phase could possibly reduce viral load and the size of latent reservoir which is a barrier to HIV-1 cure development (Cohen et al., 2010). However, it is well established that HIV-1 latency develops early in acute infection and latent reservoirs persists for the lifetime of the infected individual. While emphasis has been placed on the acute phase of infection for therapeutic interventions, the acute phase of infection has not been fully explored for effective HIV-1 cure strategies. Although the molecular mechanisms of HIV-1 latency development have not yet been fully elucidated, there are a few mechanisms that have been put forward to explain the maintenance of HIV-1 latency. Here, we discuss four mechanisms of viral latency that include chromatin configuration, nucleosome positioning, DNA methylation and transcriptional interference.

2.9.1 Chromatin Configuration and Epigenetic Modifications

Gene expression of the host and integrated provirus is controlled by the chromatin configuration in a dynamic process that alters the configuration of the chromatin to either permit or inhibit host gene expression. The nucleosomes are both functional and structural units of chromatins and consists of four pairs of core histones that includes H2A, H2B, H3 and H4 (Luger et al., 1997). Importantly, when these histones are configured around the integrated proviral core promoter that undergo epigenetic modifications that either promote transcription or promote viral silencing.

In addition, in the compact configuration, the tightly bound nucleosomes contain the viral genome in a heterochromatin state (Jordan et al., 2003). The heterochromatin configuration limits the accessibility

of the viral promoter to the host cellular transcription factors thus repressing viral gene transcription (Tripathy et al., 2011; Tamaru 2010). In contrast, in the relaxed configuration, the viral genome is in a euchromatin states (Tripathy et al., 2011; Tamaru 2010). This allows for the cellular host transcription factors to access the viral promoter and initiate viral gene transcription. These chromatin configuration changes are a result of histone epigenetic modifications.

2.9.2. Histone modifications: Acetylation and Deacetylation

The amino terminal of each histone protrudes to the exterior of the nucleosome thus making itself susceptible to epigenetic modifications. Histone modifications are inclusive of sumoylation, ADP ribosylation ubiquitination, phosphorylation, methylation and acetylation and are reversible modifications (Cain et al., 2011, Lim et al., 2009, Wang et al., 2008). However, for the purpose of this dissertation, we will focus on two major histone modifications that is, histone methylation and histone acetylation.

The proviral DNA contains two nucleosomes of importance, nucleosome-0 (nuc-0) and nucleosome-1 (nuc-1) (Rafati et al., 2011, Verdin et al., 1993). Importantly, nuc-1 overlaps with the promoter region of the LTR thereby restricting the activity of the transcription initiator site as well as restricting the ability of cellular TFs such as LBP-1, AP1 and NFAT to bind to the 5-LTR (Van Lint et al., 1997, Yoon et al., 1994). Histone acetylation and methylation play a significant role in either remodelling nuc-1 to activate transcription or maintaining the repressive effect of nuc-1, respectively. Histone acetyltransferases (HATs) are enzymes that promote acetylation of the amino terminals of the ϵ-amino on the lysine residues of H3 and H4 tails. This acetylation promotes remodelling of the nuc-1 to produce a euchromatin configuration that promotes viral gene transcription. In contrast Histone Deacetylases (HDACs) promote deacetylation of the lysine residue and restore the nuc-1 configuration thereby promoting a heterochromatin configuration that represses viral transcription (Archin et al., 2014, Archin et al., 2009, Shirakawa et al., 2013). Although there are four classes of HDACs, HDAC1 is the most dominant and is recruited to the HIV-1 LTR by TFs such as the NF-κB p50 homodimer.

The HDAC inhibitors, as the name suggest, promotes inhibition of HDACs thus promoting active gene transcription. These inhibitors such as suberoylanilide hydroxyamic acid (SAHA) presented as an attractive mechanism for HIV-1 latency reversal because they were active in multiple cell types and did not elicit global T-cell activation (Archin et al., 2009). Unfortunately, these inhibitors may be able to induce the transcription of cancer genes.

The second mechanism of histone modification includes histone methylation. The methylation of the lysine residues of histone tails is mediated by the histone lysine methyltransferases (HKMTs) that catalyse the addition of methyl groups to the amino terminal lysine residue. The histone methylation is associated with changes in the steric and hydrophobic characteristics of the histones that may either promote or repress viral gene transcription. The methylation of H3K4 induces a state of euchromatin and active transcription while methylation of H3K9, H3K27 and H4K20 induces a state of heterochromatin and repressed transcription (du Chéné et al., 2007, Friedman et al., 2011, Kim et al., 2011, Mateescu et al., 2008).

2.9.3. DNA Methylation

The integrated proviral DNA contain two CpG islands located on the 5'- and 3'- end of the HIV-1 transcription start site. The DNA methyltransferases (DMTs) catalyse the addition of 5'-methyl groups to the CpG dinucleotides. The methylation of the CpG islands is associated with transcriptional silencing and a heterochromatin configuration (Blazkova et al., 2009). During HIV-1 latency the CpG islands are hypermethylated and are rather responsible for maintaining viral latency rather than inducing viral latency. Hypermethylated CpG islands directly block the ability of transcription factors to bind to the viral promoter or they may indirectly block transaction factor binding by the addition of methyl-CpG-binding proteins (MBDs) (Bednarik et al., 1991). The MBDs further recruit HMTs and HDACs that promote a repressed heterochromatin configuration thus maintaining a state of viral latency (Blazkova et al., 2009). In addition to this, one of the hypermethylated CpG islands recruit the MBD2 and HDAC2 that enforce a transcriptionally repressive state (Kauder et al., 2009).

2.9.4. Transcriptional Interference

The mechanism of transcriptional interference is dependent on the orientation of the viral promoter integration site. Transcriptional interference can mediate suppression of viral transcription through promoter exclusion or collision. When the viral promoter is integrated into the host genome in the same orientation as a second promoter, the upstream RNA pol II transcription complex causes promoter occlusion by read through transcription that dislodges transcription factors bound to the viral promoter and promotes a state of viral transcription repression (Greger et al., 1998, Lenasi et al., 2008). In addition, this read through can also block the viral promoter from accessing the available cellular transcription factors. In contrast, when the viral promoter is integrated into the host genome in the opposite orientation to that of a second promoter, the two active RNA Pol II transcription complexes eventually collide and result in the premature termination of transcription elongation (Crampton et al., 2006, Lewinski et al., 2005). This may supress either one or both promoters resulting in transcriptional

silencing. However, although the collision of the transcriptional complexes can occur, another possibility exists in which the RNA Pol II complexes may converge and produce double stranded viral RNA that may participate in RNA interference to mediate transcriptional silencing.

The evidence put forward by the various studies suggests that the LTR is a prominent region of interest which could direct further studies that assists in cure development. Several mechanisms of HIV-1 latency development have been described, and latency development may be an interplay of several mechanisms. However, the HIV-1 LTR plays a central role in HIV-1 latency development and therefore studying the HIV- LTR presents an attractive target for future HIV-1 cure strategies. Since many studies have characterized HIV-1 subtype C in countries that include Brazil, India and China, South Africa is requiring further knowledge about HIV-1 subtype C which presents as the main subtype responsible for the pandemic. Thus, the characterization and establishment of the sequence variations exhibited by the South African subtype C will contribute to the broader aim of HIV cure developments.

CHAPTER 3: METHODOLOGY

3.1. Study design, Inclusion and Exclusion Criteria

The study population includes participants from two study cohorts namely, the Females Rising through Education, Support and Health (FRESH) acute infection and, the HIV Pathogenesis Programme (HPP) acute infection. This study strictly included plasma samples that were obtained during the acute phase of infection from patients that were antiretroviral naïve.

The FRESH cohort is an ongoing prospective study as previously described by (Dong et al., 2018). This cohort was initiated to detect acute HIV infection in young females who were at high risk for HIV infection. In doing so, this cohort enables the collection of blood samples prior to HIV infection as well as after HIV infection. Thus, contributing to the study of HIV-1 pathogenesis and the behavioural and biological risk factors associated with the risk of HIV acquisition. This study recruited young females aged between 18-23 years old that were HIV negative but at high risk of HIV acquisition and provided written consent for enrolment. The duration of surveillance was set at 48 weeks since 1st May 2015.

The FRESH study consisted of two objectives. The first objective was socioeconomic empowerment through the attendance of twice-weekly classes. The second objective was consent to participate in a twice-weekly HIV-1 finger prick test to detect acute HIV-1 infection. During the initial 19 months of surveillance, participants diagnosed with acute HIV-1 infection did not begin treatment immediately. Instead, they were monitored and referred for treatment according to the South African eligibility criteria at that time. However, in July 2014, antiretroviral therapy (ART) was administered immediately upon detection of acute HIV-1 infection according to the South African Test and Treat strategy.

Initially, 945 participants were enrolled to begin the 48-week surveillance period. Of this, 42 participants were diagnosed with acute HIV infection and were put on a 156-week monitoring programme. Thereafter, of the 42 acutely infected participants, 14 individuals remained treatment naïve while 28 individuals received early ART treatment (that is, ART was initiated during the acute stage of infection). From the 14 treatment naive participants, 1 individual had withdrawn from the study and only 13 participants remained treatment naïve and were in follow up. Of these 13 individuals, 7 patients began ART treatment while 6 remained treatment naïve. For this study, 4 treatment naïve samples were selected for analyses.

In addition, HIV-1 infected plasma samples were obtained from the HIV Pathogenesis Program (HPP) acute infection cohort previously described (Gounder et al., 2015b). Initially, a total of 22 acutely infected antiretroviral naïve participants were enrolled into the HIV Pathogenesis Programme Acute Infection Cohort based in Durban, South Africa. During screening, all participants exhibited detectable HIV-1 RNA levels but had not yet seroconverted. Therefore, these participants were described as acutely infected according to the CDC's criteria and definition of HIV-1 acute infection. The date of HIV-1 infection was estimated to be approximately 14 days before screening. Participants provided written informed consent and blood samples were collected at intervals of enrolment, at 2 weeks, 4 weeks, 2, 3 and 6 months. Thereafter blood samples were collected every 6 months following infection (Gounder et al., 2015a). The viral load and CD4 count was measured at every time visit using the Roche Cobas Taqman HIV-1 Test v2.0 (Roche Diagnostics, Branchburg, NJ, USA) and the 4-colour MultiTEST/Trucount assay (Becton Dickinson, San Jose, CA, USA), respectively (Gounder et al., 2015a). To date, 44 acutely infected participants were recruited. Plasma samples (24 samples) obtained from two study visits, at or near transmission and at one-year post-infection was made available for this study.

The 25 plasma samples (4 FRESH acute plasma samples and 21 HPP acute plasma samples) used in this study was selected based on the availability of plasma samples at the HIV Pathogenesis Programme (HPP) Laboratory. Furthermore, going forward into to the thesis, the study participants will be referred to as acutely infected antiretroviral therapy naïve patients. Approval for this study was obtained from the Biomedical Research and Ethics Committee of the University of Kwa-Zulu Natal (REF: BREC/00001051/2020).

3.2 Polymerase Chain Reaction (PCR) Amplification

First round PCR products were used as a template for nested PCR which was carried out using Platinum® Taq DNA Polymerase High Fidelity PCR Kit (Invitrogen, Thermofisher Scientific, Massachusetts, United States) as per the manufacturers' instructions. Briefly, the PCR reaction was set up by mixing 1X High Fidelity PCR Buffer, 0.1 µM sense primer (T7-U3-M-F:5'taatacgactcactatagggTTTTTAAA AGAAAAGGGGGGAC-3'_nt 9064 to 9085), 0.1 µM antisense primer (3'Sp6 R-M-R: 5'atttaggtgacactatagATTGAGGCTTA AGCAGTG GGTT-3'_nt 9614 to 9593), 0.2 µM dNTP mix, 2 µM MgSO4, 1 U Taq DNA Polymerase, 2 µL of template DNA (T/F LTR PCR1 products, which is already available in the laboratory) and this was made to a final PCR reaction volume of 50 µL with PCR grade water. The thermocycler conditions for the PCR reaction included an initial

denaturation at 94˚C for 15 seconds, 25 cycles of denaturation at 94˚C for 15 seconds, annealing at 55˚C for 30 seconds, extension at 68˚C for 60 seconds and a final extension at 68˚C for 5 minutes.

3.3 Agarose Gel Electrophoresis

The PCR products were analysed on a 1% agarose gel to confirm the size and specificity of the amplified band. The 1 Kb Plus DNA Ladder (Invitrogen, Thermofisher Scientific, Massachusetts, United States) was used as a molecular weight marker and the gel was run at 95v for 1 hour as established in the laboratory.

3.4 Gel Extraction

The nested PCR products were gel extracted using the GeneJet Gel Extraction Kit (Thermofisher Scientific, K0692) as per manufacturer's instructions. Briefly, 45 µL of the nested PCR product was run on a 1% agarose gel that was not exposed to ultraviolet light prior to extraction to protect the integrity of the DNA contained within the gel bands. In preparation for gel extraction, an empty 1.5 mL Eppendorf was pre-weighed to establish the weight of the empty Eppendorf. Thereafter, the specific PCR bands were extracted using a gel extractor (Merck, Germany, Darmstadt). The gel slice was then transferred to the pre-weighed 1.5 mL Eppendorf and thereafter the Eppendorf containing the extracted gel slice was weighed again. The weight of the gel slice was then calculated (Weight of the gel slice in Eppendorf – Weight of the pre-weighed empty Eppendorf = Weight of the gel slice) to determine the volume of binding buffer required in the subsequent step.

Thereafter, Binding buffer was added to each tube containing the gel slice in a 1:1 ratio (Add 100 µL of binding buffer per 100 mg of gel slice) and incubated at 50 - 60 °C for 10 minutes in a dry heating block. The dissolved gel slice mixture was then briefly vortexed for 10 seconds and transferred to a GeneJet column followed by centrifugation at 13 000 rpm for 1 minute. The flow through was then discarded and 100 µL of Binding buffer was then added to the columns and centrifuged at 13 000 rpm for 1 minute. Thereafter, the flow through was discarded and 700 µL of Ethanol Wash solution was added to each column followed by centrifugation at 13 000 rpm for 1 minute. The flow through was once again discarded and the empty column centrifuged for an additional 1 minute to remove the residual ethanol from the wash solution. The column was then transferred to the labelled 1.5 mL Eppendorf's and 20 µL of elution buffer was added to the centre of each column. The columns were then incubated at room temperature for 2 minutes followed by centrifugation at 13 000 rpm for 1 minute. The eluted DNA was stored at -20 ° or used immediately.

3.5 Cloning

3.5.1 Restriction digestion and ligation

Briefly, the gel extracted nested PCR product was cloned into the pGL3 Basic vector (Promega, USA, Madison) which contains the luciferase reporter gene (Gray et al., 2013). Firstly, the pGL3 Basic vector was digested with the two sticky ends restriction enzymes, Kpn1 (New England Biolabs, USA, Massachusetts) and HindIII ((New England Biolabs, USA, Massachusetts). The restriction digestion reaction was then run on a 1% agarose gel for 1 hour at 95v as established in laboratory protocol. This gel was not exposed to ultraviolet light prior to extraction to protect the integrity of the plasmid DNA contained within the gel band. In preparation for gel extraction, an empty 1.5 mL Eppendorf was pre-weighed to establish the weight of the empty Eppendorf. Thereafter, the specific PCR bands were extracted using a gel extractor (Merck, Germany, Darmstadt). The gel slice was then transferred to the pre-weighed 1.5 mL Eppendorf and thereafter the Eppendorf containing the extracted gel slice was weighed again. The weight of the gel slice was then calculated (Weight of the gel slice in Eppendorf – Weight of the pre-weighed empty Eppendorf = Weight of the gel slice) to determine the volume of binding buffer required in the subsequent step. The gel purification of the linearized pGL3 was performed using the GeneJet Gel extraction kit (Thermofisher Scientific, K0692) as per the manufacturer's instruction as we had briefly highlighted above.

The same sticky end restriction enzymes (Kpn1and HindIII) were used to digest the nested PCR product, patient derived HIV-1 LTR U3R region to create the same sticky ends. The digested patient derived U3R region was cloned into the linearized pGL3 Basic vector followed by the ligation using 1 U of T4 DNA ligase (New England Biolabs, USA, Massachusetts) as per manufacturer's instructions. Briefly, a recombinant pGL3 Basic vector was created by preparing a 20µL of the ligation reaction mixture containing 37.5 ng of insert DNA, 50 ng of plasmid DNA, 1X T4 DNA ligase buffer, 1 U of T4 DNA ligase with the reaction volume being made up to 20 µL using nuclease free water. The ligation reaction was then incubated for 10 minutes at room temperature and thereafter returned to ice for the subsequent transformation reaction.

3.5.2 Transformation

Following ligation, transformation of the recombinant pGL3 plasmid into JM109 competent *E. coli* cells (Promega, USA, Madison) was performed as per manufacturer's instructions. JM109 competent *E.coli* cells were thawed on ice for no more than one minute. Following thawing, 10 µL of the

recombinant DNA molecules was added to 50 µL of pre-chilled JM109 competent *E. coli* cells immediately followed by incubation on ice for 30 minutes. The competent *E. coli* cells and recombinant DNA was heat shocked in a water bath at 42˚C for 20 seconds and the Eppendorf tube was subsequently chilled on ice for 2 minutes. Thereafter, 450 µL of the SOC medium was added to the heat shocked cells followed by incubation at 37 °C while shaking at 250 rpm for one hour. Thereafter, 100 µL of the transformation reaction was added to the ampicillin containing LB agar plate and spread over the surface of the plates. The plates were incubated at 37˚C for 16-20 hours.

3.5.3 Plasmid Miniprep

Following the transformation reaction, the agar plates were checked after a 16- hour overnight incubation at 37 ˚C to confirm bacterial growth. The inoculated agar plates contained numerous single colonies upon successful transformation. Thereafter, ampicillin containing Lysogeny Broth (LB) Broth was prepared by aliquoting 3.0 mL of broth and 100 µg/ mL of ampicillin in a tube. Prior to inoculation of the LB broth with a bacterial colony, a master plate was prepared to store the bacterial colonies used for inoculation. A single bacterial colony was picked up from the lawn of bacteria and this was then used to inoculate a labelled master plate. Subsequently, the same bacterial colony was inoculated into a single tube of ampicillin containing LB broth. The inoculated tubes were then placed in a horizontal shaker with shaking at 230 rpm at 37˚C overnight.

Following bacterial growth in the LB broth, the LB broth had then changed to a cloudy appearance confirming bacterial growth in the LB broth. The plasmid DNA was then isolated from the bacterial colonies using the GeneJet Plasmid Mini Prep Kit (Invitrogen, Carlsbad, CA) as per the manufacturers' instructions. Briefly, the bacterial cells from the LB broth were harvested via centrifugation (Heraeus Biofuge Fresco, England, United Kingdom) at 8000 rpm for 2 minutes to form a pellet. The bacterial pellet was then resuspended in 250 µL of Resuspension Solution while taking caution to ensure a clump free suspension. Lyses of the bacterial cells was achieved by the addition of 250 µL of Lysis Solution followed by the addition of 350 µL of Neutralization Solution ensuring to mix thoroughly. The removal of the cell debris and chromosomal DNA was achieved by centrifugation (Heraeus Biofuge Fresco, England, United Kingdom) for 5 minutes at 13000 rpm

The supernatant consisting of the plasmid DNA was transferred to a GeneJet spin column followed by centrifugation (Heraeus Biofuge Fresco, England, United Kingdom) at 13000 rpm for 1 minute. Subsequently, the flow through was discarded and 500 µL of Wash Solution was added to the GeneJet

spin column followed by centrifugation (Heraeus Biofuge Fresco, England, United Kingdom) for 30-60 seconds at 13000 rpm. The wash procedure was repeated with an additional centrifugation (Heraeus Biofuge Fresco, England, United Kingdom) for 1 minute at 13000 rpm to ensure removal of the ethanol wash solution. Thereafter, the GeneJet spin column was transferred to a clean 1.5 mL Eppendorf with subsequent addition of 30 µL of pre-warmed Elution Buffer to the spin column. Following incubation for 2 minutes at room temperature the plasmid DNA was centrifuged (Heraeus Biofuge Fresco, England, United Kingdom) for 2 minutes at 13000 rpm. The DNA concentration of the prepared plasmid DNA was quantified using Nanodrop™. The plasmid DNA was then used immediately or stored at -20˚C.

3.6 Sequencing

The plasmid miniprep recombinant clones was then sequenced directly using the Big Dye Terminator v3.1 Sequencing Kit (Thermofisher Scientific, Massachusetts, United States). A sequencing master mix was prepared by the combination of 3.4 µL of PCR grade water, 2 µL sequencing buffer, 2 µL of either the 0.4 µM sense primer (T7-U3-M-F:5'taatacgactcactatagggTTTTTAAA AGAAAAGGGGGGAC-3'_nt 9064 to 9085) or the 0.4 µM antisense primer (3'Sp6 RMR: 5'atttaggtgacactatagATTGA GGCTTAAGCAGTG GGTT-3'_nt 9614 to 9593), 0.4 µL of BigDye v3.1 and 1 µL of DNA template (20ng/µL). Cycle sequencing reaction was performed separately for each primer. The reaction was centrifuged (Eppendorf 5810R) and placed in a thermocycler with the following thermocycling conditions: pre-denaturation at 96˚C for 1 minute, 25 cycles of denaturation at 96˚C for 10 seconds, annealing at 50˚C for 5 seconds, extension at 50˚C for 4 seconds and held at 4˚C. The sequencing plates were stored at 4˚C and protected from exposure to light.

Subsequently, after the sequencing reaction, 1 µL of 125mM EDTA was added to each well and mixed thoroughly. Thereafter a mixture of 1 µL of 3M NaOAc (pH 5.2) and 25 µL of 100% ice cold ethanol was added to each well. The plate was then vortexed thoroughly and centrifuged (Eppendorf 5810R) for 20 minutes at 3000 rpm. The plate was subsequently inverted onto a paper towel and centrifuged (Eppendorf 5810R) again for 1 minute at 150 x g. Following that, 35 µL of 70% ethanol was added to each well and centrifuged (Eppendorf 5810R) for 5 minutes at 3000 rpm. Subsequently the plate was inverted onto a paper towel and centrifuged for 1 minute at 150 rpm. The samples were then dried in a thermocycler for 5 minutes at 50˚C. The plate was then sealed using adhesive foil and stored at -20˚C in preparation for sequencing

3.7 Sequence Analysis

Sequence analysis was performed using Sequencher 5.1. The sequences were assembled into contigs and edited manually using Sequencher 5.1 (Gene Codes, Ann Arbor, Michigan, USA).Thereafter, a multiple sequence alignment was performed using Jalview v2.10.5 (Waterhouse et al., 2009). The sequences were then examined on Bioedit Sequence Alignment Editor v7.2.5 (Hall, 1999) to detect mutations within the U3R region of the patient LTR elements. Similarity between PCR-derived sequences and clonal sequences were determined by the construction of a phylogenetic tree using the online tool Phyml (https://www.hiv.lanl.gov/content/sequence/PHYML/interface.html). The phylogenetic tree was then viewed and edited using FigTree Tree Figure Drawing Tool v1.4.3.

3.8 Transfection into Jurkat Cells

Briefly, Jurkat cells were thawed in a 37˚C water bath for no longer than 3 minutes. The cells were grown as suspension cells in RPMI 1640 medium supplemented with 10% Fetal bovine serum (FBS), penstrep and L-Glutamine (R10 medium). The cells were grown to a viability of 95% or higher in preparation for transfection. Jurkat cells were transiently transfected with Lipofectamine 2000 transfection reagent (catalog no.11668-019, Invitrogen) as previously described by Bachu et al., (2012).

Briefly, Jurkat cells were seeded into 24-well tissue culture plates at a density of 5 $x10_5$ cells/well in 400 µL of antibiotic-free RPMI 1640 medium supplemented with 10% FBS. Thereafter, 300 ng of the recombinant LTR- containing pGL3 plasmid was prepared in 50 µL of serum-free RPMI medium. For the LTR-Tat (WT subtype C) or LTR-Tat (autologous) co-transfection assay, a plasmid pool of 300 ng of the recombinant LTR-containing pGL3 plasmid and 100 ng of WT subtype C Tat expression vector or autologous tat expression vector was prepared in 50 µL of serum-free RPMI medium. A total of 1 µL of Lipofectamine was mixed with 49 µL of serum-free RPMI medium to prepare the lipid transfection reagent. Then, the 50 µL lipofectamine-RPMI mixture was mixed with 50 µL of the recombinant pGL3 plasmid DNA. The plasmid-lipid mixture was then incubated for 20 minutes at room temperature and subsequently added to the appropriate wells. Twelve hours following the transfection, the cells were washed to remove the lipid complexes and resuspended in 500 µL of R10 medium. The transfection reaction was then incubated for 24 hours followed by a luciferase assay.

3.9. Cell Activation Assay

Jurkat cells were transiently transfected or co-transfected with the autologous Tat expression vector as described in section (3.8). Twelve hours following the transfection, the cells were washed to remove the lipid complexes and resuspended in 500 µL of R10 medium. The transient transfection was then stimulated with one of the following cell activators: TNF-α (20 ng/mL), PMA (20 ng/mL), SAHA (20 ng/mL) or Prostratin (20 ng/mL) as previously described (Bachu et al., 2012). The stimulated cells were then incubated for 24 hours after which the luciferase activity was measured.

3.10 Luciferase Assay

Bright-Glo (Promega, Madison, United States) was thawed at an ambient temperature water bath away from light. Thereafter 350 µL of the culture medium was aspirated from the 500 µL transfection reaction such that 150 µL of culture medium remained in each well. Subsequently, 100µL of Bright-Glo was added to each well followed by incubation for 2 minutes with the plate being shielded from light. The cells in each well were then mixed by two pipette strokes to lyse open the cells followed by the transfer of the 150µL of lysed cells into the wells of a black round bottom 96 well plate. The expression of the Luciferase was read immediately using the Victor Nivo Multimode plate recorder (PerkinElmer, Massachusetts, USA).

3.11 Statistical Analyses

Statistical analysis was performed using Graphpad Prism 5 software for Windows, GraphPad Software, San Diego California USA, (www.graphpad.com). The statistical significance for the transfection assay and association of viral load with LTR mutants was determined using a paired T-test. Linear regression analyses were used to determine the statistical significance and correlation coefficient of the correlation between transcription activity and markers of disease progression. The p value <0.05 was considered statistical significance

CHAPTER 4: RESULTS

4.1. Demographic Characteristics of 25 Patients from HPP and FRESH Cohort

The U3R region of 25 patient LTR elements was amplified from viral RNA obtained from patients in the HPP and FRESH acute infection cohorts. The amplification of the U3R region was performed from viral RNA obtained at two study time points per patient. Due to sample availability, the first time point was obtained close to transmission, that is, at a median of four weeks post-infection (denoted as B2) [patient plasma samples were Western blot negative but HIV-1 RNA positive] and the second time point was at one-year post infection (denoted as B7). The selection of these two time points allowed us to examine the diversification of the LTR elements by one-year post-infection. In addition, phylogenetic analysis was used to determine interpatient LTR genetic variability.

4.2. LTR Diversification of the T/F Virus During Acute Infection

To examine the LTR diversity of the 25 patients, the U3R region of LTR elements were amplified at the aforementioned timepoints and then directly sequenced from the PCR products to generate bulk sequences. A phylogenetic tree was then constructed to determine whether there is interpatient LTR genetic variability as well to determine the diversification of the T/F virus patients LTR by one-year post infection (Figure 8). The 25 patients were infected with HIV-1 subtype C viruses and the phylogenetic tree was rerooted against the Indian subtype C reference sequence (EF17 8613.1). Phylogenetic analysis demonstrates that each patient LTR sequences formed independent clusters and there was no indication of any interpatient relatedness. This further demonstrates that each patients' infection is as a result of an unrelated T/F viral variant and confirms interpatient LTR genetic variability. Phylogenetic branching showed that the LTR is subject to evolution within a patient within one-year of infection. This was shown in 14 out of 25 patients (56%), while 11 out of 25 (44%) patients maintained homogeneous LTR sequences within one-year post-infection. The analysis of *env* sequence data would be valuable to better understand the transmission pairs. However, *env* sequence analysis was beyond the scope of the study presented in this dissertation. The env sequence data of the transmission pairs will be analysed and included for the publication purposes..

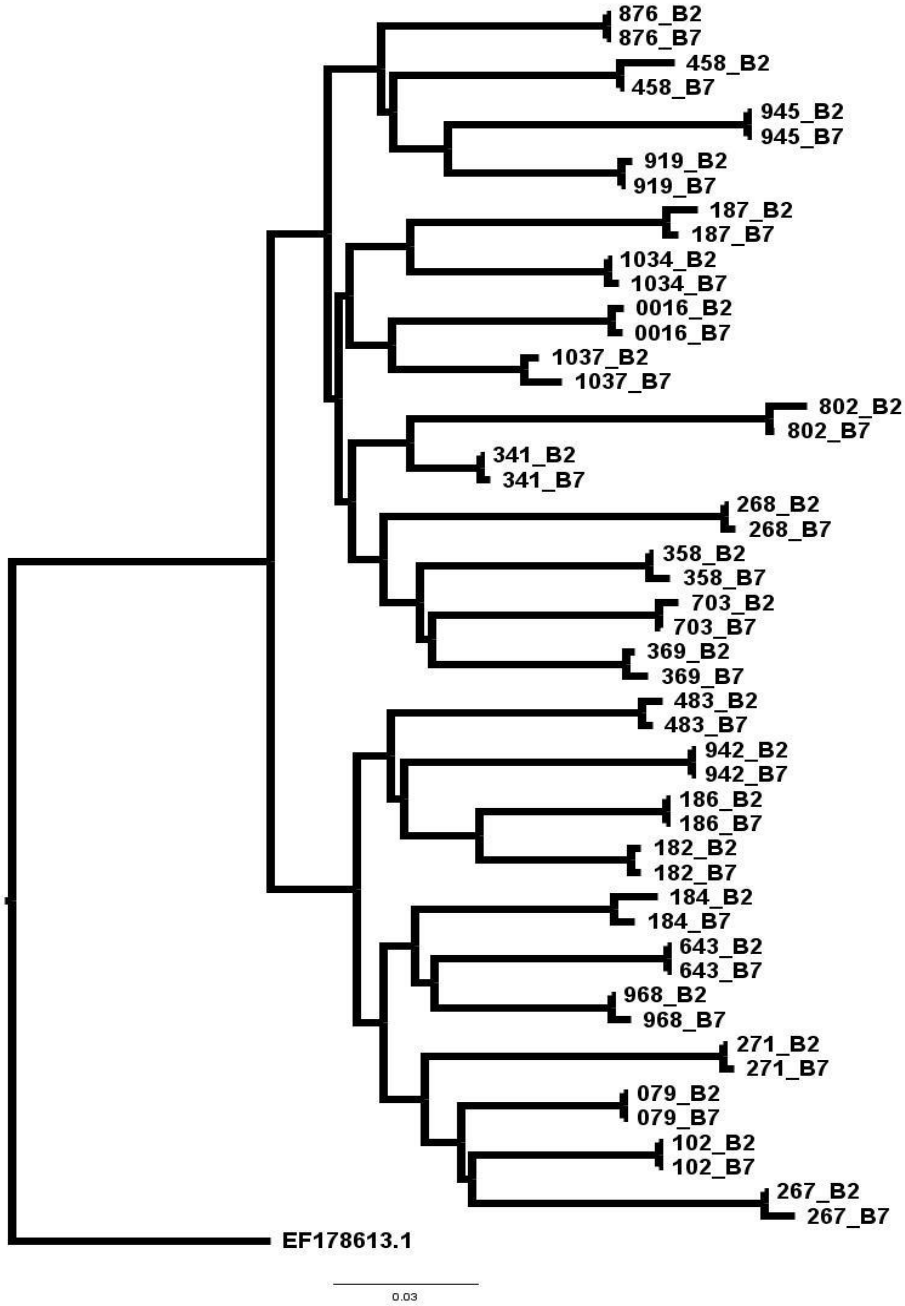


Figure 8. **Phylogenetic analysis demonstrates interpatient LTR sequence variation**. Phylogenetic analysis was performed using 25 patient derived LTR elements from the HPP and FRESH cohort. The tree was constructed using the online tool Phmyl (http://www.hivlanl.gov) and rerooted on the Indian subtype C reference sequence (EF178613.1) using FigTree Tree Figure Drawing Tool v1.4.3. The viral sequences obtained near transmission are denoted as B2 while the viral sequences obtained at one-year post infection are denoted as B7. Each patients LTR element clusters independently together thus demonstrating that the LTR element is variable between patients. Phylogenetic branching demonstrates that the LTR may evolve within one-year of infection within a patient.

4.3. The South African Subtype C U3R region is Variable at Transmission

Next we hypothesized that there may be genetic variation within the transcription factor binding sites of the U3R region as previously reported for HIV-1C circulating in India (Bachu et al., 2012). To this effect, the U3R region sequences aligned with the Indian subtype C reference sequence (EF178613.1) using Jalview v2.10.5 (Waterhouse et al., 2009) since this reference sequence contained more than 3 NF-κB binding sites. Interestingly, the transcription factor binding sites (TFBS) USF, RBE III site, the Sp1 I, Sp1 II and Sp1 III sites, the TATA Box, E- Box, 2 conserved NF-kB binding sites (NF-kB I and NF-kB II), an additional 3rd NF-kB binding site specific to subtype C and TAR region clearly aligned with the TFBS in the reference sequence. Specifically, the USF region within the modulatory domain and the TAR loop region within the R region demonstrated interpatient LTR variability. However, the core enhancer and promoter region were relatively well conserved with the exception of the Sp1 III binding site that exhibited a greater degree of interpatient variability within the core promoter region.

The USF transcription factor binds to the USF binding site. The USF transcription factor binds the E-box sequence that overlaps the -170 C/EBP region within the modulatory domain (di Fagagna et al., 1995, Schwartz et al., 2000, Tesmer et al., 1993). This interaction was reported to activate transcription in T-cells (Naghavi et al., 2001). Interestingly, our results demonstrated that the USF region in this cohort is not conserved. This region primarily exhibited variation at position 4 of the USF binding site sequence (CAC**C**TGGCCC → CAC**A/T**TGGCCC). This nucleotide change was observed in 13 out of 25 patients (52%). The C → A nucleotide change at position 4 is consistent with a previous study however, this is not a subtype C specific mutation as it was also shown to occur in subtypes D, E and F (Jeeninga et al., 2000) (**Figure 9**).

The RBE III site, located in the modulatory region, is a binding site for the cellular factor RBF-2 (Ras Response Element-2). Overall, all patient LTRs' contained the complete canonical sequence GACTGCTGA for the RBE III binding site (**Figure 9**). The RBE III site (GACTGCTGA) is well conserved in subtype C as previously reported. However, our results demonstrated conservation of the RBE III site in 22 out of the 25 (88%) patient RBE III binding sites while the RBE III site variability was observed in 3 out of the 25 (12%) patient LTR element (patient AS02-802, AS02-945 and AS02-182). In patient 802, the C nucleotide at position 6 was substituted for an A nucleotide (GACTG**C**TGA → GACTG**A**TGA), however, in patient 945 and 182 we observed a CTGAGA and CTGCTGACTTTG sequence insertion within the RBE III site, respectively.

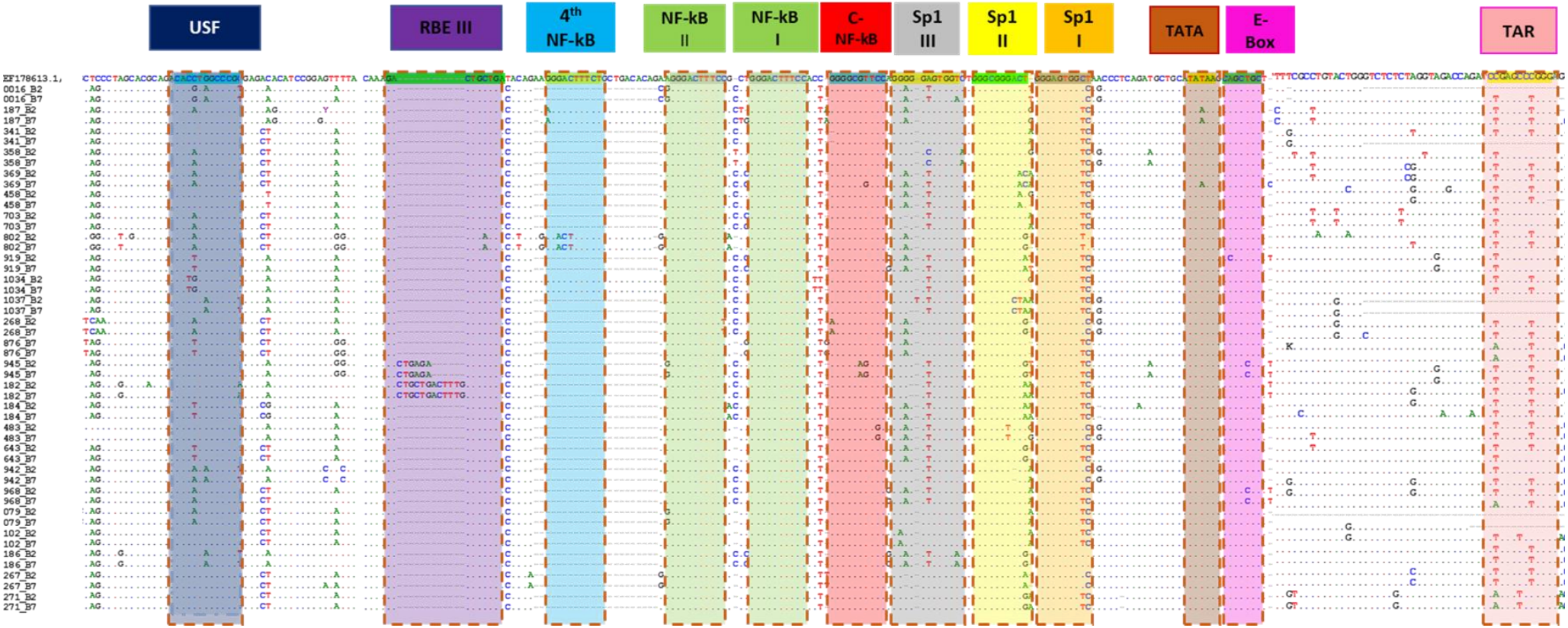


Figure 9. **Multiple sequence alignment demonstrating PCR-derived sequence of the modulatory, core enhancer and core promoter region within the patient LTR elements.** The LTR element sequences were aligned against the Indian subtype C reference sequence, EF178613.1. The dots represent nucleotide bases that are identical to the reference sequence while the dashes indicate deletions. The various coloured labelled blocks represent the transcription factor binding sites within the enhancer and promoter region. From left to right: the USF binding site is represented by blue shading, RBE III site is represented by purple shading, the 4th NF-κB (designated F-NF-κB) is represented by light blue shading while the two canonical NF-κB binding sites (designated NF-κB I and II) are shown by green shading. The 3rd-NF-κB site (designated C-NF-κB) is shown by red shading, the three Sp1 I, II and III binding site are represented by grey, yellow and orange shading respectively. The Tata box and E-box are represented by brown and dark pink shading respectively, while the TAR loop region is shaded with light pink. Nucleotide sequence analysis confirms interpatient variability within the TFBS. Overall, the T/F virus canonical sequences were relatively conserved within the core enhancer, nonetheless variation was observed within the RBE III site, Sp1 III binding site and TATA Box. Overall, the T/F virus was conserved within a patient at transmission and one-year post infection.

Upstream from the RBE III site is the enhancer region which exhibits variation due to differences in NF-κB binding site copy number. Consistent with the previous study (Baar et al., 2000), our data demonstrate the two canonical NF-kB binding sites were conserved among all 25-patient derived LTR elements. The NF-κB binding site copy number is subtype dependent. Most HIV-1 subtypes exhibit two NF-κB binding sites with subtype C exhibiting at least 3 NF-κB binding sites (Baar et al., 2000, Bachu et al., 2012, Boullosa et al., 2014, Hunt and Tiemessen, 2000, Hunt et al., 2001, Scriba et al., 2002). The identification of the 3rd NF-kB (C-NF-κB) binding site confirms that South African cohorts used in this study are infected with HIV-1 subtype C. The 3rd- NF-κB binding site was observed in 100% of patient isolates with 4 out of 25 (16%) patients having an NF-κB like site at the third NF-κB position. NF-κB-like binding sites are those NF-κB binding sites that have single nucleotide changes and thus resemble NF-κB binding sites (**Figure 9**).

Interestingly, a unique characteristic of the subtype C viruses is the occurrence of a fourth NF-κB binding site (Bachu et al., 2012). Consistent, with studies by Hunt et al., (2001), Scriba et al., (2000), Bachu et al., (2012) and Boullosa et al., (2014), our data demonstrates the occurrence of the 4th-NF-κB in South African subtype C viruses (**Figure 9**). However, while viruses containing the 4th NF-κB binding site show an increased expansion in India, our data demonstrates that viruses containing the 4th NF-κB binding site are found at a low frequency in South Africa. Specifically, the 4th NF-κB binding site was observed in 2 out of 25 (8%) patient derived LTR elements and this was observed for both the LTR elements obtained at or near transmission and as well as at one-year post infection (patients AS21-187 and AS02-1034). This finding is consistent with previously reported findings from subtype C viruses circulating in South Africa (Hunt et al., 2001, Scriba et al., 2000).

Consistent with the published data (Baar et al., 2000, Jeeninga et al., 2000), our data demonstrate that the core promoter domain consists of 3 tandem Sp1 binding sites the TATA box and E-Box. The Sp1 III binding site exhibited the most variability in comparison to the Sp1 I and Sp1 II binding sites consistent with previous studies (Bachu et al., 2012, Boullosa et al., 2014, McAllister et al., 2000, Nonnemacher et al., 2004). The subtype C consensus Sp1 III sequence was only observed in 1 out of 25 (4%) patient derived LTR elements (patient AS02-341). However, in the remaining 24 out of 25 (96%) patients, two dominant nucleotide changes existed among this set of patient sequences. The consensus nucleotide (G) at position two was substituted for an A nucleotide (GGGGAGTGG → G**A**GGAGTGG) in 13 out of 25 (52%) patients while at position 5, the consensus nucleotide (A) was substituted for a T nucleotide (GGGGAGTGG → GGGG**T**GTGG) in 15 out of 25 (48%) patients (**Figure 8**). These G2A and A5T mutations existed as either independent single nucleotide mutations in 12 out of 25 (60%) patients or as co-mutations in 11 out of 25 (44%) patients. The Sp1 I and Sp1 II

binding sites were relatively conserved to the consensus sequence, exhibiting a single nucleotide change at position 9 and 10 and at position 10, respectively.

Downstream of the Sp1 binding sites, the TATA box was relatively conserved in 24 out of 25 (96%) patient samples. The TATA Box is found upstream of the transcription start site and has a conserved nucleotide sequence in all HIV-1 subtypes with the exception of subtype E (Montano et al., 1998, Montano et al., 1997). Particularly in subtype E viruses, the TATA box sequence is polymorphic having a single nucleotide change at position three (TA**T**AA ➔ TA**A**AA). Interestingly, our data demonstrate that one patient (187) was infected with a subtype C strain harbouring this TATA box mutation (TA**T**AA ➔ TA**A**AA). Therefore, the TATA box mutant was rare (4%) among the viruses obtained from this set of PLHIV in South Africa (**Figure 9)**.

3.4. Phylogenetic Analysis Confirms Similarity between Bulk Sequence and pGL3-LTR Recombinant Clones

Here, we hypothesized the subtype C T/F viruses LTR derived clones will have the same sequences as the bulk sequences since T/F viruses are homogenous. To this effect a total of 14 representative patient derived LTR element sequences were selected to assess their transcription activity. The 14 representative patient samples selected for the downstream assays were the preliminary sequence data for this study and had demonstrated sequence variations of interest which include the 4$_{th}$-NF-κB binding site, the Sp1 III: G2A and Sp1 III:A5T mutant, the TATA box: T3A mutant.

Specifically, we selected sequences exhibiting variation within the RBE III site, NF-κB binding site and copy number, Sp1 III binding site, TATA Box, E-box and this was in addition to other mutations. These 14 LTR elements were amplified using nested PCR, cloned into the pGL3 Basic vector ((Promega, USA, Madison) and transformed into *E. coli* competent cells (Promega, USA, Madison). A single random colony was selected per patient derived LTR pGL3 recombinant clone and sequenced. The patient derived LTR sequences spanning the core enhancer and core promoter region were analysed because variation within this region has been reported to dysregulate viral gene transcription. Therefore, this phylogenetic analysis was restricted to the core enhancer and promoter region only.

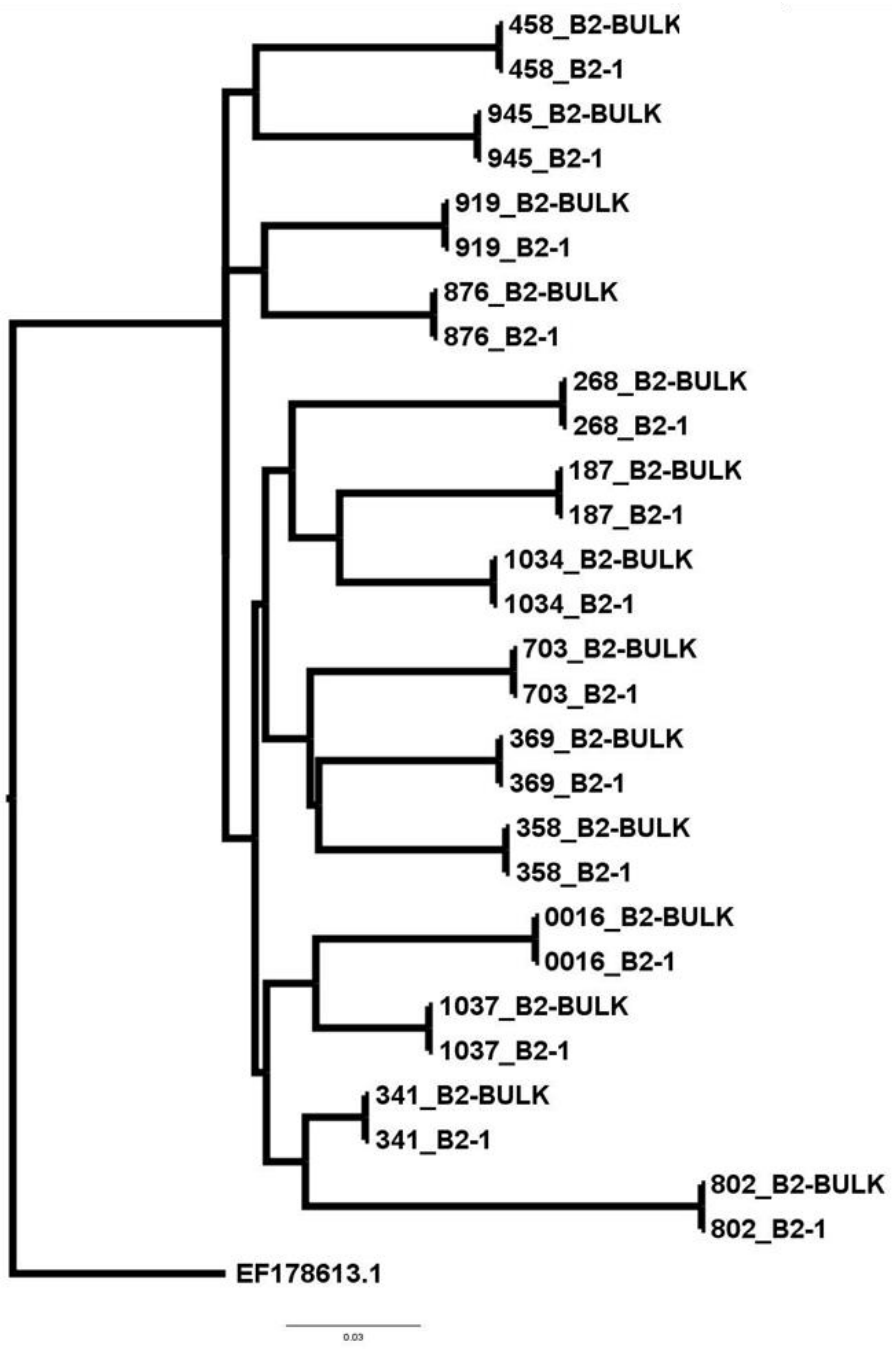


Figure 10. **Phylogenetic tree demonstrating similarity between the bulk sequence and the clonal sequences**. Phylogenetic analysis was completed using the 14 patient derived LTR elements from the HPP cohort. The tree was re-rooted on the Indian subtype C reference sequence (EF178613.1) using FigTree Tree Figure Drawing Tool v1.4.3. Both bulk and one clonal sequence was generated for each of the 14 patients and the bulk sequence and clonal sequence clustered together per patient. The patient identities are denoted with B2 indicative of viral sequences close to the transmission event.

A single randomly picked and sequenced clone (referred to as clonal sequence) together with the respective patient bulk sequence formed independent clusters. The monophyletic structure of each cluster indicates that there was no contamination or intermingling of the sequences between patients. Furthermore, the branching structure of the clonal and bulk sequence indicates homogeneity of the T/F viruses within a patient (**Figure 10**). Therefore, phylogenetic analysis confirms that the clonal sequence contains an identical nucleotide sequence to that of bulk sequence near transmission.

4.5. Nucleotide Sequences confirms the Bulk and Clonal Sequences are Identical within the Core Enhancer and Core Promoter Region

Although the phylogenetic tree confirmed that the bulk and representative clone is identical, we performed a nucleotide alignment of the bulk and clonal sequence to examine the modulatory region, core enhancer and core promoter region at the nucleotide level. The bulk sequence and clonal sequence showed 100% alignment thus demonstrating that each patients' representative clone accurately represented the modulatory region, core enhancer and core promoter of the PCR-derived sequence (**Figure 11**). This single representative clone was selected for the downstream functional assays.

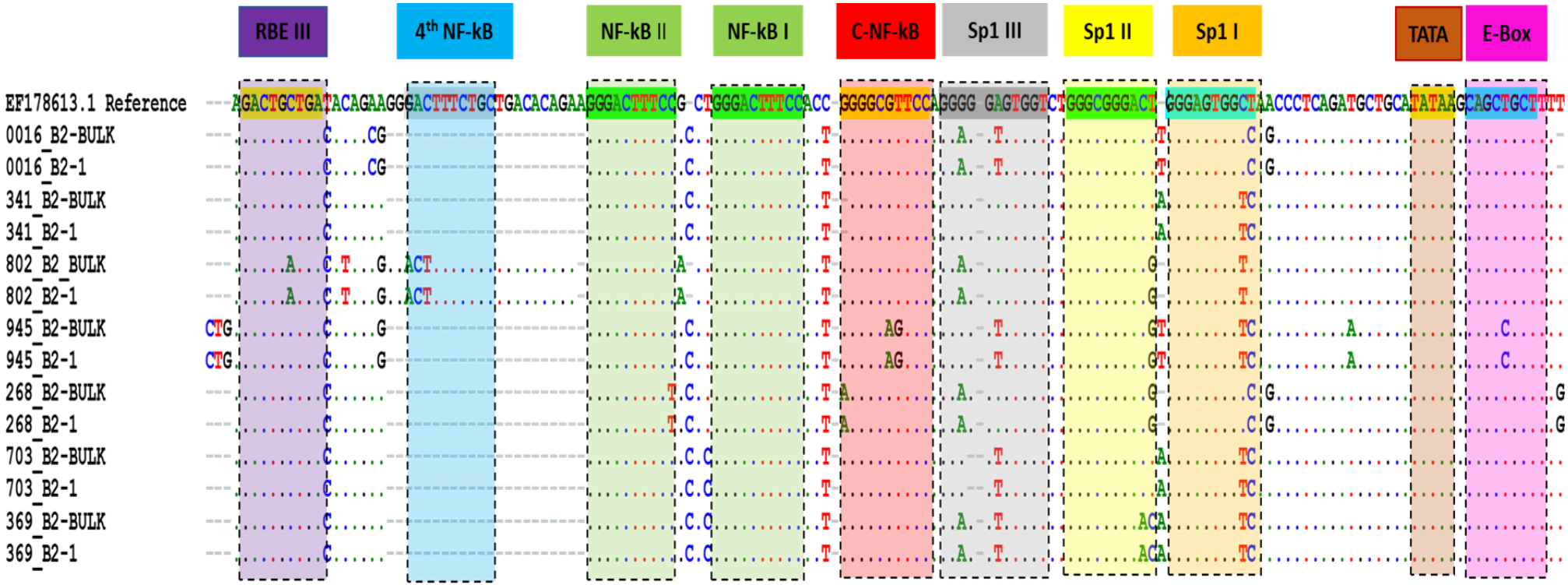


Figure 11 **Multiple sequence alignment demonstrating both clonal and PCR-derived sequence of enhancer and promoter region within the patient LTR elements.** The LTR element sequences were aligned against the Indian subtype C reference sequence, EF178613.1. The dots represent nucleotide bases that are identical to the reference sequence while the dashes indicate deletions. The various coloured labelled blocks represent the transcription factor binding sites within the enhancer and promoter region. From left to right: the RBE III site is represented by purple shading, the 4th NF-κB (designated F-NF-κB) is represented by blue shading while the two canonical NF-κB binding sites (designated NF-κB I and II) are shown by green shading. The 3rd-NF-κB site (designated C-NF-κB is shown by red shading, the three Sp1 I, II and III binding site are represented by grey, yellow and orange shading respectively, while the Tata box and E-box are represented by brown and pink shading respectively. Shown above are the PCR-derived sequence along with the respective clonal sequence for 7 patients are representative of the 14 patients that were cloned. The nucleotide alignment demonstrates that the clonal sequence is identical to the PCR-derived sequence.

4.6. Differential Transcriptional Activity of HIV-1 Subtype C T/F viruses LTR variants Between Patients

Next we hypothesized that sequence variation within the transcription factor binding sites (TFBS) may alter the ability of the LTR to bind their respective transcription factors. Especially, the Sp1 III binding site where we observed the most variation. Previously, studies have analysed the impact of intra- and inter-subtype variation on LTR functional activity. Mostly these studies have addressed the impact of subtype specific differences on LTR driven gene transcription as well as the effect of NF-κB copy number on viral gene transcription only during chronic phase of infection where the virus has already diversified (Bachu et al., 2012, Jeeninga et al., 2000). Here, we assessed the basal and Tat induced patient derived LTR transcription activity of viruses obtained from 14 patients. The patient derived LTRs were cloned upstream of the luciferase gene present in the pGL3 Basic vector and the recombinant plasmids were transfected into Jurkat cells to assess differential LTR activity. A conserved subtype C LTR was used as the wild-type or consensus subtype C LTR (WTC-LTR) obtained from the NIH Reagents Programme. The WTC-LTR was amplified and cloned into the pGL3 Basic vector followed by sequencing. The WTC-LTR sequence was confirmed against the Indian subtype C reference sequence. However, the WTC-LTR used in this study only consisted of three NF-κB binding sites.

Firstly, we compared the LTR elements basal level of gene transcription. In the absence of the Tat protein, all patient LTRs were functional promoters that demonstrated a low level of gene transcription activity (**Figure 12A**). Though most patient LTR elements demonstrated gene expression levels that were comparable to that of the WTC-LTR, in particular, two patient LTR elements (AS02-802 and AS03-268) demonstrated levels of basal gene transcription that were below that of the WT-LTR (Wild-type LTR). Interestingly, these patients contained the Sp1 III G2A mutation with one patient (AS02-802) also exhibiting an RBE III C6A mutation. This suggests that the Sp1 III G2A mutation in combination with other mutations within the core promoter and enhancer region may decrease the basal transcriptional activity of the viral promoter. In contrast, we observed 3 patient LTR elements (AS02-358, AS03-369 and AS01-919) that showed an approximate 3-4-fold higher level of basal gene expression when compared to the WTC LTR promoter. These 3 patient LTRs exhibited a Sp1 III A5C (patient AS02-358) and Sp1 III A5T and G2A (patient AS03-369 and AS01-919) mutation respectively. This further suggest that the Sp1 III A5T mutation in combination with other mutations within the LTR may enhance the basal transcriptional activity of the patient LTRs.

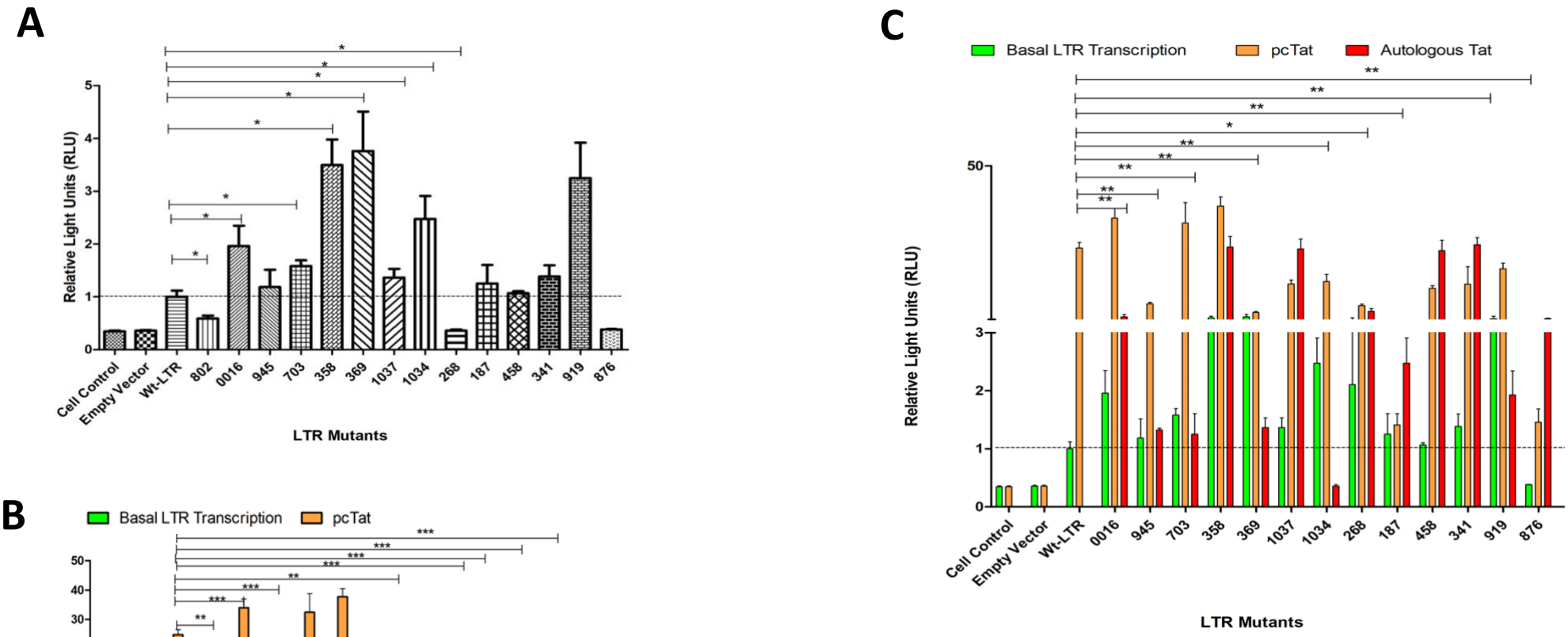
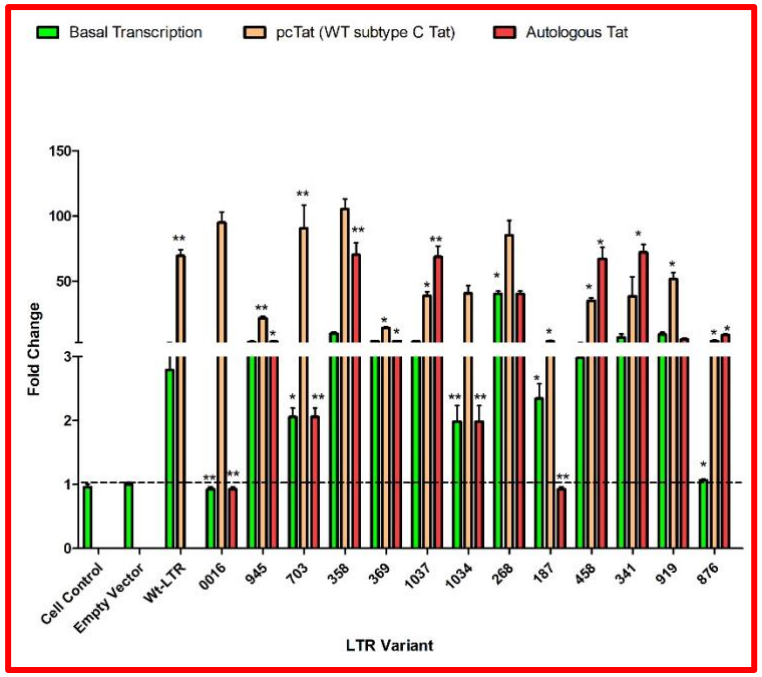
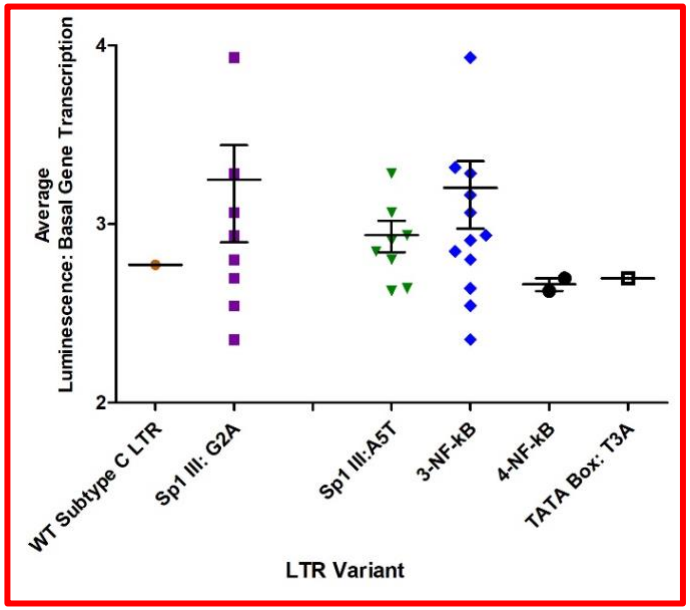


Figure 12. **Interpatient LTR genetic variability translates to differences in transcription activity.** Patient derived LTR-pGL3 basic vector recombinants were transfected into Jurkat cells and the luciferase activity was measured after a 24-hour incubation period. The transfection assay for each sample was performed in triplicates therefore the transfection data presented here are illustrative of the average relative light units (RLU). **(A)** Patient derived LTR-pGL3 basic vector recombinants were transfected into Jurkat cells alone to measure basal levels of gene transcription. **(B)** The patient derived LTR-pGL3 basic vector recombinants were co-transfected into Jurkat cells with a subtype C pcTat BL43.cc. **(C)** The patient derived LTR-pGL3 basic vector recombinants were co-transfected into Jurkat cells with the patients' autologous tat. However, the autologous tat for patient 802 was unavailable for this study therefore, patient 802 was excluded from this analysis.

D



E



F

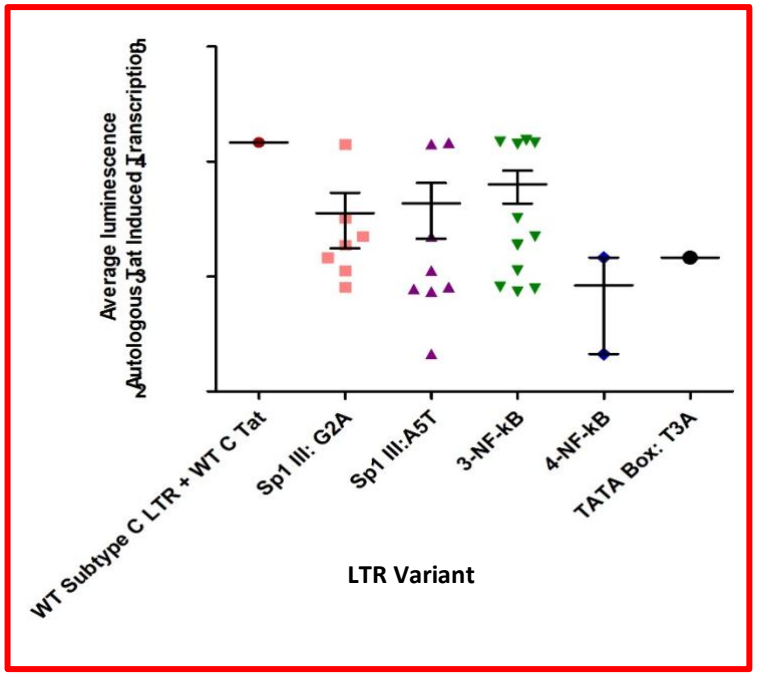


Figure 12 **Interpatient LTR genetic variability translates to differences in transcription activity Continued.** In addition to the RLU, the Fold change was also determined relative to the empty pGL3 vector. (D) Demonstrates the fold change for the LTR elements in the presence and absence of Tat. Although we examined the variation of the individual patients LTR and Tat activity on luciferase expression we then demonstrated the impact of the specific LTR variants (Sp1 III binding site, NF-kB copy number and TATA Box) on the average luciferase activity in both the absence of Tat (E) and presence of Tat.

Secondly, we compared the LTR elements gene transcription levels in the presence of a subtype C Tat obtained from NIH Reagents Programme (pcTatBL43.cc) and the respective patient autologous Tat protein. Irrespective of basal levels of gene transcription, the WT-C Tat protein was able to transactivate the majority of patient LTR elements to significantly higher levels of gene transcription (**Figure 12B**). However, 2 patients LTR (AS21-187 and AS03-369) did not show any significant level of transactivation and was comparable to the basal level of gene transcription. For patient AS21-187, this may be explained by the sequence variation observed in the TATA Box in combination with other mutations within the LTR. Patient AS21-187 exhibits the T3A mutation within the TATA Box as aforementioned. The alteration from the consensus sequence to the mutated sequence at position 3 has been reported to reduce the ability of Tat to bind and transactivate the LTR (Jeeninga et al., 2000, Montano et al., 1998).

Table 1. **Record of the mutations found within the autologous tat variation and their effect on tat function.**

AUTOLOGOUS TAT PID	TAT MUTATION
0016	L35Q/ Q39L (Enhancing)
945	P21A (Reducing) Q39L (Reducing) Q54R (Reducing)
703	P21A (Reducing)
358	P21A (Reducing) Q39L (Reducing) V4I (Reducing)
369	Q39L (Reducing)
1034	S46Y (Reducing)
268	No reported mutation
187	P21A (Reducing) Y29R (Enhancing)
458	No reported mutation
341	L35Q/ Q39L (Enhancing)
919	P21A (Reducing) Q39L (Reducing) V4I (Reducing)
876	P21A (Reducing) Y29R (Enhancing)

However, where it would be expected that the autologous Tat protein would significantly transactivate the LTR to higher levels of gene transcription, the autologous Tat induced levels of gene transcription was variable between patients (**Figure 12C**). This may be possibly explained by the interpatient Tat amino acid sequence variability. Also, it is well established that tat variability results in differential LTR transactivation activity. The genetic characterization of the autologous Tat amino acid changes was previously performed by our group (Mkhize et al 2018., unpublished). However, for the purposes of this study, we have highlighted the amino acid substitutions of each autologous tat protein (**Table 1).** The autologous Tat mutants characterised as having reducing mutations (Q39L, S46Y, P21A and V4I) did not significantly transactivate the LTR and demonstrated gene expression levels that were comparable to basal levels of gene transcription. This may be explained by the Tat protein mutations that may have a reducing effect on the ability of Tat to transactivate the LTR and hence activate transcription. In contrast, the autologous Tat proteins having enhancing mutations (L35Q/ Q39L and Y29R) significantly transactivated their respective LTRs to higher levels of gene transcription.

4.7. Variability within the LTR confers Differential Responsiveness to Extracellular Stimulation

We hypothesized that the interpatient LTR genetic varibaiity would confer differential LTR responsiveness to extracellular stimulation with TNF-α, PMA, SAHA and Prostratin. Therefore, the reponse of the LTR to extracellular stimulation was examined both in the absence of Tat (**Figure 13A**) as well as in the presence of Tat (**Figure 13B**).

The cell activation assay was performed using 12 LTR elements because the autologous Tat protein was unavailable for patients AS02-802 and AS01-876 and therefore these two samples were excluded from the cell activation assay and analysis. In the absence of Tat, all patient LTR elements responded equally well to stimulation with TNF-α and PMA which induced transcription at levels signficnalty higher than that of basal gene trasncription. Extracellular cell stimulation with PMA demonstrated the strongest response in all patient derived LTRs. However, patient isolates were least responsive to extracellular stimualtion with SAHA and Prostratin. Interestingly, 4 out of 12 patients (AS02-0016, AS02-945, AS02-1034 and AS21-187) did not respond to SAHA stimulation as they performed at a level equivalent to their respective basal level of transcription.

Importantly, one patient LTR element (AS21-187) was consistently not responding to stimulation with these cell activators except with PMA. Interestingly, we had highlighted that this specific patient (AS21-187) had exhbited the TATA Box T3A mutation in combination with other muations within the core modulatory and core enhancer domain. This highlights that the gentic varibaility within the U3 region alters the reponsiveness of the LTR to stimulation in the absence of Tat.

In the presence of the patients autologous Tat, the synergistic activation with cell activators resulted in signifcnatly higher levels of gene transcription than was observed without Tat transactivation. Importantly, the autologous Tat proteins characterised as having reducing muations (Table 1, data from a different study in the lab by a PhD student Zakithi Mkhize) demonstrated low levels of Tat induced transcription, however, when synergized with the cell activators, the levels of Tat induced transcription was signifcantly enhanced. Once again of particular note, one pateint (AS21-187) was not effciently resposnive to the synergistic cell stimulation and performed at levels below that of the WT. In all patient LTR elements, the synergistic activation with PMA yielded the strongest transcriptional repsonse.

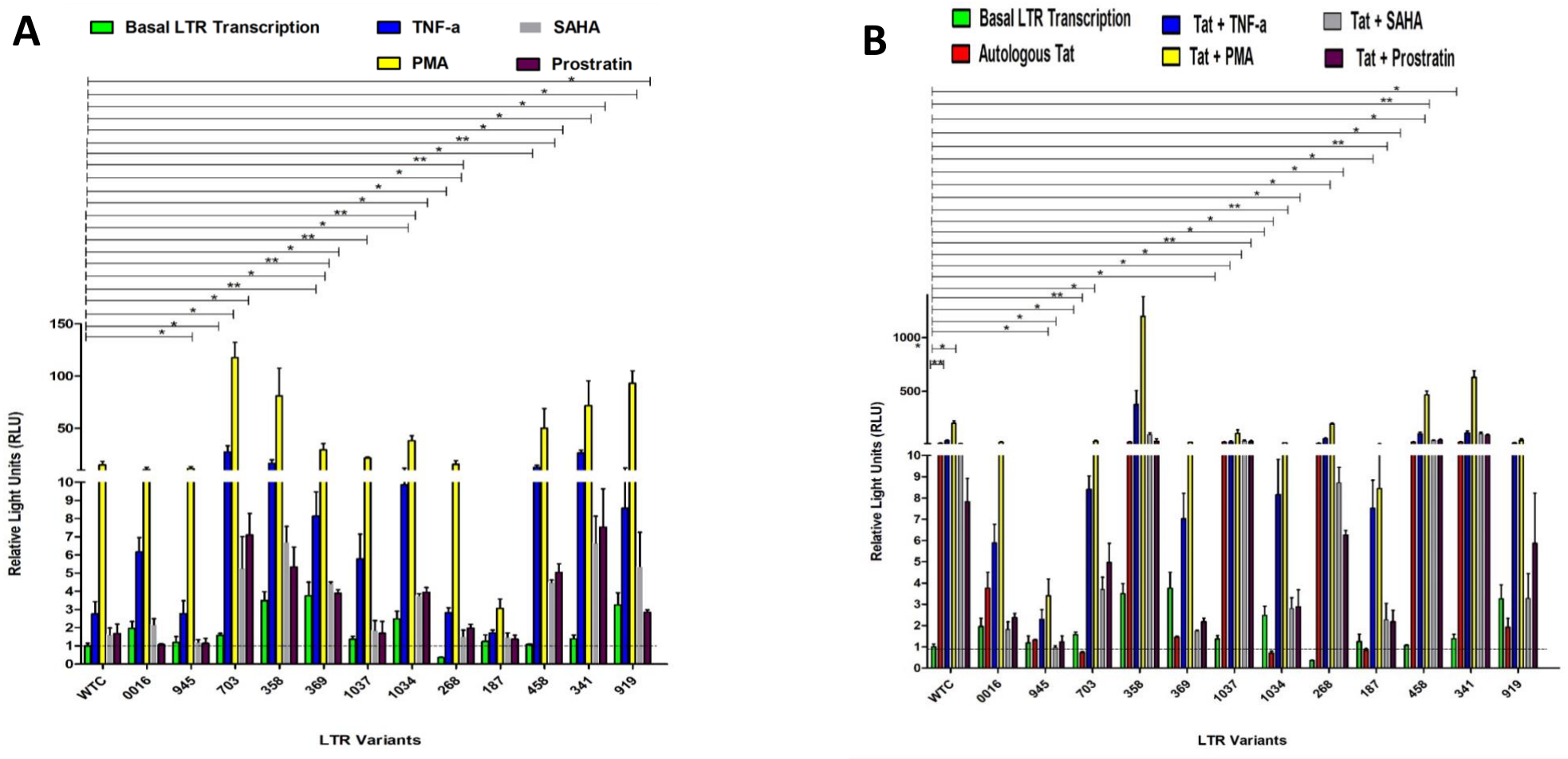


Figure 13**. Interpatient LTR genetic variability translates to differential transcription activity in the presence of cell stimulants.** Patient derived LTR-pGL3 basic vector recombinants were transfected into Jurkat cells followed by stimulation with TNF-α, PMA, SAHA and prostratin after 12 hours of incubation. The luciferase activity was measured after a 24-hour incubation period. The transfection assay for each sample was performed in triplicates therefore the transfection data presented here are illustrative of the average relative light units (RLU). **(A)** Patient derived LTR-pGL3 basic vector recombinants were transfected into Jurkat cells alone to measure basal levels of gene transcription in the presence of stimulation with TNF-α, PMA, SAHA and prostratin. **(B)** The patient derived LTR-pGL3 basic vector recombinants were co-transfected into Jurkat cells and co-stimulated with the patients' autologous tat and of the following cell stimulants: TNF-α, PMA, SAHA and Prostratin. However, the autologous tat for patient 802 was unavailable for this study therefore, patient 802 was excluded from this analysis.

4.8. The Impact of LTR Genetic Variation on Viral Load

The impact of LTR genetic variation on viral replication was then assessed by correlating the LTR mutants with viral load (**Figure 14**). At transmission (p= 0.2811), there were no differences observed between the mutation at nucleotide position 2 (G2A) of the Sp1 III binding site and viral load(**Figure 14A**). Further to that, no differences were observed at one-year post infection (p= 0.8351). Although, the Sp1 III A5T mutant exhibited a reduced mean viral load at both transmission (p= 0.3395) and one-year post infection (p= 0.2214), this was not significantly different (**Figure 14B**). The TATA Box variant was associated with a reduced viral load as expected due to the reduced transcription activity of the TAAAA variant (**Figure 14C**). However, our data showed that the 4$_{th}$-NF- κB variants, in combination with other mutations within the core enhancer and promoter domain, demonstrate a reduction in viral load (**Figure 14D**) Since these were low frequency mutations, the sample size used to analyse the TATA Box and 4$_{th}$ NF-κB variants are too small to be conclusive.

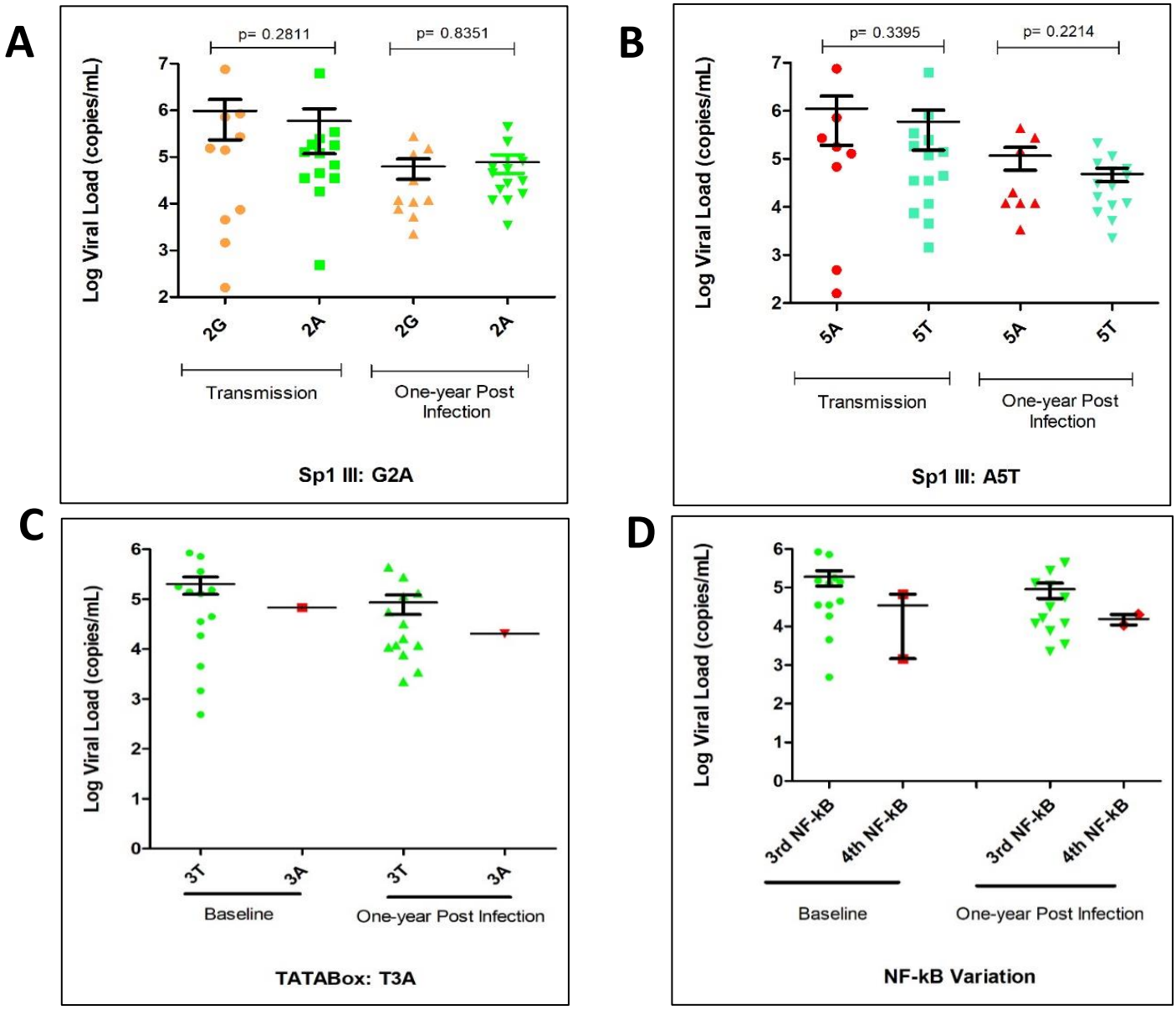
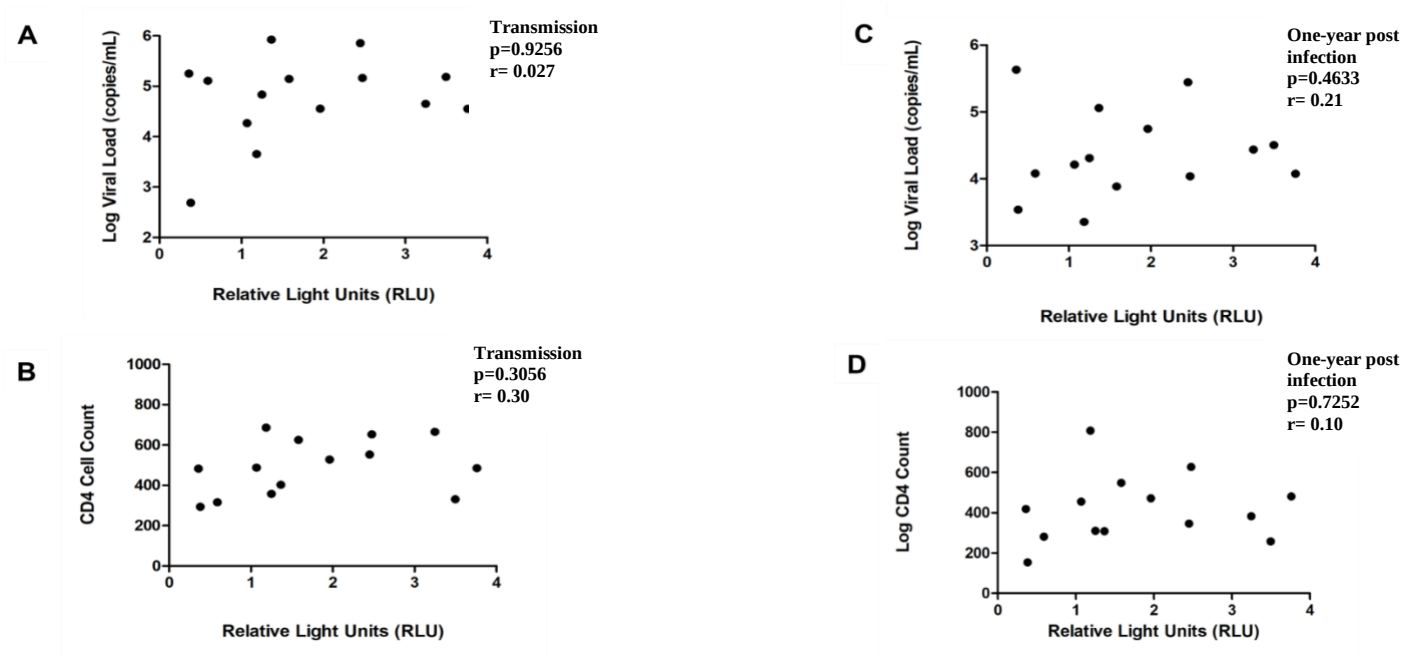


Figure 14. **The association of LTR variants with viral load at transmission and one year post infection.** Patient LTR elements were seperated into those patients who had the wild type nucleotide and those who exhibited the mutant nucleotide of the SP1 III binding site, TATAA box and those who had 3 NF-κB binding sites compared to those who had 4-NF-κB binding sites. Statical analyses was performed using Graphpad Prism 5.1 software for Windows, GraphPad Software, San Diego California USA, (www.graphpad.com). The paired T-test was used to determine signficance. The mean viral load was compared at near transmission and one year post infection. **(A)** A comparison of **t**he viral load of patients exhibiting the wild type G nucleotide and those exhbiting the mutant A nucleotide at position 2 of the Sp1 III binding site. The mean viral load of patients exhibiting the mutant nucleotide was lower than the mean viral load of patients having the wild type nucleotide. However at one year post infection there was no significant difference in viral load. **(B)** A comparison of the viral load of patients exhibiting the wild type A nucleotide and those exhbiting the mutant T nucleotide at position 5 of the Sp1 III binidng site. The mean viral load of patients having the mutant nucleotide was lower than the mean viral load of patients having however this was not signficant. (**C**) A comparison of the viral load of patients exhibiting the wild type T nucleotide and those exhbiting the mutant A nucleotide at position 3 of the TATAA box. (**D**) A comparison of the mean viral load of patients exhibiting the 3rd-NF-κB binding and those patients having the 4th-NF-κB binding site.

4.9. Correlation of LTR Transcriptional Activity with Clinical Markers

Basal LTR activity and autologous Tat induced gene expression was correlated with clinical markers such as viral load and CD4 count and statistical significance was assessed using a linear regression to determine the correlation coefficient (**Figure 15 Panel 1 and 2**). At transmission and one-year post infection, there was no significant correlation between basal LTR activity and viral load (transmission: p= 0.9256, r= 0.027, one-year post infection: p= 0.4633, r= 0.21) (**Figure 15 Panel 1: A and C**). Further to that, we did not identify a significant correlation between basal LTR activity and CD4 count at transmission (p= 0.3056, r= 0.30) and one-year post infection (p= 0.7252, r= 0.10) (**Figure 15 Panel 1: B and D)**. The autologous Tat induced LTR activity demonstrated a positive correlation with viral load at transmission (p= 0.0134, r= 0.66) however within one-year of infection this correlation was weakened and was not significant (p= 0.3905, r= 0.26) (**Figure 15 Panel 2: A and C**). Although, the autologous tat induced activity showed a negative correlation with CD4 count at transmission (p= 0.2188, r= 0.37) and one-year post infection (p= 0.1659, r= 0.41), this was also not significant (**Figure 15 Panel 2: B and D)**. The relationship between the basal and Tat induced LTR activity was weak and therefore did not show a strong prediction for the clinical markers.

Panel 1



Panel 2

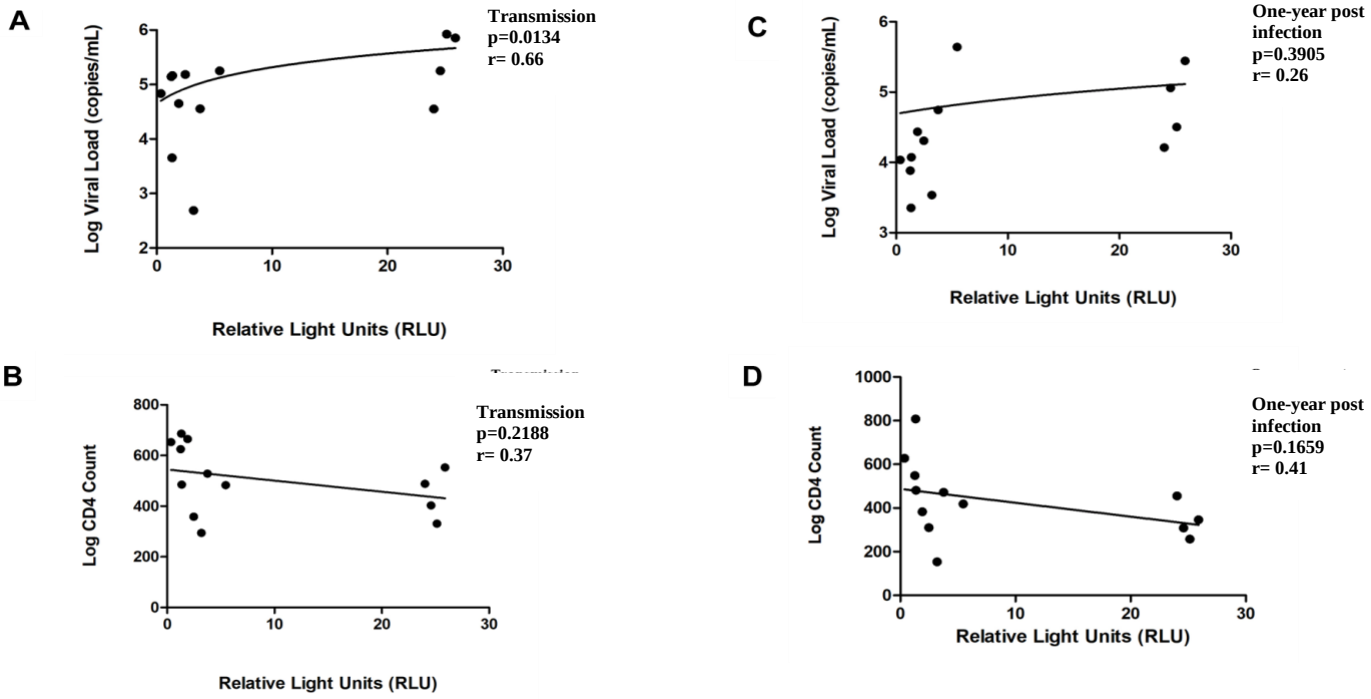


Figure 15. **No correlation of LTR activity with markers of disease progression**. Linear regression analyses were performed using Graphpad Prism 5.1 software for Windows, GraphPad Software, San Diego California USA, (www.graphpad.com)to determine a correlation between LTR basal activity measure in RLU and viral load as well as CD4 count at transmission and one-year post infection. **Panel A**: demonstrates the correlation between basal LTR activity and viral load and CD4 count at transmission (left) and one-year post infection (right). **Panel B**: demonstrates the correlation between tat induced LTR activity and viral load and CD4 count at transmission (left) and one-year post infection (right).

CHAPTER 5: DISCUSSION

It was hypothesized that the HIV-1C T/F viruses LTR genetic variation may have an impact on the viruses' transcriptional activity and may influence disease outcome in a cohort from South Africa. The data presented in this study demonstrates that the LTR genetic variation may impact the transcriptional activity of the virus however there was no distinct impact on disease progression.

A previous study reported inter-subtype LTR variation within the LTR element (Jeeninga et al., 2000) while the other studies reported intra-subtype, HIV-1 subtype C LTR variation (Boullosa et al., 2014, Jeeninga et al., 2000). However, this variation has been reported during chronic infection where the virus has already diversified due to virus own error-prone Reverse Transcriptase. Therefore, in this study we undertook to study HIV-1 subtype C genetic variation at or near (within 28 days of infection) transmission and diversity by one-year post infection.

Interestingly, our data demonstrate that sequences obtained from a total of 25 people living with acute HIV-1 infection were classified as HIV-1 subtype C virus (**Figure 9**). Sequences obtained at or near transmission from the same patient formed interpatient clusters with limited branch length differences, a stronger bootstrap support for all subjects and showed no evidence of relatedness of viruses from different patient participants (**Figure 8**). Therefore our data demonstrate that the T/F viruses LTR exhibits intra-subtype variation at transmission and limited evolution by one year post infection. In addition, Contrary to a study by Carlson et al., and colleagues (2014) who reported that the consensus like sequences for *gag*, *pol* and *env* were selected for at transmission, our data show that the T/F viruses LTR element does not exhibit a consensus like at transmission.

It is widely established that HIV-1 subtype C has expanded to contain a 3rd NF-κB binding sites compared to most subtypes, which have only 2 NF-κB binding sites and subtype AE that has one NF-κB binding site. Interestingly, the data from our study demonstrate that 100% of patient derived LTR sequences from the South African cohort contain 3 intact NF-κB binding sites. However 8% of pateints contained a 4th NF-κB binding site and 12% of patients contained the NF-κB like binding site (**Figure 9**). This observation is consistent with previous genetic characterization of the South African subtype C, which showed that 86% of patient derived LTR sequences exhibited the 3rd NF-κB binding site while 14% of patient isolates had a NF-κB like site (Scriba et al., 2002). These data demonstrate that there is no evidence suggesting that HIV-1 subtype C viruses that contain the 4th NF-κB binding site in South Africa are expanding.

Other studies from outside South Africa have also reported that HIV-1 subtype C viruses that contain the 4th NF-κB binding site as well as the NF-κB-like binding sites exists (Bachu et al., 2012, Boullosa et al., 2014, Hunt et al., 2001, Scriba et al., 2002). Contrary to the epidemic in South Africa, genetic characterization of the patient derived HIV-1 subtype C LTR in the other part of the world has depicted the rise of the 4th NF-κB and NF-κB like binding sites in India, Brazil, Mozambique and South Africa (Bachu et al., 2012, Boullosa et al., 2014, Scriba et al., 2002). Bachu et al., (2012) reported that the HIV-1 subtype C containing the 4th NF-κB was expanding in India thus becoming a dominant strain. These data suggest that the acquisition of the 4th additional NF-κB binding site confers a replicative advantage thus permitting its rapid expansion throughout India (Bachu et al., 2012). However, as aforementioned our data do not show evidence of HIV-1 subtype C viruses containing the 4th NF-κB binding site expanding in South Africa.

The occurrence of NF-κB like sites occurred in three out of twenty five (12%) patients, of which the NF-κB like binding site (GGGCGGTCC) found in patient AS903-036, referred to as 369 in this thesis was previously reported in a South African isolate TV008 (Scriba et al., 2002) (**Figure 9**). More recently, another study by Obasa et al 2019 showed that the canonical sequence of NF-κB had mutated to exhibit the sequence of the 4th NF-κB binding site in 6% of sequences within NF-κB II and in 1% of sequences within NF-κB, while the majority of the sequences exhibited the intact canonical sequence within NF-κB II (n=50; 72%) and NF-κB I (n=55; 87%) binding sites (Obasa et al., 2019). On the other hand, 49/63 (78%) exhibited the intact subtype C specific NF-κB (C- NF-κB) binding site (Obasa et al., 2019). Our study demonstrate the NF-κB I and NF-κB II binding sites from patient derived LTR sequences were conserved at transmission or near transmission and at one year post infection time points from all 25 PLHIV-1 (**Figure 9**).

The Sp1 binding sites are reported as conserved elements with the Sp1 III binding site demonstrating a greater degree of variability in comparison to the Sp1 I and Sp1 II binding site (Mbondji-Wonje et al., 2018, McAllister et al., 2000, Nonnemacher et al., 2004, Zhang et al., 1997). Consistent with this, our data demonstrated that Sp1 I and Sp1 II sequences were conserved while the Sp1 III sequence was the most variable (**Figure 9**). While Scriba et al., (2002) reported that the Sp1 II site was most conserved, our data show that the Sp1 I is more conserved than the Sp1 II binding site which is concordant with preceding studies (Mbondji-Wonje et al., 2018, McAllister et al., 2000). The Sp1 III functional capacity is dependent on the guanine residue at position 4 of the Sp1 III binding site. A single nucleotide modification at position 4 can hinder Sp1 binding and significantly weaken the tat-induced transactivation of the LTR (Berkhout and Jeang, 1992, Das et al., 2011, Kamine et al., 1991, McAllister

et al., 2000). The South African subtype C remains very well conserved at position 4 of the Sp1 binding site thereby possibly conserving optimal Sp1 factor interaction and binding.

The Sp1 III binding site exhibited variability specifically at position 2 and position 5 (**Figure 9**). One study that assessed the incidence of nucleotide changes within the subtype B Sp1 binding sites showed that the Sp1 III binding site had a 30% average non-consensus frequency at position 5 (McAllister et al., 2000). Further to that, this same study demonstrated that single nucleotide changes of the Sp1 III binding site can impair basal gene transcription (McAllister et al., 2000). The HXB2 Sp1 III sequence that is mutated at position 5 from a C-to-T nucleotide (GAGGCGTGGC → GAGG**T**GTGGC) was shown to reduce Sp1 binding efficiency (Nonnemacher et al., 2004). Furthermore, this mutation was shown to increase in frequency during disease progression and be associated with disease progression (Nonnemacher et al., 2004). Similarly, the subtype C Sp1 III presented in this study exhibits the 5T mutation (GGGGAGTGGT → GGGG**T**GTGGT) which is in agreement with previous studies (Mbondji-Wonje et al., 2018, McAllister et al., 2000, Shah et al., 2014).

This study did not find a significant difference between the mean viral loads of patients exhibiting the Sp1 III 2G nucleotide and those having the Sp1 III 2A nucleotide at one-year post infection (**Figure 14**). This may suggest that the Sp1 III G2A mutation may not be a predictor of disease outcome (viral load) during acute infection. The nucleotide substitution at position 5 of the Sp1 III binding site was previously associated with reduced activity (Nonnemacher et al., 2004). In our study, The Sp1 III: A5T identified within the Sp1 III binding site in this cohort showed decreased viral load at transmission and one-year post infection. This may be suggesting that the Sp1 III binding sites with nucleotide substitutions at position 5 may result in decreased transcription activity and hence reduced viral load. The effect of Sp1 factor binding has not been examined for subtype C viruses in South Africa. Thus, presenting a limitation to this study in that we did not examine the binding affinity of the Sp1 proteins for the mutated Sp1 III binding site. However, based on the previous findings by Nonnemacher et al., (2004), it is likely that the subtype C Sp1 III: A5T mutation may present with reduced binding affinity for the Sp1 proteins and hence reduced activity of the Sp1 III binding site.

Multiple studies have previously examined the functional activity of the LTR of subtypes A through G as well as the functional activity of viral variants differing in NF-κB copy number (Bachu et al., 2012, Jeeninga et al., 2000, Montano et al., 1998). In this study we specifically examined the functional activity of 14 patient LTR elements using a transfection assay in vitro. Even in the presence of genetic variation, all patient LTR elements were able to drive basal levels of gene transcription, although at

varying levels (**Figure 12A**). Notably, patients AS02-0802, AS03-0268 **a**nd AS02-876 demonstrated a reduced basal gene transcription that was below that of the consensus, referred to as wild-type sequence. Interestingly, these two patients both shared a Sp1 III- G2A mutation. While those patients having either a Sp1 III- A5T mutant or a co-mutation of Sp1 III- A5T and G2A performed equally well as the wild type or 3-4-fold higher than the wild type such as in patient isolates AS02-358, AS03-369, AS02-1034 and AS01-919. This data may indicate that having a Sp1 G2A mutation only may impair basal gene transcription thereby reducing the transcription and replicative fitness of the virus. Therefore, the fold differences in basal gene transcription may be significant in understanding the viral fitness of each patient.

In vivo, viral transcription is dependent on the ability of the viral protein, Tat to transactivate the LTR transcription ability or activity in order to fully transcribe the provirus genome. Therefore, we assessed the ability of the LTR to be transactivated by the HIV-1 subtype C consensus Tat *in vitro*. The subtype C pcTat that contains consensus Tat sequence was used. Our data show that irrespective of the level of basal gene transcription, the subtype C consensus Tat was able to transactivate all patient derived LTRs to significantly increase the level of gene transcription. Our data demonstrate that although the Sp1 G2A mutant may have impaired basal LTR activity, it did not hinder the ability of Tat to transactivate the LTR as compared to the previously described Sp1 III mutation at position 4 (Miller-Jenson et al., 2013).

Patient AS21-187 did not show any significant transactivation (with both the WT subtype C Tat and autologous Tat) compared to the wild type subtype C (**Figure 12A and B**) and this may be as a result of the T3A mutation within the TATA Box (TATAA → TAAAA) along with other mutations within the patients LTR. The TATA box is located 28 base pairs upstream from the transcription start site in the U3 region (Berkhout and Jeang, 1992). The Tata binding protein (TBP) together with several transcription factors form a TFIID complex binds to the TATAA sequence to initiate transcription (Van Opijnen et al., 2004). The change from a T nucleotide at position 3 to an A nucleotide significantly impairs the interaction between the TATA box sequence and TBP-TFIIB complex consequentially impairing the recruitment of RNA polymerase II (Van Opijnen et al., 2004). Therefore, there is complete requirement of a T nucleotide at the 3rd position of the TATA element. The mutated TATA box (TATAA → TAAAA) sequence impairs the ability of the LTR to initiate transcription. This therefore demonstrates that LTR genetic variation impacts LTR transactivation and HIV-1 pathogenesis.

An indication of the ability of the LTR to drive viral gene transcription *in vivo* may be imperative in determining the rate of disease progression as well as the shift between latent and productive infection. The assessment of the tat-induced gene transcription using the patient's autologous Tat protein permitted the evaluation of LTR activity in combination with tat protein variability (**Figure 12C**). Single amino acids alterations within the Tat protein can impair its ability to either recruit P-TEFb or phosphorylate RNA polymerase II resulting in impaired transcription capacity (Dey et al., 2012). However, the presence of single amino acid changes can also enhance Tat recruitment of the TBP and P-TEFb. the autologous *tat* transfection data presented here demonstrate that *tat* genetic variation can also significantly impair LTR transactivation in the case of the P21A, Q39L and V4I *tat* variants. The combination of LTR and Tat genetic variation results in differential functional activity of the LTR. Patients that have suppressed LTR transactivation activity may be more likely to shift from a productive infection to a latent infection whereas those patients that have significantly enhanced transactivation activities are more likely to transition to a productive infection rather than a latent state.

Previously, it was demonstrated that the 4-NF-κB variants have higher gene reporter expression in comparison to a 3-NF-κB LTR (Bachu et al., 2012). In this study, a comparison of the transcription capacity of variants containing 3-NF-κB binding sites versus the 4_{th} NF-κB variants showed that the 3 NF- κB variants demonstrated higher transcriptional activity (**Figure 12 A**). It is noted that three 3-NF-κB variants exhibited an approximate 1-2-fold higher transcription capacity than the 4-NF-κB variant (patient AS02-1034 and AS21-187) during basal and subtype C consensus Tat induced transcription. The inconsistency of the 4_{th} NF-κB variants to achieve higher gene reporter expression above all patient LTRs may be a possible consequence of the variability in other regions of the LTR such as the Sp1 III A5T mutation and TATA Box T3A mutation.

Furthermore, our data demonstrate that irrespective of low basal gene transcription activity, extracellular activation with TNF-α and PMA stimulates the LTR (in the absence of Tat) to significantly higher levels of gene transcription (**Figure 13A**). The patient LTR elements were more strongly activated by PMA as compared to TNF-α, SAHA and prostratin. Interestingly, regardless of the level of Tat-responsiveness, all patient isolates demonstrated a stronger response to extracellular stimulation with PMA than with TNF-α (**Figure 13B**). The Tat (autologous) induced transcription activity was significantly enhanced in the presence of PMA and TNF-α however, all isolates demonstrated stronger responsiveness to PMA than TNF-α. Patient isolates with a high level of Tat-responsiveness (358, 1037, 268, 458 and 341) demonstrated an equal responsiveness to extracellular stimulation with SAHA and Prostratin.

Interestingly, patient isolates that were less responsive to the autologous Tat (1034, 187 and 919) were more responsive to stimulation with SAHA and Prostratin (**Figure 13B**). This therefore may demonstrate that there is variation among the patient LTRs in which isolates that are least responsive to the autologous Tat may respond better to stimulation with SAHA and Prostratin than patient isolates that have higher levels of Tat responsiveness. A block in activity of the LTR and autologous Tat combined was overcome by extracellular stimulation and responded strongly to activation with PMA. Interestingly, the interpatient LTR variability resulted in variability in the transcriptional activity of the various LTR elements under stimulation within TNF-α and PMA. This variability between patients may suggest that each patients LTR is subject to varying levels of activation and the activation potential of each LTR element differs between patients.

Our data suggest that targeting the LTR in a therapeutic setting of latency reversal, would require that the genetic variability of each patients LTR be taken into account since this may have an impact on the activation potential of the LTR. The majority of the patients LTR (96%) responded equally well to extracellular stimulation except for patient AS21-187 that did not show a significant increase in both basal and autologous Tat induced activity. This may propose that all patient LTR elements may be activated via the same mechanism when stimulated with PMA but the interpatient LTR variability may be the source for the observed differences in the levels of responsiveness. Patient AS21-187 exhibited a 1-2-fold transcription activity following stimulation with TNF-α and PMA compared to the majority of patient LTRs that were successfully activated to approximately 50-100-fold. This further emphasizes that every patients LTR may respond differently to extracellular stimulation and LTR variability may impact the activation potential of the LTR.

The LTR basal transcription did not correlate with viral load and CD4 count at or near transmission and one-year post infection. However, the autologous Tat induced transcription strongly correlated positively with plasma viral load at transmission. In addition, autologous Tat induced transcription did not corelate with a decline in CD4 cell count at or near transmission and one-year post infection. Taken together our data demonstrate that during acute infection the autologous Tat induced LTR transcription contributes to an increase in plasma viral load. Therefore, during acute infection, the autologous Tat induced LTR activity may be a predictor for disease outcome (viral load).

In conclusion, our data demonstrates that the core enhancer and core promoter region of the LTR is relatively well conserved in relation to the consensus (wild type) LTR, nonetheless there was notable variation within the RBE III site, Sp1 III site and TATA Box. Furthermore, specific LTR variants may

be associated with greater capacity of the LTR to drive viral gene expression while other variants may be associated with a reduced capacity for driving viral gene expression. Interestingly, the TATA Box T3A mutation in combination with other mutations in the LTR demonstrated that LTR variability within the core promoter may significantly impair Tat induced LTR activity. However, further studies involving site directed mutagenesis may be required to assess the impact of the RBE III (C6A), Sp1 III (G2A and A5T) and TATA box mutants (T3A) alone and in combinations with other muations wihtin the South African subtype C LTR. We further show that the 3rd NF-κB binding site variants dominate the 4th NF- κB binding site variants in this South African cohort therfore, our data suggest that there is no evidence for the expansion of the 4th NF-κB variants in South Africa.

The differential interpatient basal gene expression may plausably be as a result of the variation within the RBE III site, Sp1 III site and TATA box acting in combination with other mutations present within the LTR. In additon our data demonstrated that the transcription acitivity of the LTR variants exhibit differential responsvieness to extracellular stimulation with TNF-α, PMA, SAHA and prostratin. However, the majority of LTR variants were shown to be most responsive to stimulation with PMA. Furthermore, the reuslts presented here further showed that the TATA Box T3A (TAAAA) variant, in combination with other mutaions that resulted in reduced LTR activity, was marginally induced when stimulated with PMA.

The interpatient LTR variability may have an impact on viral transcription and may impact viral fitness. Taken together, our data suggests that variation within the LTR mayimpact the transcriptional activity of the T/F virus. The data from this study may contribute to understanding the role of the LTR in establishing latency during acute infection.

CHAPTER 6: REFERENCES

ALTMAN, L. K. 1981. Rare cancer seen in 41 homosexuals. *New York Times,* 3**,** 1981.

ARCHIN, N. M., BATESON, R., TRIPATHY, M. K., CROOKS, A. M., YANG, K.-H., DAHL, N. P., KEARNEY, M. F., ANDERSON, E. M., COFFIN, J. M. & STRAIN, M. C. 2014. HIV-1 expression within resting CD4+ T cells after multiple doses of vorinostat. *The Journal of infectious diseases,* 210**,** 728-735.

ARCHIN, N. M., ESPESETH, A., PARKER, D., CHEEMA, M., HAZUDA, D. & MARGOLIS, D. M. 2009. Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. *AIDS research and human retroviruses,* 25**,** 207-212.

ARHEL, N. J., SOUQUERE-BESSE, S., MUNIER, S., SOUQUE, P., GUADAGNINI, S., RUTHERFORD, S., PRÉVOST, M. C., ALLEN, T. D. & CHARNEAU, P. 2007. HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *The EMBO journal,* 26**,** 3025-3037.

ASIBEY, E. O. Wildlife as a source of protein in Africa south of the Sahara. 1974. *Biological Conservation*, 6, 32-39.

AYOUBA, A., SOUQUIÈRES, S., NJINKU, B., MARTIN, P. M., MÜLLER-TRUTWIN, M. C., ROQUES, P., BARRÉ-SINOUSSI, F., MAUCLÈRE, P., SIMON, F. & NERRIENET, E. 2000. HIV-1 group N among HIV-1-seropositive individuals in Cameroon. *Aids,* 14**,** 2623-2625.

BAAR, M. P. D., RONDE, A. D., BERKHOUT, B., CORNELISSEN, M., HORN, K. H. V. D., SCHOOT, A. M. V. D., WOLF, F. D., LUKASHOV, V. V. & GOUDSMIT, J. 2000. Subtype-specific sequence variation of the HIV type 1 long terminal repeat and primer-binding site. *AIDS research and human retroviruses,* 16**,** 499-504.

BACHU, M., YALLA, S., ASOKAN, M., VERMA, A., NEOGI, U., SHARMA, S., MURALI, R. V., MUKTHEY, A. B., BHATT, R. & CHATTERJEE, S. 2012. Multiple NF-κB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. *Journal of Biological Chemistry,* 287**,** 44714-44735.

BAEUERLE, P. A. & BALTIMORE, D. 1988. I kappa B: a specific inhibitor of the NF-kappa B transcription factor. *Science,* 242**,** 540-546.

BAEUERLE, P. A. & BALTIMORE, D. 1989. A 65-kappaD subunit of active NF-kappaB is required for inhibition of NF-kappaB by I kappaB. *Genes & development,* 3**,** 1689-1698.

BAILES, E., GAO, F., BIBOLLET-RUCHE, F., COURGNAUD, V., PEETERS, M., MARX, P. A., HAHN,B. H. & SHARP, P. M. Hybrid origin of SIV in Chimpanzees. 2003. *Science,* 300, 5626, 1713.

BALLARD, D. W., DIXON, E. P., PEFFER, N. J., BOGERD, H., DOERRE, S., STEIN, B. & GREENE, W. C. 1992. The 65-kDa subunit of human NF-kappa B functions as a potent transcriptional activator and a target for v-Rel-mediated repression. *Proceedings of the National Academy of Sciences,* 89**,** 1875-1879.

BARRÉ-SINOUSSI, F., CHERMANN, J.-C., REY, F., NUGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C., VÉZINET-BRUN, F. & ROUZIOUX, C. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science,* 220**,** 868-871.

BBOSA, N., KALEEBU, P. & SSEMWANGA, D. 2019. HIV subtype diversity worldwide. *Current Opinion in HIV and AIDS,* 14**,** 153-160.

BEDNARIK, D., DUCKETT, C., KIM, S., PEREZ, V., GRIFFIS, K., GUENTHNER, P. & FOLKS, T. 1991. DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs. *The New biologist,* 3**,** 969-976.

BERKHOUT, B. & JEANG, K. 1992. Functional roles for the TATA promoter and enhancers in basal and Tat-induced expression of the human immunodeficiency virus type 1 long terminal repeat. *Journal of virology,* 66**,** 139-149.

BLAZKOVA, J., TREJBALOVA, K., GONDOIS-REY, F., HALFON, P., PHILIBERT, P., GUIGUEN, A., VERDIN, E., OLIVE, D., VAN LINT, C. & HEJNAR, J. 2009. CpG methylation controls reactivation of HIV from latency. *PLoS pathogens,* 5.

BOULLOSA, J., BACHU, M., BILA, D., RANGA, U., SÜFFERT, T., SASAZAWA, T. & TANURI, A. 2014. Genetic diversity in HIV-1 subtype C LTR from Brazil and Mozambique generates new transcription factor-binding sites. *Viruses,* 6**,** 2495-2504.

BRAIBANT, M. & BARIN, F. 2013. The role of neutralizing antibodies in prevention of HIV-1 infection: what can we learn from the mother-to-child transmission context? *Retrovirology,* 10**,** 103.

BROWN, P. O., BOWERMAN, B., VARMUS, H. E. & BISHOP, J. M. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proceedings of the National Academy of Sciences,* 86**,** 2525-2529.

BRUNER, K. M., MURRAY, A. J., POLLACK, R. A., SOLIMAN, M. G., LASKEY, S. B., CAPOFERRI, A. A., LAI, J., STRAIN, M. C., LADA, S. M. & HOH, R. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine,* 22**,** 1043.

BUKRINSKAYA, A., BRICHACEK, B., MANN, A. & STEVENSON, M. 1998. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *The Journal of experimental medicine,* 188**,** 2113-2125.

BUKRINSKY, M. I., SHAROVA, N., DEMPSEY, M. P., STANWICK, T. L., BUKRINSKAYA, A. G., HAGGERTY, S. & STEVENSON, M. 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci U S A,* 89.

BUONAGURO, L., TORNESELLO, M. & BUONAGURO, F. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *Journal of virology,* 81**,** 10209-10219.

CAIN, C. E., BLEKHMAN, R., MARIONI, J. C. & GILAD, Y. 2011. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics,* 187**,** 1225-1234.

CARLSON, J. M., SCHAEFER, M., MONACO, D. C., BATORSKY, R., CLAIBORNE, D. T., PRINCE, J., DEYMIER, M. J., ENDE, Z. S., KLATT, N. R. & DEZIEL, C. E. 2014. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science,* 345**,** 1254031.

CASE, K. 1986. Nomenclature: human immunodeficiency virus. *Annals of internal medicine,* 105**,** 133-133.

CASTLEY, A., SAWLESHWARKAR, S., VARMA, R., HERRING, B., THAPA, K., DWYER, D., CHIBO, D., NGUYEN, N., HAWKE, K. & RATCLIFF, R. 2017. A national study of the molecular epidemiology of HIV-1 in Australia 2005–2012. *PloS one,* 12**,** e0170601.

CDC 1981a. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR. Morbidity and mortality weekly report,* 30**,** 305.

CDC 1981b. Pneumocystis Pneumonia --- Los Angeles.

CDC 1982a. A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. *MMWR. Morbidity and mortality weekly report,* 31**,** 305.

CDC 1982b. Pneumocystis carinii pneumonia among persons with hemophilia A. *MMWR. Morbidity and mortality weekly report,* 31**,** 365.

CDC 1982c. Update on acquired immune deficiency syndrome (AIDS)--United States. *MMWR. Morbidity and mortality weekly report,* 31**,** 507.

CDC, C. F. D. C. 1982d. Opportunistic infections and Kaposi's sarcoma among Haitians in the United States. *MMWR. Morbidity and mortality weekly report,* 31**,** 353.

CHAN, D. C., FASS, D., BERGER, J. M. & KIM, P. S. 1997. Core structure of gp41 from the HIV envelope glycoprotein. *Cell,* 89**,** 263-273.

CHEREPANOV, P., MAERTENS, G., PROOST, P., DEVREESE, B., VAN BEEUMEN, J., ENGELBORGHS, Y., DE CLERCQ, E. & DEBYSER, Z. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *Journal of Biological Chemistry,* 278**,** 372-381.

CHOE, H., FARZAN, M., SUN, Y., SULLIVAN, N., ROLLINS, B., PONATH, P. D., WU, L., MACKAY, C. R., LAROSA, G., NEWMAN, W., GERARD, N., GERARD, C. & SODROSKI, J. 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell,* 85.

CHRIST, F., THYS, W., DE RIJCK, J., GIJSBERS, R., ALBANESE, A., AROSIO, D., EMILIANI, S., RAIN, J.-C., BENAROUS, R. & CERESETO, A. 2008. Transportin-SR2 imports HIV into the nucleus. *Current Biology,* 18**,** 1192-1202.

CHUN, T.-W., ENGEL, D., BERREY, M. M., SHEA, T., COREY, L. & FAUCI, A. S. 1998. Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection. *Proceedings of the National Academy of Sciences,* 95**,** 8869-8873.

CHUN, T.-W., JUSTEMENT, J. S., MURRAY, D., HALLAHAN, C. W., MAENZA, J., COLLIER, A. C., SHETH, P. M., KAUL, R., OSTROWSKI, M. & MOIR, S. 2010. Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. *AIDS (London, England),* 24**,** 2803.

CHUN, T.-W., STUYVER, L., MIZELL, S. B., EHLER, L. A., MICAN, J. A. M., BASELER, M., LLOYD, A. L., NOWAK, M. A. & FAUCI, A. S. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences,* 94**,** 13193-13197.

CLAVEL, F., GUETARD, D., BRUN-VEZINET, F., CHAMARET, S., REY, M.-A., SANTOS-FERREIRA, M. O., LAURENT, A. G., DAUGUET, C., KATLAMA, C. & ROUZIOUX, C. 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science,* 233**,** 343-346.

COFFIN, J. 1992. Genetic diversity and evolution of retroviruses. *Genetic Diversity of RNA Viruses.* Springer.

COFFIN, J. M., HUGHES, S. H. & VARMUS, H. E. 1997. *Retroviral Virions and Genomes--Retroviruses*, Cold Spring Harbor Laboratory Press.

COHEN, J., POWDERLY, W. G. & OPAL, S. M. 2016. *Infectious Diseases E-Book*, Elsevier Health Sciences.

COHEN, M. S., GAY, C. L., BUSCH, M. P. & HECHT, F. M. 2010. The detection of acute HIV infection. *The Journal of infectious diseases,* 202**,** S270-S277.

COLLABORATION, H.-C. 2010. The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *AIDS (London, England),* 24**,** 123.

COULL, J. J., ROMERIO, F., SUN, J.-M., VOLKER, J. L., GALVIN, K. M., DAVIE, J. R., SHI, Y., HANSEN, U. & MARGOLIS, D. M. 2000. The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *Journal of virology,* 74**,** 6790-6799.

CRAMPTON, N., BONASS, W. A., KIRKHAM, J., RIVETTI, C. & THOMSON, N. H. 2006. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic acids research,* 34**,** 5416-5425.

CRON, R. Q., BARTZ, S. R., CLAUSELL, A., BORT, S. J., KLEBANOFF, S. J. & LEWIS, D. B. 2000. NFAT1 enhances HIV-1 gene expression in primary human CD4 T cells. *Clinical immunology,* 94**,** 179-191.

CULLEN, B. R. 2000. Nuclear RNA export pathways. *Molecular and cellular biology,* 20**,** 4181-4187.

CULLEN, B. R. 2003. Nuclear mRNA export: insights from virology. *Trends in biochemical sciences,* 28**,** 419-424.

DAELEMANS, D., VANDAMME, A.-M. & DE CLERCQ, E. 1999. Human immunodeficiency virus gene regulation as a target for antiviral chemotherapy. *Antiviral Chemistry and Chemotherapy,* 10**,** 1-14.

DALGLEISH, A. G., BEVERLEY, P. C., CLAPHAM, P. R., CRAWFORD, D. H., GREAVES, M. F. & WEISS, R. A. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature,* 312**,** 763-767.

DAS, A. T., HARWIG, A. & BERKHOUT, B. 2011. The HIV-1 Tat protein has a versatile role in activating viral transcription. *Journal of virology,* 85**,** 9506-9516.

DAVEY, R. T., BHAT, N., YODER, C., CHUN, T.-W., METCALF, J. A., DEWAR, R., NATARAJAN, V., LEMPICKI, R. A., ADELSBERGER, J. W. & MILLER, K. D. 1999. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proceedings of the National Academy of Sciences,* 96**,** 15109-15114.

DE COCK, K. M., ADJORLOLO, G., EKPINI, E., SIBAILLY, T., KOUADIO, J., MARAN, M., BRATTEGAARD, K., VETTER, K. M., DOORLY, R. & GAYLE, H. D. 1993. Epidemiology and transmission of HIV-2: why there is no HIV-2 pandemic. *Jama,* 270**,** 2083-2086.

DE HOUWER, S., DEMEULEMEESTER, J., THYS, W., TALTYNOV, O., ZMAJKOVICOVA, K., CHRIST, F. & DEBYSER, Z. 2012. Identification of residues in the C-terminal domain of HIV-1 integrase that mediate binding to the transportin-SR2 protein. *Journal of Biological Chemistry,* 287**,** 34059-34068.

DE LEYS, R., VANDERBORGHT, B., HAESEVELDE, M. V., HEYNDRICKX, L., VAN GEEL, A., WAUTERS, C., BERNAERTS, R., SAMAN, E., NIJS, P. & WILLEMS, B. 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *Journal of virology,* 64**,** 1207-1216.

DELVIKS-FRANKENBERRY, K., GALLI, A., NIKOLAITCHIK, O., MENS, H., PATHAK, V. K. & HU, W.-S. 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses,* 3**,** 1650-1680.

DEY, S. S., XUE, Y., JOACHIMIAK, M. P., FRIEDLAND, G. D., BURNETT, J. C., ZHOU, Q., ARKIN, A. P. & SCHAFFER, D. V. 2012. Mutual information analysis reveals coevolving residues in Tat that compensate for two distinct functions in HIV-1 gene expression. *Journal of Biological Chemistry,* 287**,** 7945-7955.

DEYMIER, M. J., ENDE, Z., FENTON-MAY, A. E., DILERNIA, D. A., KILEMBE, W., ALLEN, S. A., BORROW, P. & HUNTER, E. 2015. Heterosexual transmission of subtype C HIV-1 selects consensus-like variants without increased replicative capacity or interferon-α resistance. *PLoS pathogens,* 11**,** e1005154.

DI FAGAGNA, F. D. A., MARZIO, G., GUTIERREZ, M. I., KANG, L. Y., FALASCHI, A. & GIACCA, M. 1995. Molecular and functional interactions of transcription factor USF with the long terminal repeat of human immunodeficiency virus type 1. *Journal of virology,* 69**,** 2765-2775.

DONG, K. L., MOODLEY, A., KWON, D. S., GHEBREMICHAEL, M. S., DONG, M., ISMAIL, N., NDHLOVU, Z. M., MABUKA, J. M., MUEMA, D. M. & PRETORIUS, K. 2018. Detection and treatment of Fiebig stage I HIV-1 infection in young at-risk women in South Africa: a prospective cohort study. *The Lancet HIV,* 5**,** e35-e44.

DRAGIC, T., LITWIN, V., ALLAWAY, G. P., MARTIN, S. R., HUANG, Y., NAGASHIMA, K. A., CAYANAN, C., MADDON, P. J., KOUP, R. A. & MOORE, J. P. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature,* 381**,** 667-673.

DU CHÉNÉ, I., BASYUK, E., LIN, Y. L., TRIBOULET, R., KNEZEVICH, A., CHABLE-BESSIA, C., METTLING, C., BAILLAT, V., REYNES, J. & CORBEAU, P. 2007. Suv39H1 and HP1γ are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. *The EMBO journal,* 26**,** 424-435.

ELIUK, A. S., RUITER, B. K. & BOULANGER, C. P. Classifying HIV-1 Circulating Recombinant Forms. Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP), 2011. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 1.

ELLEGÅRD, R. 2018. *Effects of Complement Opsonization of HIV on Dendritic Cells: and Implications for the Immune Response*, Linköping University Electronic Press.

ENGELMAN, A., MIZUUCHI, K. & CRAIGIE, R. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell,* 67**,** 1211-1221.

FARIA, N. R., RAMBAUT, A., SUCHARD, M. A., BAELE, G., BEDFORD, T., WARD, M. J., TATEM, A. J., SOUSA, J. D., ARINAMINPATHY, N. & PÉPIN, J. 2014. The early spread and epidemic ignition of HIV-1 in human populations. *science,* 346**,** 56-61.

FASSATI, A. & GOFF, S. P. 2001. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *Journal of virology,* 75**,** 3626-3635.

FIEBIG, E. W., WRIGHT, D. J., RAWAL, B. D., GARRETT, P. E., SCHUMACHER, R. T., PEDDADA, L., HELDEBRANT, C., SMITH, R., CONRAD, A. & KLEINMAN, S. H. 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *Aids,* 17**,** 1871-1879.

FIELDS, B., KNIPE, D. & HOWLEY, P. 2007. Fields virology. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

FIELDS, B. N., KNIPE, D. M., HOWLEY, P. M., EVERISS, K. D. & KUNG, H.-J. 1996. *Fundamental virology*, Lippincott-Raven Philadelphia^ ePA PA.

FINZI, D., BLANKSON, J., SILICIANO, J. D., MARGOLICK, J. B., CHADWICK, K., PIERSON, T., SMITH, K., LISZIEWICZ, J., LORI, F. & FLEXNER, C. 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature medicine,* 5**,** 512-517.

FINZI, D., HERMANKOVA, M., PIERSON, T., CARRUTH, L. M., BUCK, C., CHAISSON, R. E., QUINN, T. C., CHADWICK, K., MARGOLICK, J. & BROOKMEYER, R. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science,* 278**,** 1295-1300.

FLETCHER, T. M., BRICHACEK, B., SHAROVA, N., NEWMAN, M. A., STIVAHTIS, G., SHARP, P. M., EMERMAN, M., HAHN, B. H. & STEVENSON, M. 1996. Nuclear import and cell cycle arrest functions of the HIV-1 Vpr protein are encoded by two separate genes in HIV-2/SIV (SM). *The EMBO Journal,* 15**,** 6155-6165.

FRIEDMAN, H., SPECTER, S. & BENDINELLI, M. 2007. *In vivo models of HIV disease and control*, Springer Science & Business Media.

FRIEDMAN, J., CHO, W.-K., CHU, C. K., KEEDY, K. S., ARCHIN, N. M., MARGOLIS, D. M. & KARN, J. 2011. Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. *Journal of virology,* 85**,** 9078-9089.

FUJIWARA, T. & MIZUUCHI, K. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell,* 54**,** 497-504.

GALLAHER, W. R. 1987. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell,* 50**,** 327.

GALLO, R., WONG-STAAL, F., MONTAGNIER, L., HASELTINE, W. A. & YOSHIDA, M. 1988. HIV/HTLV gene nomenclature. *Nature,* 333**,** 504-504.

GALLO, S. A., FINNEGAN, C. M., VIARD, M., RAVIV, Y., DIMITROV, A., RAWAT, S. S., PURI, A., DURELL, S. & BLUMENTHAL, R. 2003. The HIV Env-mediated fusion reaction. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1614**,** 36-50.

GANCHI, P., SUN, S.-C., GREENE, W. & BALLARD, D. 1992. I kappa B/MAD-3 masks the nuclear localization signal of NF-kappa B p65 and requires the transactivation domain to inhibit NF-kappa B p65 DNA binding. *Molecular biology of the cell,* 3**,** 1339-1352.

GAYNOR, R. 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. *Aids,* 6**,** 347-364.

GEIST, V. How markets for wildlife meat and parts, and the sale of hunting privileges, jeopardize wildlife conservation. 1988. *Conservation Biology*, 2, 15-26.

GEORGESS-COURBOT, M. C.,LU, C. Y., MAKUWA, M., TELFER, P., ONANGA, R., DUBREIL, G. CHEN, Z., SMITH, S.M., GEORGES, A., GAO, F, & HAHN, B. H. 1998. Natural infection of a household pet red-capped mangabey (Cercocebus torqautus torquatus) with a new simian immunodefciency virus. *Journal of Virology*, 72 (1), 600-608.

GHOSH, S., GIFFORD, A. M., RIVIERE, L. R., TEMPST, P., NOLAN, G. P. & BALTIMORE, D. 1990. Cloning of the p50 DNA binding subunit of NF-κB: homology to rel and dorsal. *Cell,* 62**,** 1019-1029.

GOUNDER, K., PADAYACHI, N., MANN, J. K., RADEBE, M., MOKGORO, M., VAN DER STOK, M., MKHIZE, L., MNCUBE, Z., JAGGERNATH, M. & REDDY, T. 2015a. High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. *PloS one,* 10**,** e0119886.

GOUNDER, K., PADAYACHI, N., MANN, J. K., RADEBE, M., MOKGORO, M., VAN DER STOK, M., MKHIZE, L., MNCUBE, Z., JAGGERNATH, M. & REDDY, T. 2015b. High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. *PloS one,* 10.

GRAY, L. R., COWLEY, D., CRESPAN, E., WELSH, C., MACKENZIE, C., WESSELINGH, S. L., GORRY, P. R. & CHURCHILL, M. J. 2013. Reduced basal transcriptional activity of central nervous system-derived HIV type 1 long terminal repeats. *AIDS research and human retroviruses,* 29**,** 365-370.

GREGER, I. H., PROUDFOOT, N. J., DEMARCHI, F. & GIACCA, M. 1998. Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic acids research,* 26**,** 1294-1300.

GROEN, J. N. & MORRIS, K. V. 2013. Chromatin, non-coding RNAs, and the expression of HIV. *Viruses,* 5**,** 1633-1645.

GULICK, R. M., MELLORS, J. W., HAVLIR, D., ERON, J. J., GONZALEZ, C., MCMAHON, D., RICHMAN, D. D., VALENTINE, F. T., JONAS, L. & MEIBOHM, A. 1997. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *New England Journal of Medicine,* 337**,** 734-739.

GÜRTLER, L. G., HAUSER, P. H., EBERLE, J., VON BRUNN, A., KNAPP, S., ZEKENG, L., TSAGUE, J. M. & KAPTUE, L. 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *Journal of virology,* 68**,** 1581-1585.

GUTH, C. A. & SODROSKI, J. 2014. Contribution of PDZD8 to stabilization of the human immunodeficiency virus type 1 capsid. *Journal of virology,* 88**,** 4612-4623.

HAALAND, R. E., HAWKINS, P. A., SALAZAR-GONZALEZ, J., JOHNSON, A., TICHACEK, A., KARITA, E., MANIGART, O., MULENGA, J., KEELE, B. F. & SHAW, G. M. 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS pathogens,* 5.

HAHN, B. H., SHAW, G. M., ARYA, S. K., POPOVIC, M., GALLO, R. C. & WONG-STAAL, F. 1984. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature,* 312**,** 166-169.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series, 1999. [London]: Information Retrieval Ltd., c1979-c2000., 95-98.

HAMMER, S. M., SQUIRES, K. E., HUGHES, M. D., GRIMES, J. M., DEMETER, L. M., CURRIER, J. S., ERON JR, J. J., FEINBERG, J. E., BALFOUR JR, H. H. & DEYTON, L. R. 1997. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *New England Journal of Medicine,* 337**,** 725-733.

HAHN, B. H., SHAW.G. M., DE, K. M. & SHARP P. M. AIDS as a zoonosis: Scientific and public health implications. 2000. Science, 287, 607-614.

HAN, Y., LIN, Y. B., AN, W., XU, J., YANG, H.-C., O'CONNELL, K., DORDAI, D., BOEKE, J. D., SILICIANO, J. D. & SILICIANO, R. F. 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell host & microbe,* 4**,** 134-146.

HARRICH, D., GARCIA, J., WU, F., MITSUYASU, R., GONAZALEZ, J. & GAYNOR, R. 1989. Role of SP1-binding domains in in vivo transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat. *Journal of virology,* 63**,** 2585-2591.

HEMELAAR, J. 2012. The origin and diversity of the HIV-1 pandemic. *Trends in molecular medicine,* 18**,** 182-192.

HEMELAAR, J., GOUWS, E., GHYS, P. D. & OSMANOV, S. 2011. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS (London, England),* 25**,** 679.

HENDERSON, A. J., ZOU, X. & CALAME, K. L. 1995. C/EBP proteins activate transcription from the human immunodeficiency virus type 1 long terminal repeat in macrophages/monocytes. *J Virol,* 69.

HENDERSON, L. E., BOWERS, M., SOWDER, R., SERABYN, S., JOHNSON, D., BESS, J., ARTHUR, L., BRYANT, D. & FENSELAU, C. 1992. Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *Journal of Virology,* 66**,** 1856-1865.

HERNANDEZ-VARGAS, E. A. & MIDDLETON, R. H. 2013. Modeling the three stages in HIV infection. *Journal of theoretical biology,* 320**,** 33-40.

HULME, A. E., PEREZ, O. & HOPE, T. J. 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences,* 108**,** 9975-9980.

HUNT, G. & TIEMESSEN, C. 2000. Occurrence of additional NF-kappaB-binding motifs in the long terminal repeat region of South African HIV type 1 subtype C isolates. *AIDS research and human retroviruses,* 16**,** 305-306.

HUNT, G. M., JOHNSON, D. & TIEMESSEN, C. T. 2001. Characterisation of the long terminal repeat regions of South African human immunodeficiency virus type 1 isolates. *Virus Genes,* 23**,** 27-34.

INOUE, J.-I., KERR, L. D., RASHID, D., DAVIS, N., BOSE, H. R. & VERMA, I. M. 1992. Direct association of pp40/I kappa B beta with rel/NF-kappa B transcription factors: role of ankyrin repeats in the inhibition of DNA binding activity. *Proceedings of the National Academy of Sciences,* 89**,** 4333-4337.

IORDANSKIY, S., BERRO, R., ALTIERI, M., KASHANCHI, F. & BUKRINSKY, M. 2006. Intracytoplasmic maturation of the human immunodeficiency virus type 1 reverse transcription complexes determines their capacity to integrate into chromatin. *Retrovirology,* 3**,** 4.

JEENINGA, R. E., HOOGENKAMP, M., ARMAND-UGON, M., DE BAAR, M., VERHOEF, K. & BERKHOUT, B. 2000. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *Journal of virology,* 74**,** 3740-3751.

JIANG, G., ESPESETH, A., HAZUDA, D. J. & MARGOLIS, D. M. 2007. c-Myc and Sp1 contribute to proviral latency by recruiting histone deacetylase 1 to the human immunodeficiency virus type 1 promoter. *Journal of virology,* 81**,** 10914-10923.

JONES, K. A., KADONAGA, J. T., LUCIW, P. A. & TJIAN, R. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science,* 232**,** 755-759.

JONES, K. A. & PETERLIN, B. M. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annual review of biochemistry,* 63**,** 717-743.

JORDAN, A., BISGROVE, D. & VERDIN, E. 2003. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *The EMBO journal,* 22**,** 1868-1877.

JOSEPH, S. B., SWANSTROM, R., KASHUBA, A. D. & COHEN, M. S. 2015. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nature reviews Microbiology,* 13**,** 414.

KAMINE, J., SUBRAMANIAN, T. & CHINNADURAI, G. 1991. Sp1-dependent activation of a synthetic promoter by human immunodeficiency virus type 1 Tat protein. *Proceedings of the National Academy of Sciences,* 88**,** 8510-8514.

KAO, S.-Y., CALMAN, A. F., LUCIW, P. A. & PETERLIN, B. M. 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature,* 330**,** 489.

KARAGEORGOS, L., LI, P. & BURRELL, C. 1993. Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS research and human retroviruses,* 9**,** 817-823.

KARIM, S. A. & KARIM, Q. A. 2010. *HIV/Aids in South Africa*, Cambridge University Press.

KARIUKI, S. M., SELHORST, P., ARIËN, K. K. & DORFMAN, J. R. 2017. The HIV-1 transmission bottleneck. *Retrovirology,* 14**,** 22.

KARN, J. 2000. Tat, a novel regulator of HIV transcription and latency. *HIV Sequence Compendium 2000***,** 2-18.

KAUDER, S. E., BOSQUE, A., LINDQVIST, A., PLANELLES, V. & VERDIN, E. 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS pathogens,* 5.

KEELE, B. F., GIORGI, E. E., SALAZAR-GONZALEZ, J. F., DECKER, J. M., PHAM, K. T., SALAZAR, M. G., SUN, C., GRAYSON, T., WANG, S. & LI, H. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences,* 105**,** 7552-7557.

KEELE, B. F., VAN HEUVERSWYN, F.,LI, Y., BAILES, E., TAKEHISA, J., SANTIAGO, M. L., BIBOLLET-RUCHE, F., CHEN, Y., WAIN, L. V., LIEGEOIS, F., LOUL, S., NGOLE, E. M., BIENVENUE, Y., DELAPORTE, E., BROOKFIELD, J. F. Y., SHARP, P. M., SHAW,G. M., PEETERS, M. & HAHN. B. H. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. 2006. Science, 313, 5786, 523-526.

KHARSANY, A. B. & KARIM, Q. A. 2016. HIV infection and AIDS in sub-Saharan Africa: current status, challenges and opportunities. *The open AIDS journal,* 10**,** 34.

KIERAN, M., BLANK, V., LOGEAT, F., VANDEKERCKHOVE, J., LOTTSPEICH, F., LE BAIL, O., URBAN, M. B., KOURILSKY, P., BAEUERLE, P. A. & ISRAËL, A. 1990. The DNA binding subunit of NF-κB is identical to factor KBF1 and homologous to the rel oncogene product. *Cell,* 62**,** 1007-1018.

KIM, H. G., KIM, K.-C., ROH, T.-Y., PARK, J., JUNG, K.-M., LEE, J.-S., CHOI, S.-Y., KIM, S. S. & CHOI, B.-S. 2011. Gene silencing in HIV-1 latency by polycomb repressive group. *Virology journal,* 8**,** 179.

KIM, Y. K., BOURGEOIS, C. F., ISEL, C., CHURCHER, M. J. & KARN, J. 2002. Phosphorylation of the RNA polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. *Molecular and cellular biology,* 22**,** 4622-4637.

KLATZMANN, D., CHAMPAGNE, E., CHAMARET, S., GRUEST, J., GUETARD, D., HERCEND, T., GLUCKMAN, J.-C. & MONTAGNIER, L. 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature,* 312**,** 767-768.

LAU, K. A. & WONG, J. J. 2013. Current trends of HIV recombination worldwide. *Infectious disease reports,* 5.

LEARN, G. H., MUTHUI, D., BRODIE, S. J., ZHU, T., DIEM, K., MULLINS, J. I. & COREY, L. 2002. Virus population homogenization following acute human immunodeficiency virus type 1 infection. *Journal of virology,* 76**,** 11953-11959.

LENARDO, M. J. & BALTIMORE, D. 1989. NF-κB: a pleiotropic mediator of inducible and tissue-specific gene control. *Cell,* 58**,** 227-229.

LENASI, T., CONTRERAS, X. & PETERLIN, B. M. 2008. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell host & microbe,* 4**,** 123-133.

LEWINSKI, M., BISGROVE, D., SHINN, P., CHEN, H., HOFFMANN, C., HANNENHALLI, S., VERDIN, E., BERRY, C., ECKER, J. & BUSHMAN, F. 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *Journal of virology,* 79**,** 6610-6619.

LIM, P. S., HARDY, K., BUNTING, K. L., MA, L., PENG, K., CHEN, X. & SHANNON, M. F. 2009. Defining the chromatin signature of inducible genes in T cells. *Genome biology,* 10**,** R107.

LIU, J., DENG, Y., LI, Q., DEY, A. K., MOORE, J. P. & LU, M. 2010. Role of a putative gp41 dimerization domain in human immunodeficiency virus type 1 membrane fusion. *Journal of virology,* 84**,** 201-209.

LUGER, K., MÄDER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature,* 389**,** 251-260.

LUKIC, Z., DHARAN, A., FRICKE, T., DIAZ-GRIFFERO, F. & CAMPBELL, E. M. 2014. HIV-1 uncoating is facilitated by dynein and kinesin 1. *Journal of virology,* 88**,** 13613-13625.

MAJELLO, B., NAPOLITANO, G., DE LUCA, P. & LANIA, L. 1998. Recruitment of human TBP selectively activates RNA polymerase II TATA-dependent promoters. *Journal of Biological Chemistry,* 273**,** 16509-16516.

MALIM, M. H., HAUBER, J., LE, S.-Y., MAIZEL, J. V. & CULLEN, B. R. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature,* 338**,** 254-257.

MARX, J. L. 1984. Strong new candidate for AIDS agent; a newly discovered member of the human T-cell leukemia virus family is very closely linked to the immunodeficiency disease. *Science,* 224**,** 475-478.

MATEESCU, B., BOURACHOT, B., RACHEZ, C., OGRYZKO, V. & MUCHARDT, C. 2008. Regulation of an inducible promoter by an HP1β–HP1γ switch. *EMBO reports,* 9**,** 267-272.

MBONDJI-WONJE, C., DONG, M., WANG, X., ZHAO, J., RAGUPATHY, V., SANCHEZ, A. M., DENNY, T. N. & HEWLETT, I. 2018. Distinctive variation in the U3R region of the 5'Long Terminal Repeat from diverse HIV-1 strains. *PloS one,* 13**,** e0195661.

MBONYE, U. & KARN, J. 2014. Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. *Virology,* 454**,** 328-339.

MCALLISTER, J. J., PHILLIPS, D., MILLHOUSE, S., CONNER, J., HOGAN, T., ROSS, H. L. & WIGDAHL, B. 2000. Analysis of the HIV-1 LTR NF-kappaB-proximal Sp site III: evidence for cell type-specific gene regulation and viral replication. *Virology,* 274.

MCDONALD, D., VODICKA, M. A., LUCERO, G., SVITKINA, T. M., BORISY, G. G., EMERMAN, M. & HOPE, T. J. 2002. Visualization of the intracellular behavior of HIV in living cells. *The Journal of cell biology,* 159**,** 441-452.

MERVIS, R. J., AHMAD, N., LILLEHOJ, E. P., RAUM, M. G., SALAZAR, F., CHAN, H. & VENKATESAN, S. 1988. The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *Journal of virology,* 62**,** 3993-4002.

MILLER, M. D., FARNET, C. M. & BUSHMAN, F. D. 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *Journal of virology,* 71**,** 5382-5390.

MONTAGNIER, L. 1999. HUMAN IMMUNODEFICIENCY VIRUSES (RETROVIRIDAE)| General Features.

MONTANO, M. A., NIXON, C. P. & ESSEX, M. 1998. Dysregulation through the NF-κB enhancer and TATA box of the human immunodeficiency virus type 1 subtype E promoter. *Journal of virology,* 72**,** 8446-8452.

MONTANO, M. A., NOVITSKY, V. A., BLACKARD, J. T., CHO, N. L., KATZENSTEIN, D. A. & ESSEX, M. 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *Journal of virology,* 71**,** 8657-8665.

NABEL, G. & BALTIMORE, D. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature,* 326**,** 711.

NAGHAVI, M. H., ESTABLE, M. C., SCHWARTZ, S., ROEDER, R. G. & VAHLNE, A. 2001. Upstream stimulating factor affects human immunodeficiency virus type 1 (HIV-1) long terminal repeat-directed transcription in a cell-specific manner, independently of the HIV-1 subtype and the core-negative regulatory element. *Journal of General Virology,* 82**,** 547-559.

NAIDOO, V. L., MANN, J. K., NOBLE, C., ADLAND, E., CARLSON, J. M., THOMAS, J., BRUMME, C. J., THOBAKGALE-TSHABALALA, C. F., BRUMME, Z. L. & BROCKMAN, M. A. 2017. Mother-to-child HIV transmission bottleneck selects for consensus virus with lower Gag-protease-driven replication capacity. *Journal of virology***,** JVI. 00518-17.

NEEL, C., ETIENNE, L., LI, Y., TAKEHISA, J., RUDICELLI, R. S., BASS, I. N., MOUDINDO, J., MEBENGA, A., ESTEBAN, A., HEUVERSWYN, F. V., LIGEOIS, F., KRANZUSCH, P. J., WALSH, P. D., SANZ, C. M., MORGAN, D. B., NDJANGO, J. N., PLANTEIR, J., LOCATELLI, S., GONDER, M. K., LEENDERTZ, F. H., BOESCH, C., TODD, A., DELAPORTE, E., MPOUDI-NGOLE, E., HAHN, B. N. & PEETERS, M.

Molecular epidemiology of simian immunodeficiency viruss infection in wild-living Gorillas. 2010. *Journal of Virology*, 84, 3, 1464-1476.
NERMUT, M. V. & FASSATI, A. 2003. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. *Journal of virology,* 77**,** 8196-8206.
NOLAN, G. P., GHOSH, S., LIOU, H.-C., TEMPST, P. & BALTIMORE, D. 1991. DNA binding and IκB inhibition of the cloned p65 subunit of NF-κB, a rel-related polypeptide. *Cell,* 64**,** 961-969.
NONNEMACHER, M. R., IRISH, B. P., LIU, Y., MAUGER, D. & WIGDAHL, B. 2004. Specific sequence configurations of HIV-1 LTR G/C box array result in altered recruitment of Sp isoforms and correlate with disease progression. *J Neuroimmunol,* 157.
OBASA, A. E., ASHOKKUMAR, M., NEOGI, U. & JACOBS, G. B. 2019. Mutations in Long Terminal Repeats κB Transcription Factor Binding Sites in Plasma Virus Among South African People Living with HIV-1. *AIDS research and human retroviruses,* 35**,** 572-576.
OKAMURA, H., ARAMBURU, J., GARCÍA-RODRÍGUEZ, C., VIOLA, J. P., RAGHAVAN, A., TAHILIANI, M., ZHANG, X., QIN, J., HOGAN, P. G. & RAO, A. 2000. Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. *Molecular cell,* 6**,** 539-550.
PEETERS, M., COURGNAUD, V., ABELA, B., AUZEL, P.,PURRUT, X., BIBOLLET-RUCHE, F., LOUL, S., LIEGEOIS, F., BUTEL, C., KOULAGNA, D., MPOUDI-NGOLE, E., SHAW, G. M., HAHN, B. H. & DELAPORTE, E. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. 2002. *Emerging Infectious Diseases*, 8, 5, 451-457.
PEETERS, M., MUNDEKE, S. A., NGOLE, E. M. & DELAPORTE, E. Origin of HIV/AIDS and risk for ongoing zoonotic transmission from nonhuman primates to humans. 2010. *HIV Therapy*, 4, 4, 387-390.
PAUZA, C. D. 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology,* 179**,** 886-889.
PAWLICA, P. & BERTHOUX, L. 2014. Cytoplasmic dynein promotes HIV-1 uncoating. *Viruses,* 6**,** 4195-4211.
PENDERGRAST, P. S., MORRISON, D., TANSEY, W. P. & HERNANDEZ, N. 1996. Mutations in the carboxy-terminal domain of TBP affect the synthesis of human immunodeficiency virus type 1 full-length and short transcripts similarly. *Journal of virology,* 70**,** 5025-5034.
PEREIRA, L. A., BENTLEY, K., PEETERS, A., CHURCHILL, M. J. & DEACON, N. J. 2000. SURVEY AND SUMMARY A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic acids research,* 28**,** 663-668.
PERELSON, A. S., ESSUNGER, P., CAO, Y., VESANEN, M., HURLEY, A., SAKSELA, K., MARKOWITZ, M. & HO, D. D. 1997. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature,* 387**,** 188-191.
PERSAUD, D., PIERSON, T., RUFF, C., FINZI, D., CHADWICK, K. R., MARGOLICK, J. B., RUFF, A., HUTTON, N., RAY, S. & SILICIANO, R. F. 2000. A stable latent reservoir for HIV-1 in resting CD4+ T lymphocytes in infected children. *The Journal of clinical investigation,* 105**,** 995-1003.
PESSLER, F. & CRON, R. 2004. Reciprocal regulation of the nuclear factor of activated T cells and HIV-1. *Genes and immunity,* 5**,** 158.
PESSLER, F. & HERNANDEZ, N. 1998. The HIV-1 Inducer of Short Transcripts Activates the Synthesis of 5, 6-Dichloro-1-β-d-benzimidazole-resistant Short Transcripts in Vitro. *Journal of Biological Chemistry,* 273**,** 5375-5384.
PING, Y.-H. & RANA, T. M. 2001. DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *Journal of Biological Chemistry,* 276**,** 12951-12958.
PURCELL, D. & MARTIN, M. A. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of virology,* 67**,** 6365-6378.

QU, D., LI, C., SANG, F., LI, Q., JIANG, Z.-Q., XU, L.-R., GUO, H.-J., ZHANG, C. & WANG, J.-H. 2016. The variances of Sp1 and NF-κB elements correlate with the greater capacity of Chinese HIV-1 B'-LTR for driving gene expression. *Scientific reports,* 6**,** 34532.

RAFATI, H., PARRA, M., HAKRE, S., MOSHKIN, Y., VERDIN, E. & MAHMOUDI, T. 2011. Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. *PLoS biology,* 9.

RAHA, T., CHENG, S. G. & GREEN, M. R. 2005. HIV-1 Tat stimulates transcription complex assembly through recruitment of TBP in the absence of TAFs. *PLoS biology,* 3**,** e44.

ROBERTSON, D. L., ANDERSON, J., BRADAC, J., CARR, J., FOLEY, B., FUNKHOUSER, R., GAO, F., HAHN, B., KALISH, M. & KUIKEN, C. 2000. HIV-1 nomenclature proposal. *Science,* 288**,** 55-55.

ROEBUCK, K. A. & SAIFUDDIN, M. 1999. Regulation of HIV-1 transcription. *Gene Expression, The Journal of Liver Research,* 8**,** 67-84.

ROTH, M. J., SCHWARTZBERG, P. L. & GOFF, S. P. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell,* 58**,** 47-54.

RUBEN, S. M., KLEMENT, J. F., COLEMAN, T. A., MAHER, M., CHEN, C.-H. & ROSEN, C. A. 1992. I-Rel: a novel rel-related protein that inhibits NF-kappa B transcriptional activity. *Genes & development,* 6**,** 745-760.

RUTSTEIN, S. E., ANANWORANICH, J., FIDLER, S., JOHNSON, C., SANDERS, E. J., SUED, O., SAEZ-CIRION, A., PILCHER, C. D., FRASER, C. & COHEN, M. S. 2017. Clinical and public health implications of acute and early HIV detection and treatment: a scoping review. *Journal of the International AIDS Society,* 20**,** 21579.

SALTARELLI, M., HADZIYANNIS, E., HART, C., HARRISON, J., FELBER, B., SPIRA, T. & PAVLAKIS, G. 1996. Analysis of human immunodeficiency virus type 1 mRNA splicing patterns during disease progression in peripheral blood mononuclear cells from infected individuals. *AIDS research and human retroviruses,* 12**,** 1443-1456.

SCHIRALLI LESTER, G. M. & HENDERSON, A. J. 2012. Mechanisms of HIV transcriptional regulation and their contribution to latency. *Molecular biology international,* 2012.

SCHULZ, T. F., WHITBY, D., HOAD, J., CORRAH, T., WHITTLE, H. & WEISS, R. 1990. Biological and molecular variability of human immunodeficiency virus type 2 isolates from The Gambia. *Journal of virology,* 64**,** 5177-5182.

SCHWARTZ, C., CATEZ, P., ROHR, O., LECESTRE, D., AUNIS, D. & SCHAEFFER, E. 2000. Functional interactions between C/EBP, Sp1, and COUP-TF regulate human immunodeficiency virus type 1 gene transcription in human brain cells. *Journal of virology,* 74**,** 65-73.

SCRIBA, T. J., DE VILLIERS, T., TREURNICHT, F. K., ZUR MEGEDE, J., BARNETT, S. W., ENGELBRECHT, S. & VAN RENSBURG, E. J. 2002. Characterization of the South African HIV type 1 subtype C complete 5' long terminal repeat, nef, and regulatory genes. *AIDS research and human retroviruses,* 18**,** 149-159.

SCRIBA, T. J., TREURNICHT, F. K., ZEIER, M., ENGELBRECHT, S. & VAN RENSBURG, E. J. 2001. Characterization and phylogenetic analysis of South African HIV-1 subtype C accessory genes. *AIDS research and human retroviruses,* 17**,** 775-781.

SEN, R. & BALTIMORE, D. 1986. Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a posttranslational mechanism. *Cell,* 47**,** 921-928.

SHAH, S., ALEXAKI, A., PIRRONE, V., DAHIYA, S., NONNEMACHER, M. R. & WIGDAHL, B. 2014. Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT/enhancer binding protein site I. *Virology Journal,* 11**,** 92.

SHARP, P. M. & HAHN, B. H. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine,* 1**,** a006841.

SHAW, J.-P., UTZ, P. J., DURAND, D. B., TOOLE, J. J., EMMEL, E. A. & CRABTREE, G. R. 1988. Identification of a putative regulator of early T cell activation genes. *Science,* 241**,** 202-205.

SHELDON, M., RATNASABAPATHY, R. & HERNANDEZ, N. 1993. Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short RNAs. *Molecular and cellular biology,* 13**,** 1251-1263.

SHIRAKAWA, K., CHAVEZ, L., HAKRE, S., CALVANESE, V. & VERDIN, E. 2013. Reactivation of latent HIV by histone deacetylase inhibitors. *Trends in microbiology,* 21**,** 277-285.

SILICIANO, J. D., KAJDAS, J., FINZI, D., QUINN, T. C., CHADWICK, K., MARGOLICK, J. B., KOVACS, C., GANGE, S. J. & SILICIANO, R. F. 2003a. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine,* 9**,** 727-728.

SILICIANO, J. D., KAJDAS, J., FINZI, D., QUINN, T. C., CHADWICK, K., MARGOLICK, J. B., KOVACS, C., GANGE, S. J. & SILICIANO, R. F. 2003b. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine,* 9**,** 727.

SIMON, F., MAUCLÈRE, P., ROQUES, P., LOUSSERT-AJAKA, I., MÜLLER-TRUTWIN, M. C., SARAGOSTI, S., GEORGES-COURBOT, M. C., BARRÉ-SINOUSSI, F. & BRUN-VÉZINET, F. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature medicine,* 4**,** 1032-1037.

STEVENS, C. D. & MILLER, L. E. 2016. *Clinical Immunology and Serology: A Laboratory Perspetive*, F.A. Davis.

SUHASINI, M. & REDDY, T. R. 2009. Cellular proteins and HIV-1 Rev function. *Current HIV research,* 7**,** 91-100.

TALTYNOV, O., DEMEULEMEESTER, J., CHRIST, F., DE HOUWER, S., TSIRKONE, V. G., GERARD, M., WEEKS, S. D., STRELKOV, S. V. & DEBYSER, Z. 2013. Interaction of transportin-SR2 with Ras-related nuclear protein (Ran) GTPase. *Journal of Biological Chemistry,* 288**,** 25603-25613.

TESMER, V. M., RAJADHYAKSHA, A., BABIN, J. & BINA, M. 1993. NF-IL6-mediated transcriptional activation of the long terminal repeat of the human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences,* 90**,** 7298-7302.

THOMAS, M. C. & CHIANG, C.-M. 2006. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology,* 41**,** 105-178.

TOZSER, J. 2003. Stages of HIV replication and targets for therapeutic intervention. *Current topics in medicinal chemistry,* 3**,** 1447-1457.

TRICKEY, A., MAY, M. T., VEHRESCHILD, J.-J., OBEL, N., GILL, M. J., CRANE, H. M., BOESECKE, C., PATTERSON, S., GRABAR, S. & CAZANAVE, C. 2017. Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *The Lancet HIV,* 4**,** e349-e356.

TSIRKONE, V. G., BLOKKEN, J., DE WIT, F., BREEMANS, J., DE HOUWER, S., DEBYSER, Z., CHRIST, F. & STRELKOV, S. V. 2017. N-terminal half of transportin SR2 interacts with HIV integrase. *Journal of Biological Chemistry,* 292**,** 9699-9710.

TYAGI, M. & KARN, J. 2007. CBF-1 promotes transcriptional silencing during the establishment of HIV-1 latency. *The EMBO journal,* 26**,** 4985-4995.

UNAIDS 2019. UNAIDS data 2019. *UNAIDS/JC2929E http://www. unaids. org/sites/default/files/media_asset/unaidsdata-2018_en. pdf# page= 4&zoom= auto,-62,843. Accessed on,* 12.

VAN LINT, C., AMELLA, C. A., EMILIANI, S., JOHN, M., JIE, T. & VERDIN, E. 1997. Transcription factor binding sites downstream of the human immunodeficiency virus type 1 transcription start site are important for virus infectivity. *Journal of virology,* 71**,** 6113-6127.

VAN LINT, C., EMILIANI, S., OTT, M. & VERDIN, E. 1996. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *The EMBO journal,* 15**,** 1112-1120.

VAN OPIJNEN, T., KAMOSCHINSKI, J., JEENINGA, R. E. & BERKHOUT, B. 2004. The human immunodeficiency virus type 1 promoter contains a CATA box instead of a TATA box for optimal transcription and replication. *Journal of virology,* 78**,** 6883-6890.

VENNER, C. M., NANKYA, I., KYEYUNE, F., DEMERS, K., KWOK, C., CHEN, P.-L., RWAMBUYA, S., MUNJOMA, M., CHIPATO, T. & BYAMUGISHA, J. 2016. Infecting HIV-1 subtype predicts disease progression in women of sub-Saharan Africa. *EBioMedicine,* 13**,** 305-314.

VERDIN, E., PARAS, P. & VAN LINT, C. 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *The EMBO journal,* 12**,** 3249-3259.

VERHOEF, K., SANDERS, R. W., FONTAINE, V., KITAJIMA, S. & BERKHOUT, B. 1999. Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF-κB enhancer element into a GABP binding site. *Journal of virology,* 73**,** 1331-1340.

VINK, C., YEHESKIELY, E., VAN DER MAREL, G. A., VAN BOOM, J. H. & PLASTERK, R. H. 1991. Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic acids research,* 19**,** 6691-6698.

WAHEED, A. A. & FREED, E. O. 2010. The role of lipids in retrovirus replication. *Viruses,* 2**,** 1146-1180.

WALZER, P. D., PERL, D. P., KROGSTAD, D. J., RAWSON, P. G. & SCHULTZ, M. G. 1974. Pneumocystis carinii pneumonia in the United States: epidemiologic, diagnostic, and clinical features. *Annals of internal medicine,* 80**,** 83-93.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T.-Y., PENG, W. & ZHANG, M. Q. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics,* 40**,** 897.

WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-1191.

WEI, P., GARBER, M. E., FANG, S.-M., FISCHER, W. H. & JONES, K. A. 1998. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell,* 92**,** 451-462.

WILLIAMS, S. A., CHEN, L. F., KWON, H., RUIZ-JARABO, C. M., VERDIN, E. & GREENE, W. C. 2006. NF-κB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *The EMBO journal,* 25**,** 139-149.

WU, L., GERARD, N. P., WYATT, R., CHOE, H., PAROLIN, C., RUFFING, N., BORSETTI, A., CARDOSO, A. A., DESJARDIN, E. & NEWMAN, W. 1996. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature,* 384**,** 179.

XU, H., FRANKS, T., GIBSON, G., HUBER, K., RAHM, N., DE CASTILLIA, C. S., LUBAN, J., AIKEN, C., WATKINS, S. & SLUIS-CREMER, N. 2013. Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. *Retrovirology,* 10**,** 70.

YEDAVALLI, V. S., BENKIRANE, M. & JEANG, K.-T. 2003. Tat and trans-activation-responsive (TAR) RNA-independent induction of HIV-1 long terminal repeat by human and murine cyclin T1 requires Sp1. *Journal of Biological Chemistry,* 278**,** 6404-6410.

YOON, J.-B., LI, G. & ROEDER, R. G. 1994. Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. *Molecular and cellular biology,* 14**,** 1776-1785.

ZABEL, U. & BAEUERLE, P. A. 1990. Purified human IκB can rapidly dissociate the complex of the NF-κB transcription factor with its cognate DNA. *Cell,* 61**,** 255-265.

ZAGURY, J., FRANCHINI, G., REITZ, M., COLLALTI, E., STARCICH, B., HALL, L., FARGNOLI, K., JAGODZINSKI, L., GUO, H. & LAURE, F. 1988. Genetic variability between isolates of human immunodeficiency virus (HIV) type 2 is comparable to the variability among HIV type 1. *Proceedings of the National Academy of Sciences,* 85**,** 5941-5945.

ZHANG, L., HUANG, Y., YUAN, H., CHEN, B. K., IP, J. & HO, D. D. 1997. Genotypic and phenotypic characterization of long terminal repeat sequences from long-term survivors of human immunodeficiency virus type 1 infection. *Journal of virology,* 71**,** 5608-5613.

ZHANG, L., YAN, L., JIANG, J., WANG, Y., JIANG, Y., YAN, T. & CAO, Y. 2014. The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast Candida albicans. *Virulence,* 5**,** 655-664.

ZHONG, H., MAY, M. J., JIMI, E. & GHOSH, S. 2002. The phosphorylation status of nuclear NF-κB determines its association with CBP/p300 or HDAC-1. *Molecular cell,* 9**,** 625-636.

ZHOU, L., SOKOLSKAJA, E., JOLLY, C., JAMES, W., COWLEY, S. A. & FASSATI, A. 2011. Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. *PLoS pathogens,* 7.

ZHU, Y., PE'ERY, T., PENG, J., RAMANATHAN, Y., MARSHALL, N., MARSHALL, T., AMENDT, B., MATHEWS, M. B. & PRICE, D. H. 1997. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes & development,* 11**,** 2622-2632.

Appendix 1: BREC Ethics approval letter

UNIVERSITY OF
KWAZULU-NATAL
INYUVESI
YAKWAZULU-NATALI

06 March 2020

Ms Shamara Naicker (215000531)
School of Lab Med & Medical Science
Medical School

Dear Ms Shamara Naicker,

Protocol reference number: BREC/00001051/2020
Project title: Functional characterization of human immunodeficiency virus type 1 (HIV-1) subtype C transmitted/founder (T/F) viruses long terminal repeat (LTR) variants and their association with Disease outcome
Degree Purposes: MMedSc

EXPEDITED APPLICATION: APPROVAL LETTER

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application.

The conditions have been met and the study is given **full ethics approval** and may begin as from 06 March 2020. **Please ensure that outstanding site permissions are obtained and forwarded to BREC for approval before commencing research at a site.**

This approval is valid for one year from **06 March 2020.** To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2015), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be **noted** by a full Committee at its next meeting taking place on **14 April 2020.**

Yours sincerely

Prof V Rambiritch
Chair: Biomedical Research Ethics Committee