UNIVERSITY OF KWAZULU-NATAL ™

INYUVESI YAKWAZULU-NATALI

Central nervous system (CNS) derived human immunodeficiency virus type 1 (HIV-1) subtype C long terminal repeat (LTR) genetic and functional variation mediates high viral load in this compartment of tuberculous meningitis (TBM) co-infected patients

Submitted by: Wenzile Senorita Ntshangase

Supervised by: Dr Paradise Madlala

In fulfilment of the requirements for the degree: Master of Medical Science (Virology)

School of Laboratory Medicine

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.

Wenzile Ntshangase (Student)

Dr. Paradise Z. Madlala (Supervisor)

PLAGIARISM DECLARATION

I, Wenzile Senorita Ntshangase declare that:

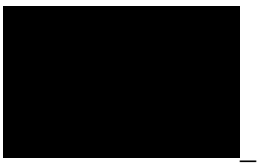
1. The research reported in this dissertation, except where otherwise indicated, is my original work.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This data does not contain other person's data, graphs, pictures, or other information, unless specifically acknowledged as being sourced from other persons.

4. This dissertation 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.
    c. This dissertation 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 reference section.

Signed: _____  _ Date: __08 June 2023_____

DEDICATION

To my grandparents: Bertina Thandiwe and Simon Mfaniseni Ntshangase

Even though you are no longer with us, you will forever be in my heart. Words can never describe enough how grateful I am of how much you dedicated your life in raising me.

My grandmother would have said, “Mgazi, Sobethu, Masiphula kaMamba, Biyela, umuntu akaqedwa, uSebenzile ntombazane kagogo, uSebenzile Mgazi omuhle”!

ACKNOWLEDGMENTS

I would like to thank the following people and organisations who have supported me throughout my Master's degree:

• Dr. Madlala for his supervision, motivation, and attention to detail.

• Prof. Ndung'u his unwavering support and insight throughout the project.

• Staff and students at the HIV Pathogenesis Programme, Hasso-Plattner Research Laboratory

• Dr. Patel and Nkosi Albert Luthuli Central Hospital for providing access to the Tuberculous meningitis study.

• The HIV Pathogenesis Programme (HPP) grant and the Poliomyelitis Research Foundation for financial assistance.

ETHICS APPROVAL

The ethics approval for this study was obtained from the Biomedical Research Committee of Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. The ethics approval reference number is as follows: **E325/05**

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

BBB: Blood–brain barrier

Bp: Base Pairs

CCR5: C-C chemokine receptor type 5

CDC: Centres for Disease Control and Prevention

CDK2: Cyclin-Dependent Kinase 2

CDK9: Cyclin-Dependent Kinase 9

CNS: Central Nervous System

Copies/ml: Copies Per Millilitre

CSF: Cerebrospinal Fluid

CT: Computerized Tomography

CXCR4: C-X-C chemokine receptor type 4

Cyc T1: Cyclin T1

DMEM: Dulbecco Modified Eagle Medium

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphate

E-box: Enhancer box

EDTA: Ethylenediaminetetraacetic acid

Env: Envelope Glycoprotein

Gag: Group Specific Antigen

GM: Growth Media

GRID: Gay-Related Immunodeficiency

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV-1: Human immunodeficiency virus type 1

HIV-2: Human immunodeficiency virus type 2

HPP: HIV Pathogenesis Programme

IQR: Interquartile Range

IN: Integrase

Inr: Initiator

IST: Inducer of short transcripts

KS: Kaposi's Sarcoma

LAV: Lymphadenopathy Associated Virus

LB: Lysogeny Broth

LEDGF/p75: Lens Epithelium-derived Growth factor

LTR: Long Terminal Repeat

Luc: Luciferase

M.tb: Mycobacterium Tuberculosis

NF-κB: Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells

NaOAc: Sodium Acetate

NELF: Negative elongation factor

NFAT: Nuclear factor of Transcription

PBS: Primer binding site

PCR: Polymerase chain reaction

PIC: Pre-integration Complex

PLHIV: People living with HIV

PPT: Polypurine tract

P-TEFb: Positive transcription elongation factor-b

PCP: Pneumocystis Pneumonia

PID: Patient Identification

Pol II: RNA Polymerase II

PR: Protease

R: Repeat region

RLU: Relative light unit

RNA: Ribonucleic acid

RNase H: Ribonuclease H

RNA pol II: RNA Polymerase II

Rev: Regulator of Virion Protein

RRE: Rev response element

RT: Reverse Transcriptase

RTC: Reverse transcription complex

SIV: Simian immunodeficiency virus

Tat: Transactivator of Transcription

TBP: Tata Binding Protein

TFs: Transcription factors

TFBS: Transcription factor binding site

U3 region: Unique 3 region

U5: Unique 5 region

U3R- Unique 3 Repeat region

ABSTRACT

Background: Human immunodeficiency virus type 1 (HIV-1) ribonucleic acid (RNA) is characteristically lower in the central nervous system (CNS) than in plasma of antiretroviral treatment naïve patients. Paradoxically, there is higher HIV-1 viral load in the cerebral spinal fluid (CSF) than plasma of treatment naïve patients co-infected with tuberculous meningitis (TBM). The mechanisms that govern high viral replication in the CNS of TBM co-infected antiretroviral therapy naïve patients remain to be determined.

Methodology: The study population comprised of 17 TBM and 3 non-TBM participants selected from an HIV-1 positive and TBM co-infected cohort. The HIV-1 viral RNA was reversed transcribed into complementary deoxyribonucleic acid (cDNA) thus the U3/R region of 3' long terminal repeat (LTR) was amplified from CSF and plasma RNA by KAPA HiFi HotStart PCR Kits (ThermoFisher Scientific, Invitrogen™, USA). The patients CSF and plasma derived LTR were subsequently cloned into a pGL3 plasmid and further transfected in Jurkat and Astrocyte cell lines to assess the LTR transcriptional activity using Bright-Glo™ Luciferase Assay System (Promega, Madison, WI, USA).

Results: CSF derived LTR had a significantly ($p<0.0001$) higher basal and Tat induced transcriptional activity compared to plasma derived LTR in Astrocyte (SVG) cell line. Similarly, CSF derived LTR had significantly higher ($p=0.0024$) Tat induced transcriptional activity compared to plasma derived LTR in Jurkat cell lines. LTR sequences containing an Adenine (A) at position 5 of the Sp1III binding site were associated with significantly high basal ($p<0.0001$) and Tat induced ($p=0.0002$) transcriptional activity compared to the LTR sequences containing a Thymine (T) at the same position when it was assessed in SVG cell. A similar case was observed in Jurkat cell lines. Consistently, CSF LTR sequences containing an A at position 5 of the Sp1III transcription binding site were associated with significantly higher HIV-1 viral load compared to LTR sequences containing a T at the same position ($p=0.0093$).

Conclusion: Our data clearly show that CSF derived LTR from TBM co-infected individuals exhibit significantly higher transcriptional. Particularly, sequences containing the A5T mutation are significantly associated with higher LTR transcriptional activity and viral load.

CHAPTER 1: INTRODUCTION

Introduction:

Since its discovery as a causative pathogen of acquired immune deficiency syndrome (AIDS), human immunodeficiency virus (HIV) continues to be a major global public health concern. Disease progression to AIDS is often accompanied by associated opportunistic infections that adds a massive weight on the global public health issue (Vaillant and Naik, 2021). Sub-Saharan Africa is the region that carries the highest burden of HIV/AIDS, and the most common opportunistic infection which is a leading cause of death amongst HIV-1 infected individuals in sub-Saharan Africa is Mycobacterium tuberculosis (M. tb) (Ghislain et al., 2021, Bruchfeld et al., 2015). Therefore, this study will be looking more into HIV/TB co-infections.

HIV-1 enters the central nervous system (CNS) within two weeks of infection and remains in this compartment throughout the lifespan of the infected individuals (Grill and Price, 2014). However, HIV-1 ribonucleic acid (RNA) is characteristically lower in the CNS than in plasma of antiretroviral treatment naïve individuals (Valcour et al., 2012). This phenomenon may be attributed to the abundance of long-lived microglia and astrocyte cells which do not facilitate efficient entry and/or sustained viral replication in the CNS (Joseph et al., 2015).

In an attempt to decipher the mechanisms underlining the phenomenon of marginally low viral replication in the CNS of HIV-1 infected individuals, a previous study reported that CNS derived HIV type 1 (1) long terminal repeat (LTR) sequences were phylogenetically distinct from the non-CNS compartment derived HIV-1 LTR (Gray et al., 2016). Specifically, they showed that CNS derived HIV-1 LTR exhibited reduced basal transcriptional activity in both Astrocytes and T cell lines (Gray et al., 2016). The reduced basal transcriptional activity associated with CNS derived HIV-1 LTR in astrocytes could possibly explain the lower HIV-1 replication in the CNS compared to plasma of treatment naïve people living with HIV-1 (PLHIV).

Paradoxically, there is higher HIV-1 RNA in the cerebral spinal fluid (CSF) than plasma of treatment naïve PLHIV who are co-infected with tuberculous meningitis (TBM) (Morris et al., 1998). Morris et al. clearly shows that TBM-co infection was associated with higher HIV-1 RNA copies in the CSF compared to co-infection with other meningitides as well as HIV-1 mono-infection (Morris et al., 1998). Furthermore, another study demonstrated that the CSF of TBM co-infected individuals have a significantly higher viral load compared to the CSF of non-TBM individuals (Seipone et al., 2018). However, Seipone *et al*. did not find any virological and host immunological biomarkers that could distinguish TBM from other meningitis (Seipone et al., 2018). Therefore, the mechanisms responsible for higher viral load in CSF of TBM co-infected individuals remains to be determined.

It is known that the HIV-1 deoxyribonucleic acid (DNA) genome is 9.7 kb long, comprising of nine viral genes and is flanked by exactly the same LTR at both the 5' and 3' end (reviewed in (Groen and Morris, 2013)). The HIV-1 LTR is approximately 640 bp long and is divided into 3 regions namely the unique 3 (U3), repeat (R) and unique 5 (U5) regions (Mbondji-Wonje et al., 2018). The U3 region is further divided into 3 domains namely the core promoter, core enhancer and modulatory domain (Mbondji-Wonje et al., 2018). Briefly, the core promoter domain is defined by the presence of a TATA box, E-box, an initiator element, and 3 specificity protein 1 (Sp1) transcription factor binding sites. The core enhancer domain consists of the nuclear factor – kappa B (NF-κB) binding sites and the modulatory domain is composed of activator protein 3 (AP-3), nuclear factor of activated T cells (NFAT) and CCAAT enhancer binding protein (C/EBP) (Kilareski et al., 2009, Mbondji-Wonje et al., 2018). While the 3' LTR acts in transcription termination, the 5' LTR acts as the promoter that drives viral gene expression (Mbondji-Wonje et al., 2018).

Inter and intra-subtype specific genetic variation that translate to differential LTR transcriptional activity have been mapped to the U3 region (Bachu et al., 2012). The majority of HIV-1 subtypes including the prototype, HIV-1 subtype B contain two NF- kB binding sites whereas the circulating recombinant form_AE (CRF01_AE) has one NF-κB binding site (Montano et al., 1998). On the other hand, HIV-1 subtype C LTR harbour three to four NF-κB binding sites (Bachu et al., 2012a, Boullosa et al., 2014, Obasa et al., 2019). Bachu *et al.* reported that in India, subtype C viral strains

comprising of four NF-κB were gradually expanding and replacing the standard subtype C viruses which consists of three NF-κB.

Interestingly, Bachu *et al*. showed that subtype C LTR containing four NF-κB binding sites exhibit enhanced transcriptional activity compared to the standard subtype C LTR, which contains only three NF-κB binding sites or to other LTR elements that exhibit either two or one NF-κB binding sites. They further demonstrated that subtype C LTR with four NF-κB binding sites translate to higher viral replication capacities (Bachu et al., 2012). However, it is not known whether LTR genetic variation may be the cause of higher viral load in the CNS compartment.

1.1 Aim:

The aim of the study is to characterize the genetic and functional variation in CNS and plasma derived HIV-1 subtype C LTR element and associate this with markers of disease progression such as viral load in these compartments of TBM co-infected individuals.

1.2 Hypothesis:

We hypothesized that LTR genetic variation in the CNS may be the cause of higher HIV-1 viral load in the CNS compared to plasma compartment of TBM co-infected individuals.

1.3 Objectives:

1. Amplify, sequence and cloned CSF and plasma derived HIV-1 subtype C LTR elements from the TBM co-infected cohort; this is to study inter-compartmental LTR sequence diversity and functional activity.
2. Determine the transcriptional activity of CSF and plasma derived LTR elements using a transcriptional assay – This is to study functional differences between CSF and plasma derived LTR elements.
3. Associate the transcriptional activity of HIV-1 LTR elements with markers of disease progression such as viral loads from CSF and plasma of TBM cohort.

CHAPTER 2: LITERATURE REVIEW

2.1 Background

Amongst the different narrated histories, one states that the ancestor of human immunodeficiency virus (HIV) was first discovered in central Africa from the early 1920's in Kinshasa in the Democratic Republic of Congo (Wise, 2014). Epidemiologic and phylogenetic analysis of archived HIV positive blood samples indicated that the expansion of HIV commenced via the cross-species barrier from chimpanzees to humans, whereby the genetically similar simian immunodeficiency virus (SIV) was transmitted into the Congolese human population (Faria et al., 2014).

A familiar mode of transmission is believed to be that the Congolese communities hunted chimpanzees that were already infected with the SIV (Sharp and Hahn, 2011). As a consequence, they consumed or rather contacted SIV containing contaminated blood (Sharp and Hahn, 2011). The zoonotic transmission and infection by the SIV virus resulted in SIV evolving to HIV in the human host (Williams and Burdo, 2009), this will be described in detail in the following sections.

The spread of the virus is believed to may have started from Kinshasa along infrastructure routes (roads, railways, and rivers) via migrants and the sex trade (Gallagher, 2014). It is documented that HIV-1 may have spread from Africa to Haiti in the 1960s, then in 1970 proceeded from the Caribbean to New York City and later in the decade landed in San Francisco (Gallagher, 2014). The international travel from the United States could have also encouraged the spread of the virus across the globe (Gallagher, 2014). Although HIV commenced in the 1970s in the United States, it was only recognized publicly in the early 1980s from the first cases of HIV reported in USA (Williams and Burdo, 2009).

In June 1981, a weekly report was published by the Centre for Disease Control and Prevention (CDC) which outlined the occurrence of *Pneumocystis carinii* pneumonia (PCP) in five sexually active homosexual North-American young men (CDC, 1982). The PCP is an infection known to be caused by the fungus *Pneumocystis jirovecii,* as a result patient suffer from fluid build-up in the

lungs and show symptoms like those of fever and chest discomfort (CDC, 1982). Of the five PCP cases reported, these patients did not have any relation with each other, however, these individuals were reported to be using inhalant drugs (CDC, 1982).

Nonetheless, PCP is an opportunistic infection and most patients who had PCP had a medical condition that weakens their immune system. Unfortunately, two of these five men had already died at the time the CDC report was published (CDC, 1981). Following the PCP event, a subsequent report was published describing a case of a rare malignancy called Kaposi sarcoma (KS) observed within 26 homosexual men in July 1981(CDC, 1981). The KS was said to be an aggressive cancer that was most common in elderly men. This form of aggressive cancer began to spread, infecting about 41 homosexual men. Of these 41 individuals, eight died in less than 2 years after being diagnosed (CDC, 1981).

In an attempt to manage the outbreak of PCP and KS infection, in August 1981 a clinical case definition was then set, which was going to determine if the person's illness is considered as a case to be included as part of the outbreak. With the increasing number of infections recorded, the majority cases were amongst homosexual men which led them to classifying the disease as Gay Related Immunodeficiency (GRID) (CDC, 1981). In early 1982, there was a significant increase in the cases reported within the heterosexual population (CDC, 1982a, CDC, 1982b).

Such observations then proved that the disease was not only restricted within the homosexual population, as a result in 1982 the CDC renamed the disease as Acquired Immune Deficiency Syndrome (AIDS) (CDC, 1982a, CDC, 1982b). Therefore in 1982, the case definition of AIDS was defined by the susceptibility to opportunistic infections, low CD4+ T cell count (< 200 cells/mm^3) and inflamed lymph nodes. However, the cause of the condition remained elusive (CDC, 1982b).

In 1983 following the AIDS case definition, a virologist Luc Montagnier and co-workers isolated the Lymphadenopathy-associates virus (LAV) which was recognized to be common in individuals suffering from AIDS and was classified into the Lentivirus genus, Retroviridae family (Barré-

Sinoussi et al., 1983). Eventually in 1984, Gallo and co-workers named it human T lymphotropic virus type III (HTLV-III) (Gallo et al., 1984).

Furthermore, in the same year they confirmed that the HTLV-III could be the cause of AIDS (Gallo et al., 1984). After tremendous research on the LAV, it was only in 1986 where the International Committee of Taxonomy of Viruses (ICTV) renamed the virus as Human Immunodeficiency Virus (HIV) (Fauquet, 1999).

2.2 Epidemiology

Since it was discovered in 1983 as a causative agent of AIDS (Barré-Sinoussi et al., 1983), HIV continues to be a major global public health concern. According to the recent joint United Nations Programme on HIV/AIDS (UNAIDS) 2020 global report, 37.7 million people were living with HIV, 1.5 million of individuals became newly infected with HIV and 680 000 people died from AIDS-related illnesses globally (UNAIDS, 2020). The most acquired opportunistic infection in HIV/AIDS patients is Tuberculosis (TB) caused by Mycobacterium tuberculosis (M. tb), and it is the leading cause of death in people living with HIV (PLHIV) worldwide (UNAIDS, 2020) (WHO, 2020). As aforementioned, Sub-Saharan Africa is the region with the greatest burden of the HIV/TB co-infection, where TB infection is responsible for approximately 70% of AIDS related deaths worldwide (WHO, 2020). Therefore, HIV/TB co-infection is going to be discussed in the subsequent sections since it is the focus of this study.

There are two types of HIV: HIV type 1 (HIV-1) and type 2 (HIV-2) with a few genetic differences that set them apart (Figure 1) (Clavel et al., 1986). While HIV-2 is mainly restricted to West Africa, HIV-1 is globally distributed and accounts for 90% of the global HIV infections (Aldrich and Hemelaar, 2012). HIV-1 is divided into four distinct lineages, group M (major), O (outlier), N (non-major), and P (newly identified) (Figure 1), each of which resulting from an independent cross-species transmission event (reviewed in (Bbosa et al., 2019)).

The Group M viruses are further sub-divided into nine genetically different subtypes A-D, F-H, J and K with approximately 98 circulating recombinant forms (CRFs) and 63 unique recombinant forms (URFs) that have been identified to date (Figure 1) (Bbosa et al., 2019, Hemelaar et al., 2019). The CRFs is a recombinant event between pure subtypes which results in the formation of new strains of HIV-1 whereas a unique recombinant is isolated from a single HIV-1 infected patient (Keng et al., 2009) two individuals (Capoferri et al., 2020).

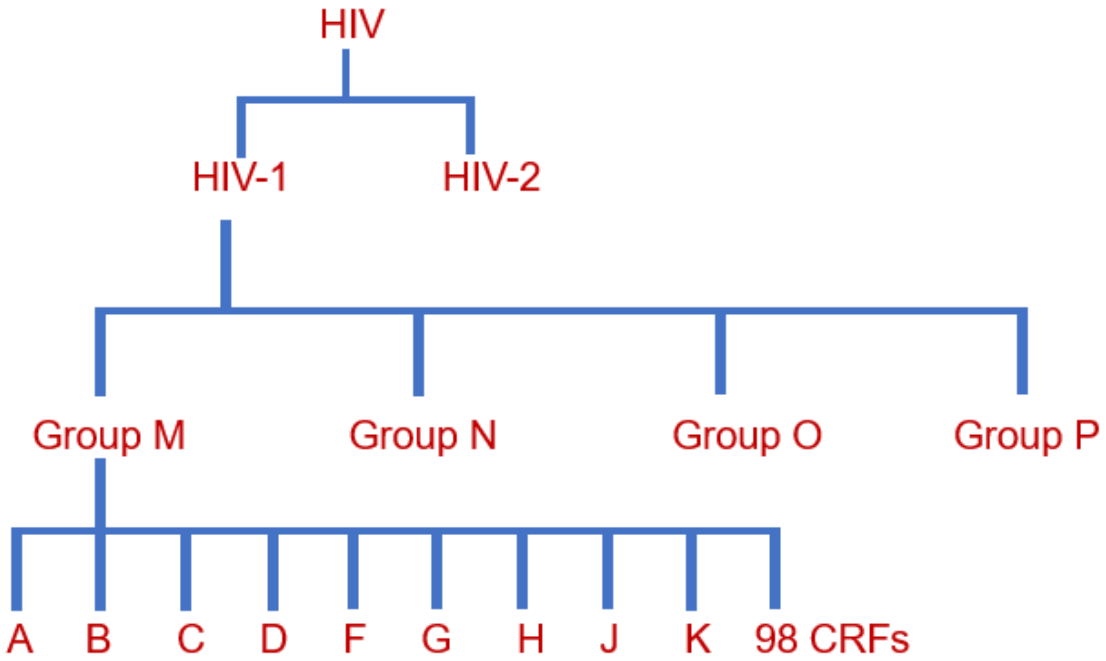


Figure 1: HIV diversification. HIV is classified into two types: HIV-1 and HIV-2. HIV-1 is divided into four distinct groups: Group M (major), O (outlier), N (non-major), and P (newly existing). Group M is further subdivided into nine genetically different subtypes A-D, F-H, J and K with approximately 98 circulating recombinant forms (CRFs). Reproduced with modification from https://www.hiv.uw.edu/pdf/screening-diagnosis/epidemiology/core-concept/all.

However, the HIV-1M subtypes and CRFs are unevenly distributed worldwide (Hemelaar et al., 2019). Subtype A is predominant in Eastern Europe, Central Asia and Central and Eastern Africa, while subtype B is predominant in North America, West and Central Europe and Australia (Figure 2) (Hemelaar et al., 2019). Subtype D is predominant in and East Africa, whereas subtype G is mainly present in West and Central Africa (Figure 2) (Hemelaar et al., 2019). Subtypes F, H, J and K have remained fixed at low levels and are scattered in various countries (Hemelaar et al., 2019). There are three CRF's which significantly adds to the epidemic CRF01_AE, CRF02_AG and CRF07_BC (Figure 2).

The CRF01_AE is abundant in Southeast Asia while CRF_02AG is established in West Africa. Furthermore, CRF07_BC is found in China (Figure 2) (Hemelaar et al., 2019).

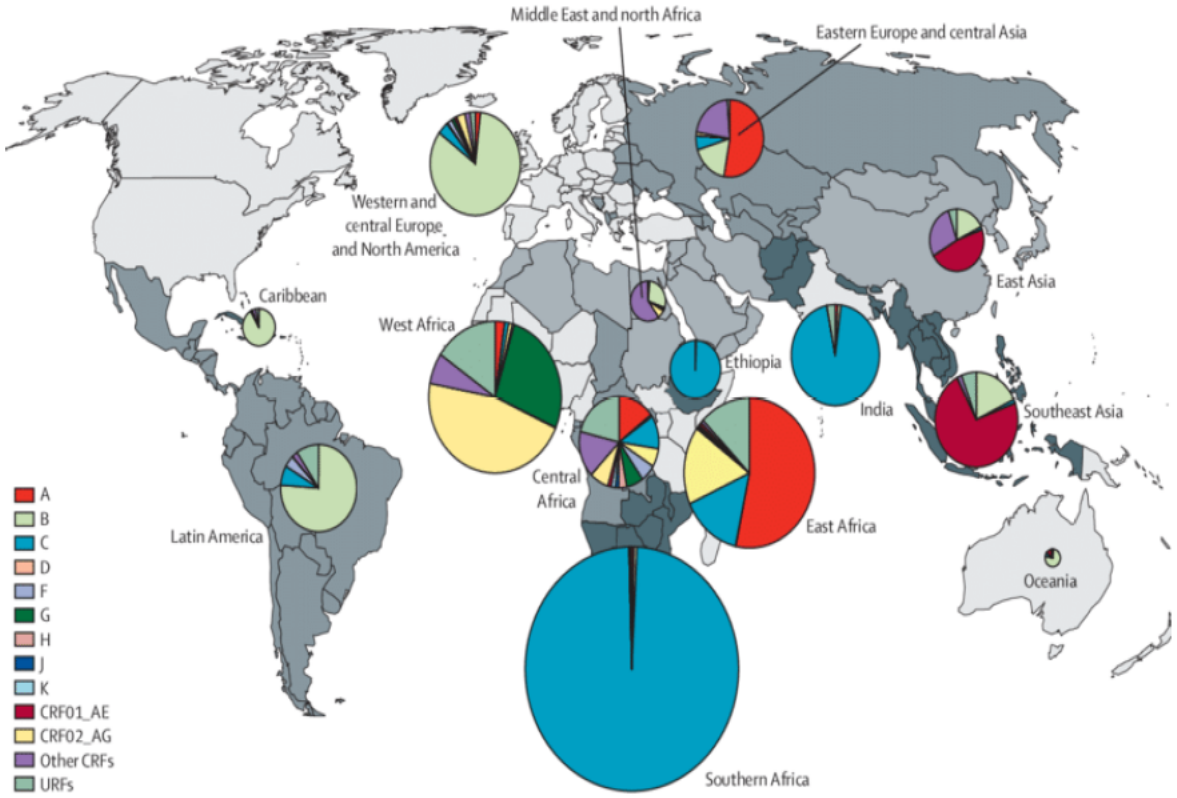


Figure 2: Global distribution of HIV-1 subtypes and the circulating recombinant forms. The HIV-1 subtypes are unevenly distributed across the globe. Each HIV-1 subtype is denoted by a specific colour. Subtype A is represented by a red colour and appears to be dominant in Eastern Europe, Central Asia and Central and Eastern Africa. Subtype B is represented by a light green colour and is dominant in North America, West, and Central Europe. On the other hand, CRF01_AE is shown by a burgundy colour and dominates in Southeast Asia. On a larger scale subtype C represented by a blue colour is dominant in Ethiopia and India however, it is highly concentrated in Sub-Saharan Africa. Subtype C contributes to approximately 50% of the global HIV-1 pandemic (Hemelaar et al., 2019).

On the other hand, HIV-1 subtype C is circulating in Ethiopia, India and predominantly found in Southern Africa (Hemelaar et al., 2019). Interestingly, HIV-1 subtype C accounts for roughly 50% of the worldwide HIV-1 infections (Hemelaar et al., 2019). As a consequence, Sub-Saharan Africa endures a huge burden of the HIV/AIDS pandemic (Bbosa et al., 2019). Therefore, this study will be focusing on HIV-1 subtype C since it is carried out in South Africa.

2.3 The Virology of HIV-1

Human immunodeficiency virus type 1 is classified into the Lentivirus genus, Retroviridae family (Coffin et al., 1986). The viral genome exists with two single stranded ribonucleic acid (RNA) molecules within the virus core. The HIV-1 RNA genome comprises of nine viral genes which will be described in the following sections. The RNA genome is spanned by an incomplete long terminal repeat (LTR) sequence at both the 5' and 3' end (Figure 3) (Montagnier, 1999). Briefly, the LTR of the RNA genome at the 5' end is composed by the U5 (unique, 5′ end) and R (repeat) region but lacks the U3 (unique, 3′ end) region whereas the 3' end LTR is composed by the U3 and R region but lacks the U5 region (Montagnier, 1999). The incomplete LTR structure of the HIV-1 RNA is repaired during a process known as reverse transcription during HIV-1 replication, which will be discussed below.

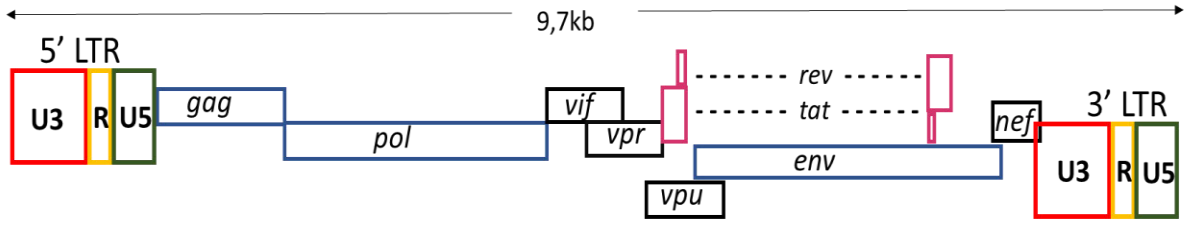


Figure 3. The DNA genomic structure of HIV-1. The HIV-1 DNA genome is approximately 9.7 kb long. It is composed by 9 genes that are flanked by identical LTR (red, yellow, and green) sequences on both the 5' and 3' end. Three structural genes; *gag*, *pol* and *env* (blue) two regulatory genes; *rev* and *tat* (purple) and 4 accessory genes; *vif*, *vpr*, *vpu* and *env* (black). Source: Adapted from (Suzuki et al., 2011).

On the other hand, the full-length HIV-1 DNA is 9.7 kilobase (Fassati and Goff, 2001). As aforementioned, the HIV-1 genome is made up of nine genes. These nine HIV-1 genes are divided into three groups: structural, accessory and regulatory genes (Figure 3) (Freed, 1998). Structural genes comprise of specific antigen (*gag*); polymerase (*pol*) and envelop (*env*) genes (Freed, 1998). The accessory group consists of four genes, viral infectivity factor (*vif*), viral protein R (*vpr*), viral protein U (*vpu* or *vpx* in the case of HIV-2) and the negative regulation factor (*nef*) genes (Gallo et al., 1988). Lastly, the regulatory group is made up by the regulator of virion expression (*rev*) and the trans-activator of transcription (*tat*) genes (Stevens and Miller, 2016).

2.4 The HIV-1 Proteome

Although HIV-1 has only nine genes, it codes for a total of 16 viral proteins (Stevens and Miller, 2016). Briefly, the structural genes; *gag* gene encodes for the polyprotein p55 precursor, which is cleaved to produce four main structural proteins (Blood, 2016). These four main structural proteins encompass the p6 protein and p7 Nucleocapsid protein (NC), the canonical p24 Capsid protein (CA protein) and p17 Matrix protein (MA) (Blood, 2016). The *pol* gene codes for four viral enzymes produced from a precursor protein, Integrase (IN), Reverse transcriptase (RT), Protease (PR) and RNase H (Hill et al., 2005). The *env* gene encodes Glycoprotein 160 (Gp 160) precursor protein, which is cleaved to produce the Gp 120 Surface protein (SU) and Gp 41 Transmembrane protein (TM) (Hill et al., 2005).

The regulator genes; *tat* gene codes for the Trans-activator of transcription (Tat) whereas the *rev* gene codes for the Regulator of expression of virion (Rev) regulatory proteins (Gallo et al., 1988).

Ultimately, the accessory genes; *nef* codes for the Negative regulating factor (Nef), *vif* gene codes for the Viral infectivity factor (Vif), *vpr* gene codes for the Virus protein r (Vpr) in HIV-1. In the context of HIV-2 vpr is replaced with the *vpx* gene coding for the Virion-associated protein. Lastly the *vpu* gene codes for the Virus protein unique (Vpu) (Gallo et al., 1988).

2.5 The HIV-1 Replication cycle

In order for the HIV-1 virus to establish infection, it undergoes a series of consequential steps within a susceptible cell, collectively known as replication cycle (Tozser, 2003). Briefly, the replication cycle includes attachment, entry, uncoating, reverse transcription, nuclear import, integration, gene transcription, assembly, budding and viral maturation (Tozser, 2003). The HIV-1 replication cycle will be briefly discussed in the subsequent paragraphs and the reverse transcription step will be discussed in order to underline the significant role of the 5’ LTR, which is the focus of this study. More details will be given to the transcription step, which is driven by the viral promoter, 5’ LTR.

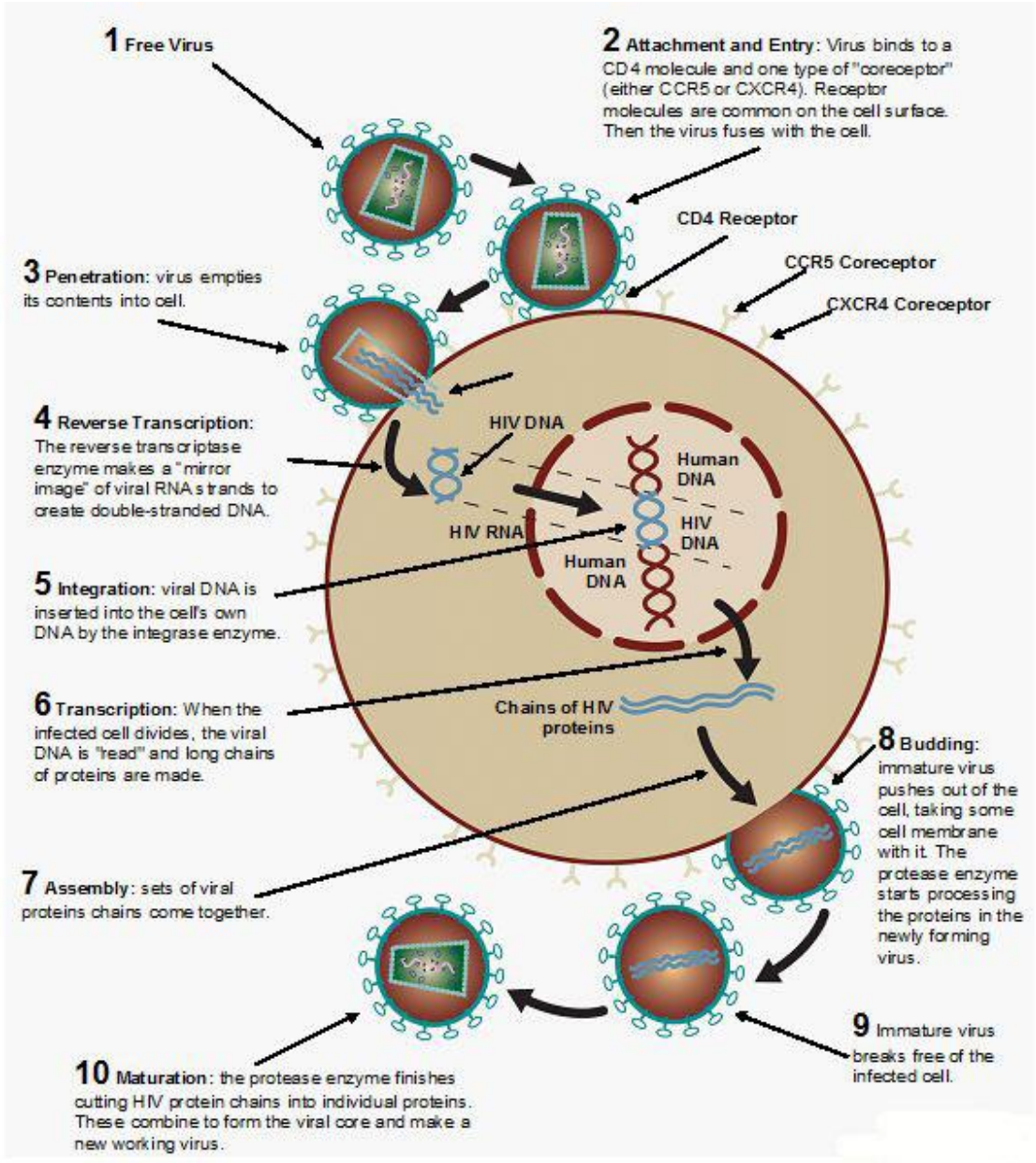


Figure 2.5. The HIV-1 replication cycle. HIV-1 undergoes a series of consequential steps to complete the process of replication. These steps will be discussed in the following section. Taken from (https://www.aidsmap.com/about-hiv/hiv-lifecycle).

2.5.1 Attachment and Entry

The HIV-1 virus first attaches to the hosts cell cluster of differentiation 4 (CD4) receptor of a susceptible cell via its gp120 protein (Ferguson et al., 2002). This interaction triggers a conformational change in the gp120 protein thus exposing a hidden co-receptor binding site, which allows gp120 to bind to the host coreceptor C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4) (Krebs et al., 2001). The CCR5 chemokine receptor is found

on the surface of macrophages, dendritic cells, Langerhans cells and on a subset of helper T lymphocytes (Krebs et al., 2001). On the other hand, the CXCR4 chemokine receptor is commonly found on T cell subsets (Freed, 1998). Binding to the co-receptor further induces additional conformational changes on the gp 41 thus exposing a fusion peptide (Freed, 1998). This promotes the attachment of gp41 fusion protein to the host cell membranes thus pulling the viral envelop closer to the host cell membrane (Gallo et al., 1988). In this way a membrane fusion pore is created on the host membrane and this action facilitates the entry of the viral core into the cytoplasm of the host cell (Chan et al., 1997).(Fassati and Goff, 2001).

2.5.2 Uncoating

Following cell entry, disassembly of the capsid enclosing the HIV-1 core occurs, a process known as uncoating that is poorly understood (Fassati and Goff, 2001). Several models have been proposed as to where does uncoating take place within the host cell. The first model proposes that uncoating takes place immediately following cell entry close to the plasma membrane (Fassati and Goff, 2001). This allows the viral RNA to be available reverse transcription (Fassati and Goff, 2001). The second model proposes that following post-entry the viral capsid remains assembled up until the HIV-1 incoming complexes reach the nuclear membrane (Arhel, 2010). However, it is presumed that uncoating occurs gradually during transport towards the nuclear pore complex (NPC).

Uncoating takes place in the cytoplasm, or at the nuclear envelope, since the viral cores is too large to pass through nuclear pores (Burdick et al., 2020). However, recent evidence indicates that intact viral cores are imported into the nucleus through a mechanism involving interactions with host protein cleavage and polyadenylation specificity factor 6 (CPSF6) (Burdick et al., 2020). In the nucleus, completion of reverse transcription takes place before uncoating, then uncoating is initiated in less than 1.5 h before integration near the genomic integration sites (Burdick et al., 2020, Dharan et al., 2020, Selyutina et al., 2020). The subsequent steps that are discussed below are linked to the first theory of uncoating.

2.5.3 Reverse Transcription

Reverse transcription is driven by the reverse transcriptase (RT) enzyme, with a recent report indicating that cDNA synthesis is completed inside the nucleus (Müller et al., 2021). The viral core contains two identical viral RNA with viral enzymes: reverse transcriptase (RT), integrase (IN) and protease (PR) and other proteins (Fassati and Goff, 2001).

In the course of the microtubule-based transport towards the nucleus, the RT facilitates the synthesis of the complementary double stranded deoxyribonucleic acid (cDNA) (Iordanskiy et al., 2006). Thereafter, the viral RNase H from the RT domain begins to degrade the RNA strands and at the same time the host DNA Polymerase facilitates the synthesis of a complete double stranded DNA (Nermut and Fassati, 2003). This all takes place within the reverse transcription complex (RTC) which mainly consists of the two single stranded HIV-1 RNA genome, IN, Vpr, MA and CA and cellular components such as dNTPs. Once the viral DNA synthesis is completed, the RTC matures into a pre-integration complex (PIC) (Bukrinskaya et al., 1998). In this case the PIC now comprises of the HIV-1 DNA along with the associated viral IN, proteins and cellular proteins such as the Transportin (TRN-SR2) and Lens Epithelium derived Growth factor (LEDGF/p75), which allows tethering of the HIV-1 DNA into the host chromatin (Cherepanov et al., 2003).

2.5.3.1 Construction of the Complete LTR

As aforementioned, the RNA genome has incomplete LTR sequence at either end of the genome (Ooms et al., 2007). The 5'LTR contains only the R and U5 regions located just before the primer binding site (pbs), whereas the 3' LTR end contains only the U3, and R regions located right after the poly-purine rich tract (ppt) (Figure 4) (Zhang et al., 2018). The identical R regions at both the ends of the RNA genome allows the synthesis of the double-stranded viral DNA from the single-stranded viral RNA template (Hu and Hughes, 2012).

The synthesis of a complete LTR happens via a strand transfer mechanism in sequential steps (Figure 4) (Coffin et al., 1997). Briefly, the synthesis of a complementary minus strand DNA is achieved from the plus strand single stranded RNA genome that serve as a template. Reverse transcription is initiated when the tRNA that acts as the primer binds to the complementary primer

binding site (pbs) near the U5 region of the 5' end (Figure 4) (Step 1). Different tRNA primers are used by different retroviruses (Esposito et al., 2012). Some lentiviruses use tRNA1,2 Lys as a primer whereas HIV-1 use the host cell tRNA3 Lys as a primer (Esposito et al., 2012). The cellular tRNA3 Lys has 18 nucleotides that are complementary to the pbs of the plus strand RNA (Esposito et al., 2012). Subsequently, the enzyme RT then transcribes the U5 and R regions of the RNA generating a complementary minus-strand DNA (Figure 4) (Step 2). Following, RNase H degrades the U5 and R regions on the 5' end of the RNA releasing the tRNA primer to jump to the 3' end of the viral genome, thus the newly synthesised DNA strands binds to the complementary R region on the RNA (Figure 4) (step 3). The minus strand DNA synthesis facilitated by the RT is extended in the 3' to 5' direction until the R region of the 5'end of the RNA (Figure 4) (step 4).

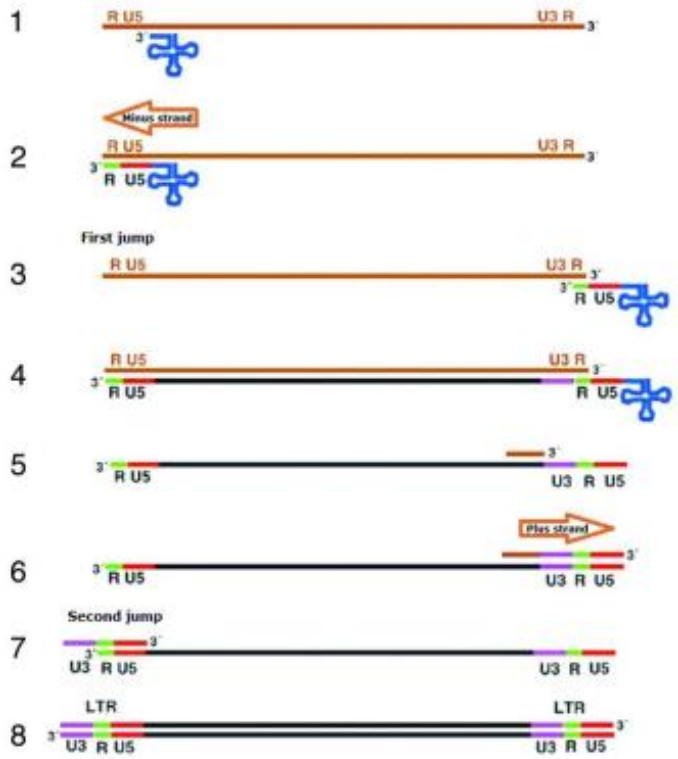


Figure 4. Construction of the complete LTR. Step 1: At the 5' end tRNA base pairs with the pbs next to the U5 region. Step 2: minus strand DNA is synthesized by RT transcribing the U5 and R region. Step 3: Newly synthesized minus strand DNA jumps to the 3' end of the RNA template base paring with the R region. Step 4: Minus DNA strand synthesis continues in the 3' to 5' direction. Step 5: RNase H degrades the plus strand template RNA except for the base paired ppt fragment. Step 6: synthesis of the plus-strand DNA whereby the RT transcribes the U3, R and U5 region of the minus strand template DNA. Step 7: newly synthesized plus-strand DNA jumps to the 3' end of the minus-strand DNA pairing with the R and U5 region. Step 8: synthesis of a complementary double stranded DNA flanked by a complete identical LTR sequence at 5' and 3' end. Taken from (Zhang et al., 2018).

Subsequently the RNase H degrades the plus strand template RNA excluding the RNA fragment that is annealed to the polypurine tract (PPT) of the minus strand near the U3 region in the 5' LTR (Figure 4) (Step 5). The second DNA strand synthesis is initiated using the remaining PPT fragment of viral RNA as a primer in the 5' to 3' direction, transcribing the U3, R and U5 region of the first DNA strand (Figure 4) (Step 6). The short newly second synthesized DNA jumps and binds to the PBS on the first strand (Figure 4) (step 7). The two DNA strands are extended to form a complete double-stranded DNA copy that is flanked by a complete and identical LTR sequence at the 5' and 3' end (Figure 4) (Step 8). The structure and function of the complete LTR sequence will be described in detail below since this is the focus of the study.

2.5.4 DNA Nuclear Import and Integration

The newly synthesised viral DNA as a component of PIC is imported from the cytoplasm into the nucleus through nuclear pore complexes (NPCs) and stably integrates into the host cell genomic DNA (reviewed in (Christ and Debyser, 2013)). Briefly, the NPCs are specialized channels of the nuclear membrane made up of a group of proteins called nucleoporins (Nups), that allows movement across nuclear membrane (Wente and Rout, 2010). This process of nuclear import is mediated by a protein known as Transportin SR2 (TRN-SR2) (Christ et al., 2008). The TRN-SR2 interacts directly with HIV-1 integrase (IN) to drive HIV-1 nuclear entry (Tabasi et al., 2021).

The HIV-1 IN is a 32 kDa protein which is composed of three distinct domains (Tsirkone et al., 2017). This includes the N-terminal domain (NTD), which exists with a Zn^{2+}-binding motif (His-His-Cys-Cys), a catalytic core domain (CCD) where the active site is found and the C-terminal domain (CTD) (Tsirkone et al., 2017). The CCD and CTD forms a dimer, which binds directly to N-terminal half of TRN-SR2 protein through the HEAT repeats (Tsirkone et al., 2017). This interaction mediates the PIC nuclear import into the nucleus, where again the TRN-SR2 interacts with the Ras-related nuclear protein (Ran) which then allows the release of the PIC (Taltynov et al., 2013, Tsirkone et al., 2017). This is accomplished via the Ran GTPase cycle which regulates nuclear import (Tabasi et al., 2021).

Within the cytoplasm, the IN removes the di-nucleotides (Guanine (G) and Thymine (T)) at both 3' ends of the viral double stranded DNA (Fujiwara and Mizuuchi, 1988, Pauza, 1990, Roth et al., 1989). As a result, active 3' hydroxyl (-OH) groups are created which once inside the nucleus causes a nucleophilic attack on the host DNA (Vink et al., 1991, Engelman et al., 1991). The host protein LEDGF/p75 tethers the viral double stranded DNA to the host chromatin (Poeschla, 2008). The nucleophilic attack of the 3'-OH groups on the viral DNA along with the IN allow the stable integration of the viral DNA into the cellular DNA (Cherepanov et al., 2003, Fujiwara and Mizuuchi, 1988). The stably integrated viral DNA is referred to provirus/proviral DNA (Guth and Sodroski, 2014).

2.5.5 Expression of HIV-1 Viral Genes

Following its stable integration, the proviral DNA then serves as a template DNA for transcription of viral genes (Guth and Sodroski, 2014). Transcription is initiated at the 5' LTR and is terminated at the 3' LTR of the viral DNA (Guth and Sodroski, 2014). The viral gene transcription that is controlled by the LTR is explained in more details below since it is the focus of the study.

2.5.5.1 Structure and Function of the HIV-1 LTR

As aforementioned, the LTR consists of 3 regions: the U3, R and U5 regions and is 640 bp long (Figure 5) (reviewed in (Groen and Morris, 2013)). The U3 region is further divided into 3 domains namely the core promoter, core enhancer and modulatory domain (Gaynor, 1992).

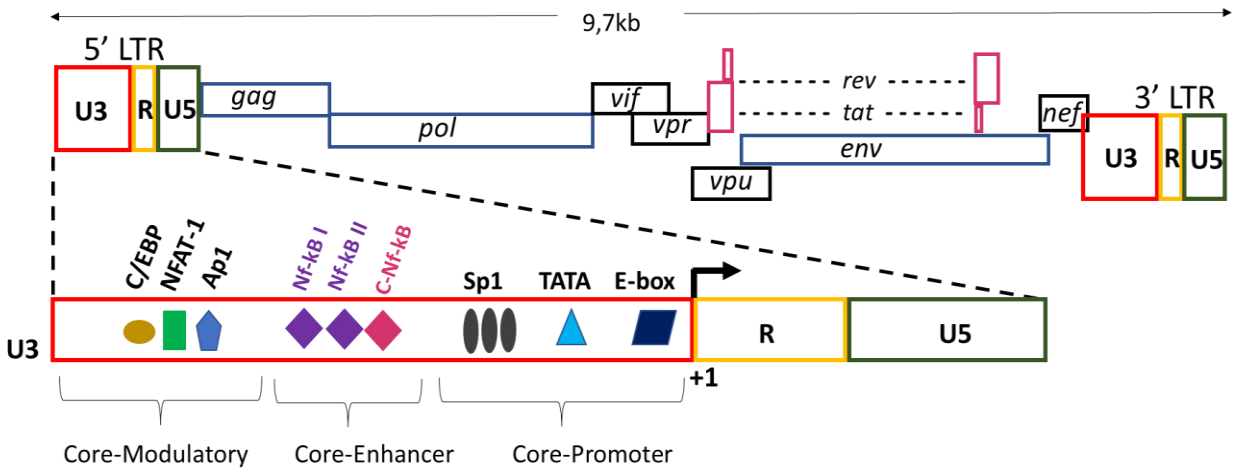


Figure 5. Zooming into the complete structure of the 5' LTR. The HIV-1 DNA genome is 9.7 kb and is flanked by identical LTR sequence at both the 5' and 3' end. The HIV-1 LTR is approximately 640 bp long is divided into 3 regions namely the unique 3 (U3), repeat (R) and unique 5 (U5) regions. The U3 region is further divided into 3 domains namely the core promoter, core enhancer and modulatory domain. The core promoter domain is defined by the presence of a TATA box, E-box, an initiator element, and 3 specificity protein 1 (Sp1) transcription factor binding sites. The core enhancer domain consists of nuclear factor – kappa B (NF-κB) binding sites and the modulatory domain is composed of activator protein 1 (Ap1), nuclear factor of activated T cells (NFAT) and CCAAT enhancer binding protein (C/EBP).

The U3 is populated by a number of transcription factor binding site (TFBS). The core modulatory domain consists of the C/EBP, NFAT-1 and Ap1 TFBS. The core enhancer domain has a subtype-specific number of NF-κB TFBS (Figure 5) (Suzuki et al., 2011). The core promoter domain consists of the initiator element (+1), E-box, TATA box and Sp1 TFBS. Nonetheless, the 5' LTR acts as the viral promoter that drives viral gene transcription, whereas the 3' LTR acts in transcription termination.

2.5.5.2 Function of Host Cellular Factors in HIV-1 LTR

Gene expression of HIV-1 is known to be regulated at the transcriptional level by using the cellular RNA polymerase II (RNAPII) machinery (reviewed in (Dutilleul et al., 2020)). Transcription of HIV-1 is initiated at the U3/R region in the 5'LTR hence it is regulated by cellular transcription factors which interacts to the binding sites identified within the 5' LTR (Colin and Van Lint, 2009).

Overall, the LTR region is spanned by numerous TFBS that are recognized by both constitutive and inducible transcription factors such as Sp1, Ap1, NFAT and NF-κB etc (Suñé and García-Blanco,

1995). These transcription factors have the ability to enhance or repress HIV-1 transcription (Suñé and García-Blanco, 1995). The end results of viral transcriptional activity are dependent on the balance between activators and/or repressors interaction with the 5' LTR (Colin and Van Lint, 2009). Only a few transcription factors have been described as key elements to initiate HIV-1 transcription including NF-κB, Sp1 and the TATA-box whose essential role will be discussed in detail.

Within 5' LTR enhancer region located just upstream of the HIV-1 core promoter binds the NF-κB transcription factor, which has been shown to strongly increase HIV-1 promoter activity (Bachu et al., 2012b). The NF-κB transcription factor is a heterodimer made up of the p65/p50 subunits (Baeuerle and Baltimore, 1988). In an un-infected cell, the p50/p65 is sequestered into the cytoplasm and is inactive due to its interaction with the inhibitory protein IκBα (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha), diminishing its nuclear localization signal (reviewed in (Dutilleul et al., 2020)). However, when the cell is now infected by HIV-1, following MAPK activation or via the PKC signalling pathway, the IκB kinases (IKKα and IKKβ) phosphorylate IkBα to degrade it (reviewed in (Dutilleul et al., 2020)). Then the active heterodimer p50/p65 is imported into the nucleus where it recognizes its binding sites along the host cellular genome and the HIV-1 5'LTR (reviewed in (Dutilleul et al., 2020)). Thereafter, the heterodimer complex is able to recruit histone acetyltransferases (HATs) to promote histone acetylation and therefore HIV-1 transcription (Mbonye and Karn, 2014). The histone acetyltransferase is histone modifier, who's role is to modify the Lysine residue on histone 3 and histone 4 within the chromatin (Cron et al., 2000). Upon the recruitment histone acetyltransferases, the chromatin is modified in such a way that it adopts an open configuration which allows accessibility of the TFBS by other viral and host transcription factors (Cron et al., 2000).

Additional cellular factors allow a more efficient transcriptional activation (Gaynor, 1992). Among these additional factors, the Sp1 is an important cellular transcriptional factor needed to initiate HIV-1 transcription and has been shown to synergically have a protein–protein interaction with NF-κB (Gaynor, 1992). The Sp1 transcription factors translocate from the cytoplasm and physically interacts with the 5' LTR core promoter region through three G-C rich TFBS binding sites which are located downstream of the NF-κB binding sites (Gaynor, 1992). Once the cell is stimulated, the

Sp1 acts synergistically with NF-κB via protein–protein interactions to activate viral transcription (Perkins et al., 1993). In addition, it has been identified that Sp1 and NF-κB both interact with the TATA binding protein (TBP) and some TBP associated-factors (TAFs), such as TAFII250 for NF-κB (through the p65 subunit) and TAF110 for Sp1, that form part of the RNAPII transcriptional factor D (TFIID) (Perkins et al., 1993). In conclusion, NF-κB and Sp1 are the specific transcription factors to recruit the RNAPII-dependent transcriptional machinery to the HIV-1 promoter region (Perkins et al., 1993).

The 5'LTR core promoter also contains a TATA box element, with the sequence TAT/AA (Van Opijnen et al., 2004). It has been shown that the TATA box binds the TATA Binding Protein (TBP) transcription factor, an RNAPII transcription factor TFIID subunit (Van Opijnen et al., 2004). This interaction induces the formation of the transcription pre-initiation complex (Van Opijnen et al., 2004). The transcription factor TFIID is composed of the TATA binding protein (TBP) and multiple TBP associated factors (TAFs) (Gaynor, 1992). The pivotal role of TFIID is to stabilize the platform of the construction of additional members of the transcriptional pre-initiation complex (Zhang et al., 2016). Moreover, an initiator (Inr) element which contains a highly conserved binding site called RBEI known to bind the RBF-2 complex (TFII-I and USF1/2) is found downstream of the TATA box (reviewed in (Dutilleul et al., 2020)). All together these transcription factors promote viral transcription.

Furthermore, another viral protein known as Trans-Activator of Transcription (Tat) plays a vital role in gene expression (Debaisieux et al., 2012). The protein is 86 and 101 amino acids long depending on the subtype (Debaisieux et al., 2012). Tat is a regulatory protein which enhances the efficiency of viral transcription (Debaisieux et al., 2012). It performs this role by binding to the TAR element of the 5'LTR. The detail role of Tat in gene expression will be explained in the subsequent sections.

2.5.5.3 Basal Transcription

Once the HIV-1 DNA is integrated into the host cellular genome, the cellular RNA polymerase II (RNAP II) machinery is hijacked to regulate HIV-1 gene expression at the transcriptional level (reviewed in (Dutilleul et al., 2020)). At the U3/R junction in the 5'LTR HIV-1 transcription is

initiated and is aided by cellular transcription factors which binds in TFBS identified within the 5'LTR (Colin et al., 2014).

Following integration, transcription is initiated at the initiator element (+1) in the R region of the HIV 5' LTR. Solely the 5' LTR has the ability of initiating and driving basal LTR transcription; however at low levels when Tat is absent (Colin et al., 2014). Briefly, initiation of HIV-1 transcription takes place when the RNA Pol II binds to the core promoter region of the 5' LTR inclusive of the TATA box, initiator (Inr) at the transcription start site +1, NF-κB and Sp1 transcription factors (Colin and Van Lint, 2009). The binding of cellular transcription factors to their binding site is enough to aid the recruitment of the transcription pre-initiation complex and thus to initiate HIV-1 transcription (Colin and Van Lint, 2009). The formation of the transcription pre-initiation complex is formed by the binding of TBP which is a TFIID subunit, to the TATA box (Colin et al., 2014). This process is regulated by the interaction between transcription factors such as Sp1 and NF-κB and other TAFs subunits of TFIID (Gaynor, 1992).

Thereafter, the RNA Pol II dependent transcriptional factors (TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH) are recruited, forming part of the pre-initiation complexes which facilitate transcription initiation (Pereira et al., 2000). The RNA Pol II carboxy-terminal domain (CTD) is composed of 52 repeats of a seven amino acids sequence (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) (Pereira et al., 2000). During transcription initiation the RNA Pol II CTD which is part of the pre-initiation complex is phosphorylated by the cyclin-dependent kinase 7 (Cdk7) on serine 5 (Ser5) residues (Gaynor, 1992). The phosphorylation of the RNA Pol II causes the RNA Pol II release from the promoter which supports transcription initiation (Gaynor, 1992).

Basal transcription can only effectively initiate transcription however, transcription elongation is deficient which as a result produces short viral mRNA transcripts (Liu et al., 2015). The short viral mRNA transcripts are a consequence RNA Pol II pausing. The RNA Pol II pausing is due to the binding of the negative elongation factor (NELF) and the DRB (5,6-dichloro-1-β-d-ribofuranosyl benzimidazole) sensitivity-inducing factor (DSIF) pausing factors (Liu et al., 2015). The NELF pauses the transcriptional elongation in the presence of DSIF only, which blocks the RNA Pol II (Liu et al., 2015). The newly synthesized short mRNA transcripts encode for the viral regulatory

proteins Tat and Rev (Daelemans et al., 1999). At the end, the rev and tat spliced mRNA's are then transported to the cytoplasm to be translated into the respective proteins and thus imported back into the nucleus (Daelemans et al., 1999). In the nucleus the Rev protein plays an essential role in the regulation of HIV-1 protein expression, whereby it is involved in the export of unspliced and incompletely spliced mRNAs (Cullen, 2000). Nevertheless, transcriptional elongation is not efficient in the absence of the Tat (Liu et al., 2015).

2.5.5.4 Tat-induced HIV-1 Transcription

The Tat protein promotes the RNA Pol II to free from the HIV-1 promoter to enable transcriptional elongation (Chavali et al., 2019). Briefly, the Tat protein that is synthesized in the cytoplasm gets imported into the nucleus where it plays an essential role as an activator of transcription (Raha et al., 2005). The Tat binds at uridine-rich (U-rich) bulge of the TAR RNA loop through its basic ARM (arginine-rich RNA-binding motif) domain (Chavali et al., 2019). The TAR RNA loop is a short RNA transcript located at the R-region of 5'-LTR (Zhu et al., 1997).The binding of the Tat protein to the TAR element mediates the recruitment of the positive transcription elongation factor (P-TEFb) (Zhu et al., 1997). The P-TEFb complex is made up by the regulatory subunit Cyclin T1 (Cyc T1) and catalytic subunit cyclin-dependent kinase 9 (CDK9) (Zhu et al., 1997). The role of P-TEFb complex is to drive the phosphorylation of the C-terminal domain (CTD) of RNA Pol II required for activation of transcriptional elongation of the HIV-1 mRNA (Kim et al., 2002). A combination of multiple heptad repeats of amino acids (tyrosine 1 – serine 2 – proline 3 –threonine 4 – serine 5 – proline 6 – serine 7) forms the CTD of RNA Pol II (Sikorski and Buratowski, 2009). Phosphorylation takes place on serine 2 which is phosphorylated by the CDK9 subunit and serine 5 and 7 which are phosphorylated by the TFIIH subunit protein (Sikorski and Buratowski, 2009). Moreover, the DSIF and NELF are also phosphorylated by CDK9 which converts the pausing factor into a positive elongating factor, thereby inducing their release from the HIV-1 promoter (Kiernan et al., 1999).

The newly synthesized mRNA transcripts are either un-spliced, fully spliced, or partially spliced (Saltarelli et al., 1996). The HIV mRNA's are grouped into three classes: the un-spliced RNA genome that encodes Gag or Gag-Pol; the partially spliced transcripts that codes for Vif, Vpr, Env and Vpu and fully spliced mRNAs encoding Tat, Rev, and Nef viral proteins (Kiernan et al., 1999).

On the other hand, the un-spliced mRNA's are also packed as viral genome for newly virions (Saltarelli et al., 1996). Nonetheless, the splice or un-spliced mRNA's need to be exported from the nucleus to the cytoplasm for viral protein translation (Kim et al., 2002).

The Rev regulatory protein produced during the early stages of transcription is able to translocate into the nucleus and binds to the Rev response element (RRE) present in the single spliced and unspliced viral mRNA transcripts (Cullen, 2000). This Rev-RRE interaction induces the recruitment of the host chromosomal region maintaince1 (CRM1) and the Ran guanosine-5-triphosphate (Ran-GTP) nuclear export machinery (Cullen, 2000). The association of Crm1/Ran-GTP nuclear export machinery with the Rev-RRE complex forms a Rev-RRECRM1-RanGTP RNA complex which mediates the exports of viral mRNA transcripts from the nucleus through nuclear pore complexes and are released into the cytoplasm (Suhasini and Reddy, 2009). The conversion of Ran-GTP to Ran-GDP allows the dissociation of the complex enabling the release of the mRNAs into the cytoplasm (Suhasini and Reddy, 2009).

Subsequently the following step in transcription is the export of the spliced and un-spliced viral gene transcripts from the nucleus to the cytoplasm (Saltarelli et al., 1996). The fully spliced viral mRNAs, encoding the early proteins Tat, Rev and Nef can travel into the cytoplasm via the conventional cellular mRNA export pathways (Guzman, 2015). Whereas on the other hand, the un-spliced (Gag or Gag-Pol) and the partially spliced (encoding Vif, Vpr, Vpx, Vpu, Env) mRNA require the interaction with the viral protein Rev for export (Guzman, 2015).

2.6 HIV-1 Translation, virion Assembly and Budding

Once exported to the cytoplasm, the capped and polyadenylated viral mRNAs are translated into viral proteins (Guzman, 2015). The structural and enzymatic GAG and GAG-Pol polyproteins RNA are translated to the capsid (CA), matrix (MA), nucleocapsid (NC), protease (PR), reverse transcriptase (RT), integrase (IN) and p6 proteins (Guzman, 2015). Synthesis of Tat, Rev, Nef, Vif and Vpr proteins are established from alternatively spliced mRNAs whereas the Env and Vpu proteins are translated from the same spliced mRNA (Guzman, 2015).

During the late stage of the viral life cycle, HIV-1 Gag assembles into a spherical immature capsid, and undergoes budding, release from the infected cell and final maturation (Guzman, 2015). Assembly takes place at the plasma membrane of the cell to produce spherical membrane-bound virions (Guzman, 2015). Additionally, the viral PR cleaves polyproteins to form matured structural proteins, enzymes, and the viral PR (Gomez and Hope, 2005). The commencement of the virus assembly is when core structural proteins; p17 matrix protein and the p24 capsid protein encapsulate the newly synthesized viral RNA genome with a viral capsid (Guzman, 2015). Viral budding occurs at the cholesterol rich plasma membrane thus immature virions then bud through the host cell after acquiring a viral envelope from viral glycoproteins gp120 and gp41, which are transported to the surface of the cell through cellular secretory mechanisms (Guzman, 2015). From here immature viruses undergo maturation process to form matured virions.

2.7 HIV-1 Infection in the CNS

Following HIV-1 infection, HIV-1 RNA is marginally lower in the CNS than in the plasma of HIV-1 mono-infected individuals, individuals infected with HIV only (Valcour et al., 2012). The CNS compartment has a limited availability of HIV-1 target cells in the CNS, such as the CD4+ T cells and macrophages, whereas the most abundant long-lived microglia and astrocyte cells do not facilitate efficient entry or allow sustained viral replication (Joseph et al., 2015). It is plausible that the lower HIV-1 RNA in the CNS of mono-infected individuals may be a consequence of these long-lived cells. Studies show that HIV-1 infection in microglial cells is latent if rather that they meet the two criteria of a true latency reservoir which describes; 1) the presence of HIV-1 integrated DNA in long lived cells and, 2) the existence of mechanisms which allow the virus to persist for long period in latent cells ((reviewed in (Wallet et al., 2019, Eisele and Siliciano, 2012)). On the other hand, the astrocyte is the most abundant cell type in the CNS however, recent studies showed that HIV-1 is not detected able, and that the infection appears to be non-productive in the Astrocytes (Churchill et al., 2009) (Gorry et al., 2003).

HIV-1 enters the CNS within two weeks of infection (Rojas-Celis et al., 2019). There are several theories that have been proposed as to how HIV-1 crosses the blood brain barrier (BBB) and enters the CNS thus infecting the cells of the CNS. One is the “Trojan horse” theory, which proposes that infected monocytes or CD4+T lymphocyte are able to migrate through the endothelial cells making

up the BBB and enter the CNS (Spudich and González-Scarano, 2012). Another theory proposes that free floating HIV-1 viral particles can travel freely to the CNS by transcytosis, which is mediated by the viral envelope protein gp120 (Banks et al., 2001). This mechanism does also take place when the BBB is exposed to an increase in proinflammatory cytokines (Lafortune et al., 1996).

2.7.1 CNS Tissue Specific HIV-1 Replication

In addition to T cells, HIV-1 infects other cell types such as macrophages that may contribute to the viral reservoir and spread into the CNS by crossing the BBB (Karris and Smith, 2011). On the other hand, the CNS compartment has fewer CD4+ T cells, which are primary HIV-1 target cells while macrophage and microglia cells are the most abundant in this compartment (Joseph et al., 2015).

Once the virus enters the CNS, it then subsequently infects the cells of the CNS, which include the long-lived microglial cells (Joseph et al., 2015). Microglial cells are the resident tissue macrophages of the CNS and plays a crucial role to the innate immune system (Rio-Hortega, 1939). As aforementioned, in addition to CD4 receptor, HIV-1 entry into the host cell requires the presence of either CXCR4 or CCR5 coreceptor (Joseph et al., 2015). However, CCR3 and CCR5 co-receptors are required for efficient infection of the CNS cells. Microglial cells express both the CCR3 and CCR5 co-receptors on their surface, making these cells to be susceptible to HIV-1 infection and support productive HIV-1 replication. (Joseph et al., 2015).

Perivascular macrophages are also the resident cells and part of the immune system in the brain (Konsman et al., 2007). The perivascular macrophages transmit and modulate peripheral inflammatory signals between the immune system and the brain (Konsman et al., 2007). In addition, perivascular macrophages express CXCR4 and CCR5 co-receptors making them to be susceptible to HIV-1 infection (Konsman et al., 2007).

The role of the neuron's is to process and transmit information but it remains controversial whether HIV infect these or not. (Cantó-Nogués et al., 2005) (Ref 1997). (Rottman et al., 1997). However,

the presence of these HIV-1 co-receptors on the surface of neurons suggest that there is a possibility that HIV-1 infection can be established in these cells (Rottman et al., 1997).

Astrocytes cells are also another cell type population of the CNS (Liddelow and Barres, 2015). These cells are the most abundant glial primary cells of the CNS (Liddelow and Barres, 2015). They play an essential role in the homeostasis and provide structural support of neurons (Liddelow and Barres, 2015). The astrocyte cells do not express the well characterised HIV-1 receptor such as CD4+ and co receptors such as CXCR4 or CCR5 on their cell surface which hinders HIV-1 infection (Churchill et al., 2009). Nonetheless, viral proteins and nucleic acids can be detected in cultured astrocyte cells (Churchill et al., 2009). In addition, Li *et al.* reported that in vitro experiments show that Astrocyte cells are able to be infected through a cell-to-cell contact HIV-1 transmission from the surrounding cells (Li et al., 2015).

2.7.2 HIV/TB co-infection in the CNS

A major public health issue has been posed by the HIV-1-associated opportunistic infections of the CNS. Tuberculosis (TB) is the most common opportunistic infection in HIV-1 infected individuals in sub-Saharan Africa. The co-infection of HIV-1 and TB has been reported to exacerbates both diseases. As aforementioned, Sub-Saharan Africa is the region with the greatest burden of the HIV/TB syndemic, and accounts for approximately 70% of AIDS deaths worldwide (WHO, 2020). Individual's that are mono-infected with HIV-1 show relatively lower CNS RNA load compared to the peripheral compartment. But a paradoxical event takes place when individuals are co-infected with Mycobacterium tuberculosis (M. tb), where studies show that the CNS of co-infected individuals have higher HIV-1 RNA load compared to CNS of non-TBM individuals (Seipone et al., 2018). However, the underlying mechanism is poorly understood.

2.8 Impact of 5'LTR Genetic Variation on HIV-1 Pathogenesis

A previous study has reported that inter- and intra-subtype specific genetic variations exists across the LTR (Bachu et al., 2012c). The majority of HIV-1 subtypes including the prototype HIV-1 subtype B contain two NF-κB binding sites (Jeeninga et al., 2000). On the other hand, recombinant

CRF01_AEhas one NF-κB binding site (Montano et al., 1998). One trait that best differentiates HIV-1 subtype C strains from most HIV-1 subtypes is the presence of an additional NF-κB (Bosso et al., 2021). The majority of subtype C viruses contain three NF-κB sites while other studies have reported that some of the subtype C strains have gained a fourth NF-κB (Bachu et al., 2012b, Boullosa et al., 2014). Previous studies reported that subtype C strains with three to four NF-κB binding were circulating in other global regions including India, China, Mozambique, Brazil, and South Africa over the past decade (Bachu et al., 2012c, Boullosa et al., 2014, Bbosa et al., 2019). Moreover, the subtype C viral strains were also gradually expanding and has taken over the epidemic in India (Bachu et al., 2012c).

The TATA box, which is in the core-promoter domain of the U3 region of the LTR has the TATAA canonical sequence (Montano et al., 1998). However, some studies have also reported that genetic variation exists within the TATA box. Specifically, the TATA box sequence in recombinant CRF01_AE viruses is mutated to contain a TAAAA sequence instead of the TATAA canonical sequence described for all other subtypes (Montano et al., 1998). The TATAA box is flanked by E-Box elements, which have been shown to contain genetic variation (Baar et al., 2000). For example, Baar *et al*. showed that subtype G have mutation on E box (i) from T to A at position 5, which occurs together with mutation on E box (ii) from T to C at the same position (Baar et al., 2000). Furthermore, the core promoter domain of the U3 region of the LTR contains three Sp1 binging sites (reviewed in (Groen and Morris, 2013)). A previous study have also reported variation within the Sp1 binding sites (Qu et al., 2016). Specifically, Qu *et al*. showed that the Chinese subtype B 5'-LTR exhibits a mutation from a Cytosine (C) to Thymine (T) at position 5 of the Sp1III TFBS (Qu et al., 2016).

The low fidelity of error prone RT enzyme and host immune pressures results in genetic variation within the HIV-1 genome, including the LTR thus creating viral quasi species (reviewed in (Sampathkumar et al., 2012)). There is increasing evidence that genetic variation of the transcription binding sites may result to differential LTR transcriptional activities (Bachu et al., 2012). This may subsequently affect the HIV-1 replication rate (Mbondji-Wonje et al, 2018). Interestingly, the data from Bachu *et al*. demonstrated that the HIV-1C viruses containing four NF-κB binding sites are associated with stronger LTR transcriptional activity, improved replicative fitness and significantly higher viral loads in plasma compared to the standard subtype C virus that contains three NF-κB

(Bachu et al., 2012c). A different study also showed that genetic variation within the Sp1III site translates to differential transcriptional activity (Montano et al., 1998, Nonnemacher et al., 2004, Gray et al., 2016). Specifically, Nonnemacher *et al*. reported that a change from a C to T at position 5 has be studied to lead to an increase in disease progression (Nonnemacher et al., 2004).

Most studies have reported on HIV-1 mono infection, where they observe that mono-infected individuals exhibit marginally lower HIV replication in the CNS compared to the periphery. However, the effect of HIV-1 LTR genetic variation has not been reported in the TBM co-infection that may result in paradoxically higher viral replication in the CNS compared to the periphery. Therefore, understanding the effects of HIV-1 LTR genetic variation on viral gene transcription will enhance our understanding of HIV-1 replication in the CNS of TBM co-infected individuals.

CHAPTER 3: METHODOLOGY

3.1 Study design, Inclusion and Exclusion Criteria

The study participants were made available from a larger prospective cohort of tuberculosis meningitis (TBM) patients (Patel et al., 2013). Briefly, 235 consecutive patients with suspected meningitis were prospectively recruited between 1 January 2008 and 31 December 2011. Patients with a meningitis illness who were referred from local district general hospitals were investigated at the tertiary hospital at Nkosi Albert Luthuli Central Hospital and included in the study.

All patients in this cohort that were clinically assessed for tuberculosis meningitis underwent; a computerised tomography (CT) scan to exclude contraindications to a lumbar puncture, and had blood drawn for routine tests including a serum fluorescent treponemal antibody test, a venereal disease research laboratory test, a human immunodeficiency virus (HIV) enzyme-linked immunosorbent assay, and a CD4 count. Approximately 15 ml of cerebrospinal fluid (CSF), obtained by lumbar puncture, was processed for the following tests: microscopy (Gram stain and fluorescent staining for acid-fast bacilli [auramine]); bacterial culture; Mycobacterium tuberculosis (M.tb.) culture (Bactec MGIT 960; BD); fungal culture; cryptococcal latex agglutination test; Roche Amplicor Mycobacterium Tuberculosis PCR Test (Roche Diagnostic Systems) (Amplicor PCR); routine chemistry (protein, glucose, chloride); viral Polymerase chain reaction (PCR) for cytomegalovirus, varicella zoster virus, and herpes simplex; venereal disease research laboratory test; fluorescent treponemal antibody test; and test for cysticercus antibodies. An uncentrifuged specimen and, volume permitting, a centrifuged sample of CSF was biobanked using Xpert MTB/ resistance to rifampin (RIF) analysis.

Subsequently, patients were categorized into three groups from the test results obtained; n = 59 definite TBM (positive CSF M.tb culture), n = 64 probable-TBM (did not meet the standards for being definite TBM) and n = 81 non-TBM (CSF M.tb. culture negative) (Figure 3.1). Patients who had positive CSF M.tb culture and/or Amplicor PCR fell into the definite TBM group, whereas the probable-TBM group were patients that did not meet the standards for being definite TBM but were treated empirically with anti-TB drugs. On the other hand, the non-TBM group were patients with

a negative CSF M.tb. culture and had an additional confirmatory alternate diagnosis and responded to therapy in the absence of anti-TB treatment.

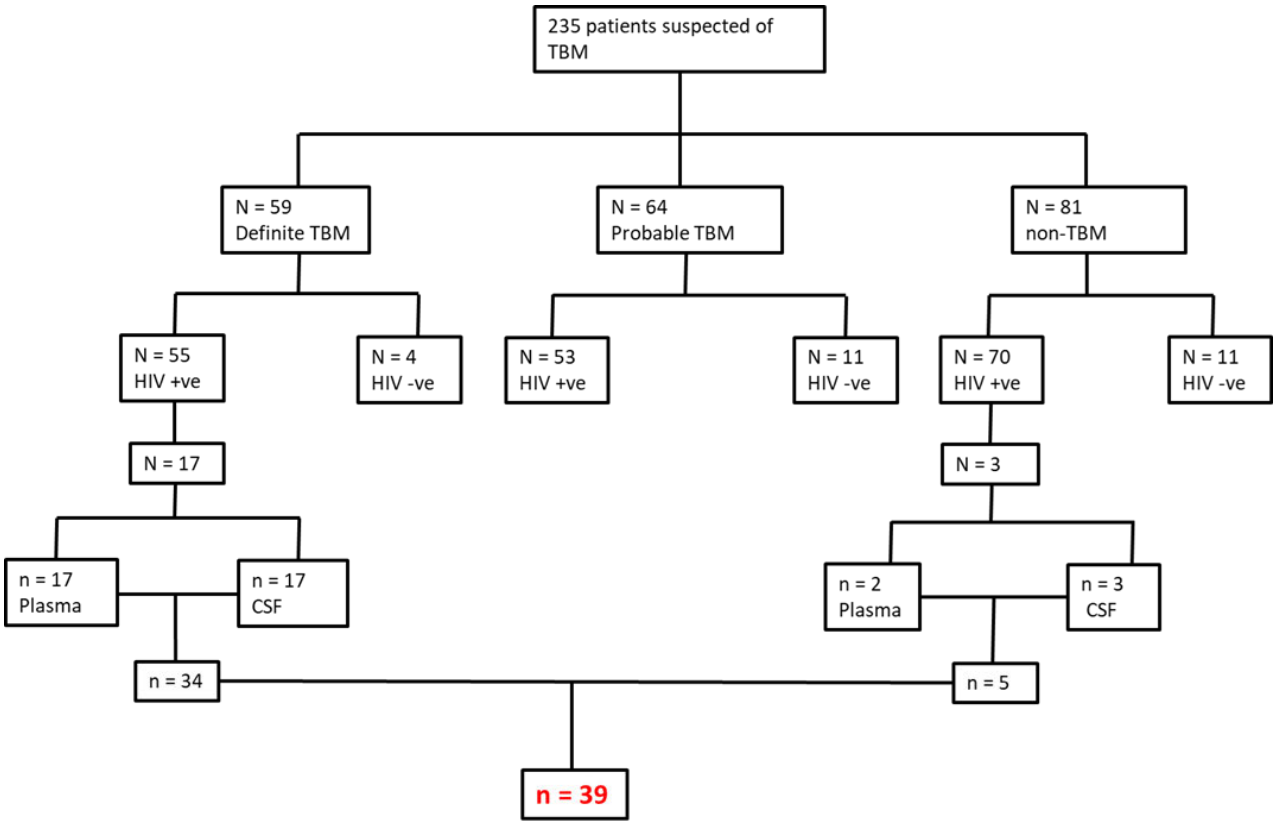


Figure 3.1: Cohort design and patient selection. The cohort began with 235 TBM co-infected treatment naïve patients. After clinical assessment 59 had definite TBM, 64 probable TBM, and 81 non-TBM and patients were further grouped into HIV positive (+) and HIV negative (-) groups. From the HIV/TBM+ group 17 patient samples of 17 matching CSF and plasma making a total of 34 were used for this study. From the HIV+/non-TBM group only 3 samples were used, however one samples did not have a matching plasma. N represents the number of patients and n represents the number of samples (Patel et al., 2013).

Samples obtained from CSF and plasma compartments of 20 human immunodeficiency virus (HIV) chronically infected, antiretroviral therapy (ART) naïve participants were made available, for the purpose of this study. The TBM samples were obtained from a total of 17 TBM patients, whilst non-TBM samples were obtained from 3 non-TBM patients. One participant from the non-TBM group did not have a matching plasma resulting in a total of 39 (Figure 3.1).

3.1.1 Baseline Demographic and Clinical Characteristics

The demographics of the study participants selected for this study included age and sex (Table 3.1). The TBM group had a median age of 28 years with an interquartile range [IQR] of 20 – 40 years, while 53% of the participants were males. The non-TBM group had a median age of 34 years with an IQR of 29-38 years, while 33 % of study participants were males.

The TBM and non-TBM group were matched for plasma CD4 count and viral loads with a median plasma CD4 count of 117 cells/µl [IQR, 14 – 677] and 113 cells/µl [IQR, 36 – 532] respectively (Table 3.1). The median plasma viral load for the TBM group was 390 650 copies/ml [IQR, 289 217 – 1 710 603] while it was 102 025 copies/ml [IQR, 4317 – 213 260] for the non-TBM group. The median CSF viral load of the TBM group was significantly higher 630 291 copies/ml [IQR, 8133 – 10 000 000] compared to the non-TBM group 10 631 copies/ml [IQR, 5109 – 14 230]) (p = 0.020). Furthermore, the TBM group had a higher CSF viral load 630 291 copies/ml [IQR, 8133 – 10 000 000] compared to the plasma viral load 390 650 copies/ml [IQR, 289 217 – 1 710 603]. On the other hand, for the same groups the median lymphocyte count in the CSF were 6 cells/µl [IQR, 2 – 122] and 2 cells/µl [IQR, 0.24 – 6] respectively.

Table 3.1: Demographic and Clinical Characteristics of the study participants

	Characteristics	*TBM*	*Non-TBM*	*p-value*
Demographics	**No of study participants**	**17**	**3**	**N/A**
	Gender, male (%)	**9/17 (53)**	**1/3 (33)**	**0.006#**
	Median Age, years (IQR)	**28 (20-40)**	**34 (29-38)**	**0.221**
Clinical	**Median Plasma Viral Load copies/ml (IQR)**	**390650 (IQR, 289217-1710603)**	**102 025 (IQR, 4317-213260)**	**0.169**
	Median CSF Viral Load copies/ml (IQR)	**630291 (IQR, 8133-10000000)**	**10631 (IQR,5109-14230)**	**0.020#**
	Median peripheral CD4 counts cells/µl (IQR)	**117 (IQR, 14-677)**	**113 (IQR, 36-532)**	**0.695**
	Median CSF CD4 counts cells/µl (IQR)	**6 (IQR, 2-112)**	**2 (IQR, 0.24-6)**	**0.112**

IQR: Interquartile range. Significant differences between the groups are shown in bold p values #Fisher's exact test (Taken from (Ramruthan, 2018)

3.2 Ethics

Approval for this study was obtained from the Biomedical Research and Ethics Committee of the University of Kwa-Zulu Natal (REF: E325/05). Written informed consent was obtained from the study participants and the patients were anonymous and recorded as patient identity (PID). The study was carried out at the HIV Pathogenesis Programme (HPP) laboratories at the University of KwaZulu Natal's Nelson R. Mandela School of Medicine.

3.3 Sequencing of the CSF and plasma derived HIV-1 subtype C LTR elements from the TBM and non-TBM cohort PLHIV.

3.3.1 Complementary deoxyribonucleic acid (cDNA) synthesis

The HIV-1 ribonucleic acid (RNA) was extracted by a previous student (Ramruthan, 2018). Continuously, Complementary DNA (cDNA) Synthesis was performed from HIV-1 RNA previously extracted from matching CSF and plasma obtained from 17 TBM and 3 non-TBM groups however, 1 CSF sample did not have a matching plasma from the non-TBM group, making a total of 39 samples. The extracted viral RNA was reverse transcribed to complementary DNA (cDNA) utilising SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (ThermoFisher Scientific, Invitrogen™, USA), as per manufacturer's instructions. Briefly, a reaction mix comprising of 1X reaction Buffer, 0.1 micromolar (µM) sense primer (5'-U3-M-F:5'CAGGTACCTTTAAGACCAATGAC-3' nt 9013 to nt 9035) and anti-sense primer (3'-R-M-R: 5'- ACTYAAGGCAAGCTTTATTGAG-3' NT 9680 to nt 9609) LTR primers, 1 unit (U) SuperScript™ III RT/Platinum™ Taq Mix, 1 microliters (µl) of 5 nanogram (ng)/µl CSF or plasma derived RNA and PCR grade water to make it to a final volume of 25 µl was prepared. The reaction mixture was then subjected to one step cDNA synthesis and PCR amplification conditions, which were 55 (degree Celsius) °C for 30 min and 94°C for 2 min pre-incubation, followed by 40 cycles of denaturation at 94°C for 15 sec, annealing at 55°C for 30 sec, extension at 68°C for 2 min, final extension step at 68°C for 5 min) and hold at 8°C. The resulting first round PCR (PCR1) products was used for nested PCR, which is discussed in the next section.

3.3.2 Nested polymerase chain reaction (PCR)

The first round PCR product was used as template for nested PCR to amplify the HIV-1 3' LTR U3/R region using KAPA HiFi HotStart PCR Kits (ThermoFisher Scientific, Invitrogen™, USA) as per the manufacturer's instructions. Briefly, a 25 µl reaction mixture consisted of 1X High Fidelity PCR Buffer, 0.3 millimolar (mM) KAPA dNTP mix; 0.3 µM forward primer (5'T7LTRKpn15'taatacgactcactatagggTTGGTACCTTTAAAAGAAAAGGGGGGAC-3'_nt 9064 to 9085) and 0.3 µM reverse primer (3'Sp6LTRHindIII5'-atttaggtgacactatagAGCTTTATTGAGGCAAGCTTAGTGGGTT-3'_nt9622-9593), 1 U KAPA HiFi HotStart enzyme, 5 ng of the PCR1 template and the volume was made up to a final volume with PCR grade water. The thermocycler (Bio-Rad T100 Thermal Cycler, United States) conditions were set to initial denaturation 94°C for 5 min, followed by 25 cycles of (94°C for 15 sec for the denaturation phase, 60°C for 30 sec for the annealing phase, the extension phase at 72°C for 1 min, final extension at 72°C for 5 min) and ultimately the indefinite hold at 8 °C. The specificity and the size of the second round PCR (PCR2) products which is approximately 640 bp was verified on a 1% Agarose gel tracked with GelRed at 95 volt (V) for 45 min, compared against GeneRuler™ 1 Kb (ThermoFisher scientific™, Lithuania, EU) and visualized under ultraviolet light (UV light). The PCR2 product was quantified by NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Delaware, USA), and stored at -20°C freezer until further use.

3.3.2.1 Polymerase chain reaction (PCR) clean-up

The PCR2 products clean-up was performed using QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA, USA) as per manufactures instructions. Briefly, the 20 µl of the PCR2 product was mixed with 100 µl of Buffer PB, then the mixture was transferred into a QIAquick spin column collection tube and centrifuged at 21767 x g (gravity) for 1 min. The filtrate in the collection tube was discarded and 750 µl of Wash Buffer PE was added to the spin column and thereafter centrifuged for 1 min at 21767 x g. An extra spin was performed to remove residual wash Buffer. Thereafter, 30 µl of 10 mM elution Buffer (EB) (pH 8.5) was used to elute the purified DNA into a sterile 1.5 ml tube. The yield of the purified DNA was quantified using the NanoDrop™ 2000 spectrophotometer and stored at the -20°C freezer until further use.

3.3.3 Sequencing

Following clean-up of the PCR2 products, CNS and plasma derived HIV-1 3' LTR U3/R region was directly sequenced using Big Dye Terminator v3.1 Sequencing Kit (Thermofisher Scientific, Massachusetts, United States) as per the manufactures instructions. Briefly, two separate sequencing reaction mixes were prepared for each sample one with a forward primer and the other one with a reverse primer. The sequencing mix was made of 1X Big Dye sequencing Buffer, 3.2 picomolar (pM) forward primer (5'T7LTRKpn1 5'taatacgactcactatagggTTGGTACCTTTAAAAGAAAAGGGGGGAC-3'_nt 9064 to 9085) or pM reverse primer (3'Sp6LTRHindIII 5'-atttaggtgacactatagAGCTTTATTGAGGCAAGCTTAGTGGGTT- 3'_nt 9622 to 9593), which contains the T7 promoter sequence (in small letters), 0.4 µl of BigDye v3, 20 ng of either CSF or plasma patient derived PCR2 template and was made to a 10 µl reaction volume using PCR grade water. The reaction mix was aliquoted into the corresponding well of a 96-well plate (Applied Biosystems). Subsequently, the sequencing reaction was performed in the thermocycler under the following thermocycling conditions: pre-denaturation at 96˚C for 1 min, followed up by 25 cycles of (denaturation at 96˚C for 10 sec, annealing at 50˚C for 5 sec, extension at 60˚C for 4 min) and ultimately the indefinite hold at 4˚C.

3.3.4 Sequence Clean-up

Immediately after sequencing, 1 µl of 125 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0) was added to each well of the 96-well plate containing the sequencing reaction product. This was followed by an addition of 26 µl of 3M Sodium Acetate solution (NaOAc) (pH 5.2) (150 µl of NaOAc into 3750 µl 100% ethanol) to precipitate the DNA. The 96-well plate was then sealed with a sticky plastic cover, vortex briefly and incubated at room temperature (RT) for 10 min. Subsequently, the 96-well plate was centrifuged (Eppendorf centrifuge 5810R, Merck, Germany) at 2061 x g for 20 min. To remove the supernatant, the 96- well plate was inverted onto a paper towel thus centrifuged for 5 min at 150 x g. Next, 35 µl of 70 % ethanol was added to each well and the 96-well plate which was subsequently sealed and centrifuged at 1159 x g for 5 min. Thereafter, the plate was inverted on a paper towel and centrifuged at 150 x g for 1 min. Next the paper towel was removed, and the 96-well plate was dried in the thermocycler at 50˚C for 5 min. The plate was then sealed with an adhesive foil, labelled, and stored at the -20˚C freezer prior to analysis in the ABI 3130xl Genetic Analyzer (Applied Biosystems).

3.3.5 Sequence analysis

The reverse and forward primer sequences of HIV-1 3'LTR U3/R region were cleaned, edited, and assembled together to form contigs using Sequencher 5.1 software (Gene Codes, Ann Arbor, USA, Michigan). Subsequently, a multiple sequence alignment was created using Jalview v2.10.5 (Waterhouse et al., 2009). The sequences were then examined on Bioedit Sequence Alignment Editor v7.2.5 (Hall, 1999) for mutations within the patient derived LTR U3/R sequences. With these sequences a phylogenetic tree was constructed using the online tool Phyml (https://www.hiv.lanl.gov/content/sequence/PHYML/interface.html). The examined patients in this study were infected with HIV-1 subtype C viruses and thus the phylogenetic tree was rerooted against the Indian subtype C reference sequence (EF17 8613.1). The phylogenetic tree was then viewed and edited using FigTree Tree Figure Drawing Tool v1.4.3.

3.4 Cloning

3.4.1 Second round polymerase chain reaction (PCR2) products and pGL3 plasmid Restriction Digest and Ligation

The purified PCR2 products were cloned into the pGL3 Basic vector (Promega, USA, Madison). The pGL3 plasmid was chosen for cloning because it contains the firefly luciferase (luc) reporter gene, which is under the control of the promoter where the LTR containing the promoter will be ligated (Gray et al., 2013). Briefly, the pGL3 Basic vector underwent restriction digestion with the two sticky ends restriction enzymes, Kpn1 and HindIII (New England Biolabs, USA, Massachusetts). The restriction digestion master mix was composed of 10000 ng of circular pGL3 plasmid, 1 U of each restriction enzyme, 1 X CutSmart Buffer and was made to a 30 µl final volume using nuclease-free water.

The reaction was then placed in the thermocycler at 37 °C for 2 hours (hrs) and then heat inactivated at 65 °C for 20 min. The entire 30 µl restriction digestion reaction was then run on a 1% agarose gel for 1 hr at 95V using a 1 Kb Plus DNA Ladder (Invitrogen, Thermofisher Scientific, Massachusetts, United States). The restriction digestion controls including 1 µl of the undigested and digested plasmid were included for tracking our band or fragment of interest. Once the gel was done running,

the gel portion with the 1 µl of undigested and digested samples was cut out of the gel and viewed under UV and immediately the band corresponding to the band of interest was excised using a gel extractor (Merck, Germany, Darmstadt). The other gel portion with 30 µl restriction digestion reaction was not exposed to ultraviolet light to protect the integrity of the DNA from UV light, which mutates DNA. The two gel pieces were then placed back together and the position of the excised gel band on the confirmation gel portion was used to make a blind excision of the band of interest from the gel piece with 30 µl restriction digestion reaction. Prior to DNA gel extraction, an empty 1.5 ml Eppendorf was pre-weighed to get the weight of the empty 1.5 ml Eppendorf. Following, the extracted PCR bands or 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 mass 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 establish 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. Briefly, a 1:1 ratio volume of binding buffer to the gel slice (volume: weight) was used and then this gel mixture was incubated at 50-60 °C for 10 min to melt the gel. Thereafter, 800 µl of the solubilized gel solution was transferred into the GeneJET purification column and centrifuge for 1 min 21767 x g. The flow-through was discarded and the same collection tube was re-used. The DNA in the column was washed with 700 µl of wash buffer and centrifuged for 1 min. The resulting flow-through was discarded and the empty GeneJET purification column had an additional spin for 1 min to completely remove residual wash buffer. The GeneJET purification column was transferred into a clean 1.5 ml microcentrifuge tube and 30 µl of elution buffer was added to the centre of the purification column membrane without touching the membrane and centrifuge for 1 min. The eluted purified plasmid DNA was then quantified using the NanoDrop® 2000.

The PCR2 products of patient derived HIV-1 LTR U3/R region were digested with the same sticky end restriction enzymes (Kpn1and Hind III) as described above to create the same sticky ends. The restriction digestion master mix was prepared as mentioned above for the plasmid DNA; however, 150 ng of the PCR2 was digested. The PCR2 products were used directly from restriction digestion. The digested CSF or plasma derived LTR U3/R region was cloned into the linearized pGL3 Basic

vector by T4 DNA Ligase (New England Biolabs, Massachusetts) as per the manufacturer's instructions. Briefly, a recombinant pGL3 Basic vector was established 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, 1U of T4 DNA ligase with the reaction volume being made up to 20 µl using nuclease-free water. The ligation reaction was then placed in the cycler for 10 min at 25 °C, 20 min at 65 °C and thereafter returned to ice for the subsequent transformation reaction.

3.4.2 Transformation of HIV-1 LTR-pGL3 plasmid into JM109 E. coli competent cells.

The recombinant pGL3 plasmid was then transformed into JM109 *E. coli* (Promega, USA, Madison) cells as per the manufacturer's instructions. Briefly, frozen JM109 *E. coli* cells contained in a tube were thawed on ice and 10 µl of the recombinant plasmid was subsequently added to 50 µl of the thawed JM109 competent *E. coli* cells and then immediately incubated on ice for 30 min. The reaction was thereafter heat-shocked at 42 °C for 45-50 sec in the water bath without shaking and placed back into ice for 2 min. Thereupon, 900 µl of the Super Optimal with Catabolite (SOC) recovery medium was added to the reaction tube followed by incubation at 37 °C while shaking at 250 revolutions per minute (rpm) for 1 hr. Next, 100 µl of the transformation reaction was spread over the surface of the already prepared 100 microgram (µg) / millilitre's (ml) ampicillin containing agar plate with subsequent incubation at 37 °C for 16-20 hours. The selection marker considered was the ampicillin resistant gene contained in the pGL3 plasmid.

3.4.3 Colony Polymerase chain reaction (PCR)

Single colonies were picked from the plate on the following day for colony PCR to confirm that cloning was successful. Briefly, colonies from ampicillin plates were picked using a pipette tip. The tip used to pick the colony was first touched once on a master plate and subsequently the tip was placed in 10 µl of nuclease-free water in a 96-well plate. The clones were mixed with water by pipetting up and down and the diluted plasmid clones were boiled at 95°C for 10 min. The colony lysate was used as the DNA template for PCR reaction, which was prepared as described above for nested PCR. Briefly, the boiled plasmid clones were used as template DNA. Exactly 9 µl of the master mix reaction was aliquoted into PCR tubes and 1 µl of the template DNA was added and the reaction was placed in the thermocycler under the same nested PCR conditions mentioned above to

amplify the patient derived insert of interests. The patient derived insert was confirmed by the presence of the band at approximately 640 bp on a 1 % agarose gel stained with GelRed and visualized under ultraviolet light. Positive clones were further inoculated into a 250 ml of Luria broth (LB), containing 250 µl of 100 µg/ml ampicillin. The inoculum was then incubated vertically at 37°C in a 230 rpm shaking incubator overnight.

3.4.4 Plasmid Maxiprep

The overnight LB broth cultures that had changed to a cloudy appearance confirming bacterial growth were taken out of the incubator and placed in the 4 °C fridge for 2 hrs. Thereafter the cultures were transferred into 50 ml falcon tubes and centrifuged (Heraeus Biofuge Fresco, England, United Kingdom) at 4000 x g for 30 min at 4°C. The resulting supernatant was discarded. Following this, the plasmid DNA clones were harvested by Qiagen Plasmid Maxiprep Kit (ThermoFisher Scientific, USA) as per the manufacturer's instructions. Briefly, the resulting pellet from spinning was resuspended in 10 ml of Buffer P1.

Immediately after, 10 ml of Buffer P2 was added, mixed vigorously, and incubated at room temperature (RT) for 15 min. Another 10 ml of Buffer P3 was added, mixed together with the content of the tube vigorously and incubate on ice for 20 min. This was then centrifuged at 4000 x g for 1 hr at 4°C and the resulting supernatant was poured through a sheet of gauze placed on the top of the QIAGEN-tip 500 that was equilibrated with 10 ml of Buffer QBT. The QIAGEN-tip 500 was washed twice with 30 ml of Wash Buffer on the vacuum manifold. The QIAGEN-tip 500 was place inside a new 50 ml tube using the blue tip holder and thus the DNA was eluted with 15 ml Buffer QF which was warmed at 65°C. Afterwards 10.5 ml RT isopropanol was added to the eluted DNA and centrifuged at 4000 x g for 2 hrs to precipitate the DNA. Subsequently, the resulting supernatant was discarded, and 5 ml of 70 % RT ethanol was used to resuspend the pellet and also centrifuged at 4000 x g for 1.5 hrs to wash the precipitated DNA. Lastly, the supernatant was discarded, and the resulted pellet was air-dried for 10 min and re-dissolved in 200 µl of nuclease-free water.

3.5 Transfection into Jurkat and Astrocyte Cell lines

3.5.1 Maintenance and Preparation of Jurkat cell

Jurkat cells are T lymphocyte cell lines derived from an acute Leukaemia T cell (Audiffred et al., 2010, Walzel et al., 2002). This cell line was chosen because it represents the cells of the peripheral compartment. Jurkat cells were cultured in growth media consisting of Roswell Park Memorial Institute (RPMI) 1640 (Gibco BRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% Fetal bovine serum (FBS) (Gibco BRL, Life Technologies, Cergy Pontoise, France), 5 % Penicillin Streptomycin (Pen-Strep) (Sigma–Aldrich, St. Louis, MO, USA), 1 % HEPES (Gibco BRL, Life Technologies, Cergy Pontoise, France) and 1 % L-Glutamine (L-Glut) (R10 medium) (Sigma–Aldrich, St. Louis, MO, USA). This growth media is called R10.

Briefly, a vial of frozen Jurkat cells was removed from the liquid nitrogen tank and placed in dry ice. The frozen vial of Jurkat cells was thawed by placing the vial in a prewarmed water bath at 37˚C with a gentle swirling until there was just some little ice left in the vial. The thawing process was no longer than 2 min and the vial was carefully not immersed in water. Once the cells were thawed, the vial was sprayed with ethanol and the cell suspension was quickly pipetted into a 15 ml tube with 9 ml of pre-warmed growth media (R10). The tube with the cell content was centrifuged for 10 min at 290 x g (Eppendorf centrifuge 5810R, Merck, Germany) to remove dimethyl sulfoxide (DMSO). The resulting supernatant was discarded, and the pellet was resuspended by pipetting up and down in 10 ml of pre-warmed growth media and the cell content was centrifuged for 10 min at 290 x g.

Next, the supernatant was discarded, and the cell pellet was re-suspended by gently pipetting up and down in 2 ml of pre-warmed growth media. Subsequently the 2 ml cell suspension was transferred into a T-25 culture flask containing 3 ml pre-warmed growth media making a total volume of 5 ml. The following day the cell culture was topped up with 5 ml of pre-warmed growth media making a total volume of 10 ml. The next day, the cells were transferred from the T25 into the T75 flask using a serological pipette and topped up with 5 ml of pre-warmed growth media and was incubated for 24 hrs at 37°C 5% carbon dioxide (CO_2). After 24 hrs, the T75 cell culture flask was viewed on the Zoe® fluorescent imager (Bio-Rad) to look for good cell growth prior to cell splitting. The cells

were counted before splitting and assessed the 95-100 % confluency. Cell confluency is understood to be the percentage area covered by cells in a flask (Haenel and Garbow, 2014). Briefly, using a serological pipette cell contained in the flasks were mixed by pipetting up and down and 1 ml was aliquoted into an Eppendorf tube. In another Eppendorf tube, 10 µl of Tryphan blue (BIO-RAD Laboratories, Hercules, CA) was aliquoted. In the tube with 1 ml cell suspension, 10 µl of the cell suspension was pipetted out and mixed with the 10 µl of Tryphan blue.

From the Tryphan blue cell suspension mixture, 10 µl was pipetted into one side of the slide for cell count and the other 10 µl was pipetted into the other side of the slide. The slide was inserted in the TC20™ Automated Cell Counter (BIO-RAD Laboratories, Hercules, CA) to determine the cell count per ml and viability. Knowing the cell count per ml was useful in determining the number of cells to be seeded. Following this step, the cells in the flasks were then seeded in new T75 culture flasks with approximately 35.0 x 10^4 cells in 25 ml pre-warmed growth media and incubated for 24 hrs at 37^{o}C in 5% CO_2. These cells were maintained in culture by passaging them once a week until passage 15 and after which, that a new vial of fresh cells was requested for ongoing experiments.

3.5.1.1 Transfection in Jurkat cell

Briefly, once the Jurkat cells reached 95% viability and more, they 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 pGL3 plasmid containing either consensus or patient derived LTR along with 100 ng of pTarget (empty Tat) vector was prepared in 50 µl of serum-free RPMI medium. For the co-transfection assay which is the Tat induced transcriptional activity, the pGL3-LTR along with consensus pTarget containing HIV-1 subtype C Tat was prepared in 50 µl of serum-free RPMI medium. For the positive control, a pGL3 and pTarget plasmid vectors containing consensus LTR and Tat respectively were used and for the negative control, an empty pGL3 and pTarget expression vectors were used. Lastly, a cell control containing Jurkat cells only was included.

A total of 1 µl (1µg/ml) of polyethylenimine (PEI) was diluted with 49 µl of serum-free RPMI medium to prepare the lipid transfection reagent. The 50 µl polyethylenimine (PEI)-RPMI mixture

was mixed together with 50 µl of the recombinant pGL3 plasmid DNA. Subsequently, the plasmid-lipid mixture was incubated for 20 min at RT and thereafter it was added to the appropriate wells thus incubated at 37ºC in 5% CO_2. Twelve hours post-transfection, the cells were washed twice with Phosphate Buffer Saline (PBS) (pH7.2) (GIBCO Laboratories, Grand Island, NY) to remove the lipid complexes and was further resuspended in 500 µl of R10 and returned to the incubator at 37ºC in 5% CO_2. The cells were then incubated for 12 hrs prior to the luciferase assay. This experiment was done in triplicates and repeated three times.

3.5.2 Maintenance and Preparation of Astrocyte cells

The Astrocyte (SVG) cells are adherent cells, and this cell line was chosen because it represents the cells of the central nervous system (CNS) compartment. The Astrocyte cell line used for this study is a line of human fetal glial cells that supports John Cunningham (JC) virus multiplication which were established in Australia (Major et al., 1985). The cells were culture in growth media consisting of Dulbecco modified Eagle medium (DMEM) (Gibco BRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% FBS (Gibco BRL, Life Technologies, Cergy Pontoise, France), 0.5 % Pen-Strep (Sigma–Aldrich, St. Louis, MO, USA), 1 % HEPES (Gibco BRL, Life Technologies, Cergy Pontoise, France), 1 % L-Glut (Gibco BRL, Life Technologies, Cergy Pontoise, France) 1 % light sensitive N2 supplement (Sigma–Aldrich, St. Louis, MO, USA).

Briefly, a vial of SVG cells was removed from the liquid nitrogen tank and thawed. The vial of SVG cells was thawed by placing the vial in a 37˚C water bath until there was just a small bit of ice left in the vial/ for no longer than 2 min being careful to not immerse entirely in water. Once the cells were thawed, the vial was sprayed with ethanol and the cell suspension was quickly pipetted into a 50 ml tube using a serological pipette with 30 ml of pre-warmed un-supplemented DMEM. The cell suspension was centrifuged to remove DMSO for 10 min at 290 x g and the resulting supernatant was discarded. Following this, the cell pellet was re-suspended in 5 ml of fresh pre-warmed growth media and transferred into a T75 culture flask and was topped up with 10 ml pre-warmed growth media thus incubated for 24 hrs at 37ºC.

After 24 hrs of incubation, the old growth media was replaced with 15 ml of fresh warmed up medium and the flask was placed back in the incubator 2-3 days. After 3 days culture flask containing cells were viewed on the fluorescent imager to assess growth and a 90-100% confluency. The cells were then split by decanting the supernatant and rinsing the cell monolayer with 5 ml of PBS furthermore, the cells were then trypsonized with 3 ml of 0.25% Trypsin-EDTA (GIBCO BRL, Life Technologies, Cergy Pontoise, France). Subsequently the cells were strictly incubated at RT for 30 sec after which the trypsin solution was removed, and the flask was further incubated at 37^{o}C in 5% CO_2 for 4 min. Immediately afterwards, the flask was filled with 10 ml of growth media and 1 ml of the cell suspension was aliquoted into an Eppendorf tube for cell count. The cell count was performed as described for Jurkat cells above. Precedingly, the cells were seeded into new T75 culture flasks with a cell density of 1 x 10^6 cells/ml and 90% and above confluency in 10 ml of pre-warmed growth media prior to incubation at 37^{o}C in 5% CO_2 for 2-3 days. Cells were maintained up until the 15th passage and thereafter, a new fresh vial was requested.

3.5.2.1 Transfection in Astrocyte cell lines

As previously described by Gray *et al*. 2016 a 96-well plate was used to plate 5000 cells per well in a total volume of 100 µl of DMEM supplemented with 10% FBS only and incubated for overnight at 37^{o}C in 5% CO_2. The next day, old medium was replaced with new 100 µl fresh DMEM supplemented with 10% FBS and returned to the incubator and left to stay overnight. The following day, the plasmid-lipid mixture was prepared in a similar way as mentioned above for Jurkat cells. After 12 hrs of post-transfection, the old media was replaced with 100 µl of complete growth media and returned to the incubator for 20 hrs prior to luciferase assay. This experiment was done in triplicates and repeated three times.

3.5.3 Luciferase Assay

3.5.3.1 In Jurkat cell line

Bright-Glo™ Luciferase Assay System (Promega, Madison, WI, USA) was thawed at an ambient temperature water bath away from light. The cells in the 24-well plate were removed into a 15 ml tube and washed using PBS, a residual volume of 300 µl was left behind and used to resuspend the cell pellet. A volume of 100 µl of the cell suspension was aliquoted in each of triplicate wells of a

transparent 96-well plate. Following, the light was switched off and 100 µl of bright-Glo was added to each well with cell suspension and mixed by pipetting up and down thoroughly. This solution was incubated for 2 min and immediately 150 µl was transferred into the black flat bottom 96-well plate (Greiner Bio-One, Monroe, NC) and directly measured in the Victor Nivo Multimode plate reader (PerkinElmer, Massachusetts, USA).

3.5.3.2 In Astrocyte cell line

After 20 hrs, approximately 100 µl of supernatant was removed from each of the triplicate wells leaving a 100 µl layer of supernatant to cover the cells in each well. Next, 100 µl of Bright-Glo™ Luciferase Assay System was added on the cells and incubated at RT for 2 min, in the dark. Following incubation, the contents of the plate was mixed by pipetting, thereafter 150 µl of the lysate was transferred to corresponding wells of a 96-well black bottom microplate thus measured in the Victor Nivo Multimode plate reader.

3.6 Statistical analysis

Statistical values were automatically generated 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, Mann-Whitney. The p value <0.05 was considered statistical significance.

CHAPTER 4: RESULTS

4.1. Phylogenetic tree analysis demonstrates genetic diversity of the long terminal repeat (LTR)

The key demographic and clinical characteristics of the 20 patients included in this study are presented in Table 1. The tuberculous meningitis (TBM) and non-TBM group had a median age of 28 and 34 with an interquartile range [IQR] of 20 -40 and 29-38 years, respectively. The TBM and non-TBM group showed a median plasma CD4 count of 117 cells/µl [IQR, 14 – 677] and 113 cells/µl [IQR, 36 – 532] respectively. The median plasma viral load for the TBM group was 390 650 copies/ml [IQR, 289 217 – 1 710 603] while it was 102 025 copies/ml [IQR, 4317 – 213 260] for the non-TBM group. On the other hand, the TBM and non-TBM group had a median lymphocyte count in the cerebrospinal fluid (CSF) of 6 cells/µl [IQR, 2 – 122] and 2 cells/µl [IQR, 0.24 – 6] respectively. Yet in agreement with a previous report (Morris et al., 1998), the median CSF viral load of the TBM group was significantly higher 630 291 copies/ml [IQR, 8133 – 10 000 000] compared to the non-TBM group 10 631 copies/ml [IQR, 5109 – 14 230]) ($p = 0.020$). Furthermore, the TBM group had a higher CSF viral load 630 291 copies/ml [IQR, 8133 – 10 000 000] compared to the plasma viral load 390 650 copies/ml [IQR, 289 217 – 1 710 603].

A total of 39 (19 plasma and 20 CSF) samples obtained from 20 (17 TBM, 3 non-TBM) patients were used to investigate LTR genetic variation in the TBM cohort (Figure 4.1). From 39 human immunodeficiency virus type 1 (HIV-1 LTR), 20 CSF and 19 plasma samples were obtained, one plasma sample was not available from one of the non-TBM patients (X235). The phylogenetic tree was rerooted against an Indian subtype C reference sequence (EF178613.1) to compare with the 20 study participants which were infected with HIV-1 subtype C viruses.

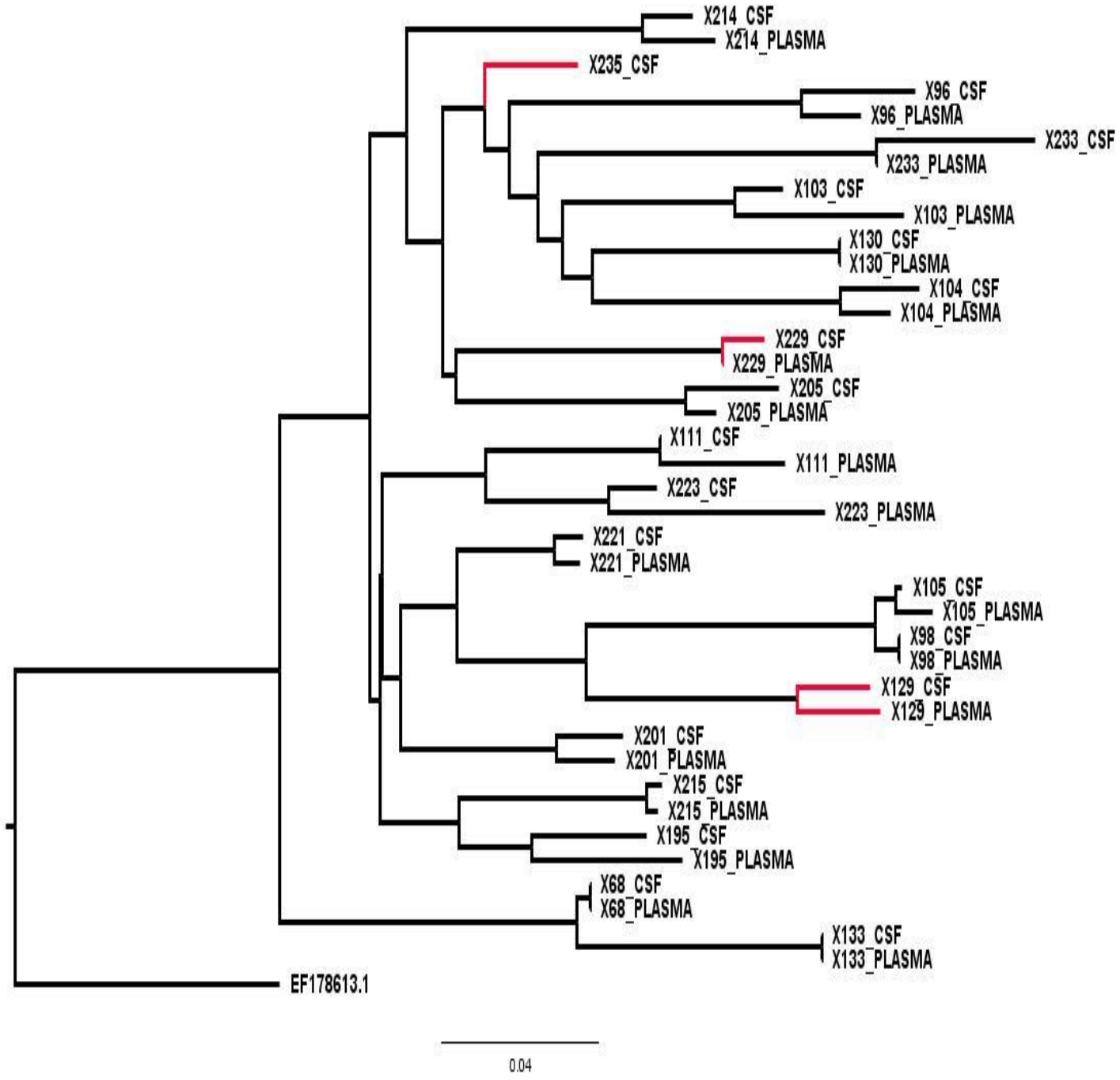


Figure 4.1. Phylogenetic tree analysis of 20 PID's from the TBM cohort. The phylogenetic tree was constructed using the online tool Phmyl (http://www.hivlanl.gov) and was rooted to a subtype C LTR consensus sequence (EF 17 8613.1) along with the 20 matching CSF and 19 plasma patients derived LTR using FigTree Figure Drawing Tool v1.4.3. The TBM sequences are marked in black and the non-TBM sequences are marked in red on the phylogenetic tree. The matched CSF and plasma derived LTR from each patient clustered together and were related to the subtype C consensus sequences. A scale of 0.04 represent 4% of nucleotide sequence divergence.

Our data demonstrated that matched CSF and plasma derived LTR from each patient clustered together (Figure 4.1). This indicated that the CSF and plasma LTR were isolated from the same patient and that there was no cross-contamination and intermingling between the patients. All PID's had matching CSF and plasma derived LTR except X235 who's plasma sample was not available. There were 12 out of 19 (63%) matching CSF and plasma derived LTR: denoted as X214, X96, X233, X103, X104, X205, X111, X223, X105, X195, X201 from the TBM group and X229 from the non-TBM group which, exhibited distinct differences in branch length between the two compartments suggesting that genetic variation between the CSF and plasma compartment exists. On the other hand, the data showed that 7 out of 19 (37%) matching CSF and plasma derived LTR denoted as X221, X98, X130, X68, X215, X133, X129 exhibited no branch length differences. These data suggest that the CNS and peripheral compartments may be infected with the viral strain containing the same LTR.

4.2 Multiple sequence alignment of CSF and plasma derived LTR elements

Next, we hypothesized that branch length difference observed in the phylogenetic tree could be due to sequence variation within the transcription factor binding site (TFBS). To this extent, the CSF and plasma derived LTR U3/R were aligned with the Indian subtype C reference sequence (EF178613.1) on Jalview v2.10.5 (Waterhouse et al., 2009). The data from this study demonstrate that the TFBS USF, RBE III, Sp1I, Sp1II and Sp1III sites, TATA Box, E- Box, NF-κB binding sites (H NF-κB and HH NF-κB) and C-NF-κB binding site specific to subtype C and TAR region aligned clearly with the TFBS in the reference sequence (Figure 4.2). This confirmed that the CSF and plasma derived LTR were indeed from HIV-1 subtype C viruses.

The LTR U3/R analysis showed no unique genetic variation between CSF and plasma derived LTR. Instead, our data showed that the TFBS NF-κB (H_κB and HH_κB), E-Box and the TATA box were relatively conserved while the 4th NF-κB (F_κB) and RBE III site contained deletions in the LTR sequences. None of the LTR analysed in this study exhibited the 4th NF-κB site as seen in previous studies (Bachu et al., 2012b, Boullosa et al., 2014, Obasa et al., 2019). The X215 CSF and plasma derived LTR sequences showed an insertion (CTGCTGACA) at the end of the canonical sequence of HH_κB. On the other hand, the C_ κB binding site specific to subtype C exhibited polymorphism at position 7 and 8 from Thymine (T) to Guanine (G) in some patients. The Sp1I binding site in our

data showed nucleotide changes at position 9 and 10 from Cytosine (C) to T (27/39) and T to C (35/39) respectively whereas, the Sp1II exhibited a mutation at position 10 from T to C. The Sp1III binding site also showed notable genetic variation at position 2 from G to Adenine(A), and position 5 from A to T. The G2A/A5T mutation were observed to occur independently and in other LTR elements co-existed together. Our data show that the 5' E box (i) did show minor genetic variation at position 6 from G to A.

The TAR element exhibited subtype C specific polymorphisms. Within the TAR element, nucleotide changes at position 24 and 30 from a C to T were observed. A nucleotide change from G to A at position 50 was also seen within the TAR element. These data suggest that genetic variation within the TFBS existed in CSF and plasma derived LTR from TBM co-infected patients.

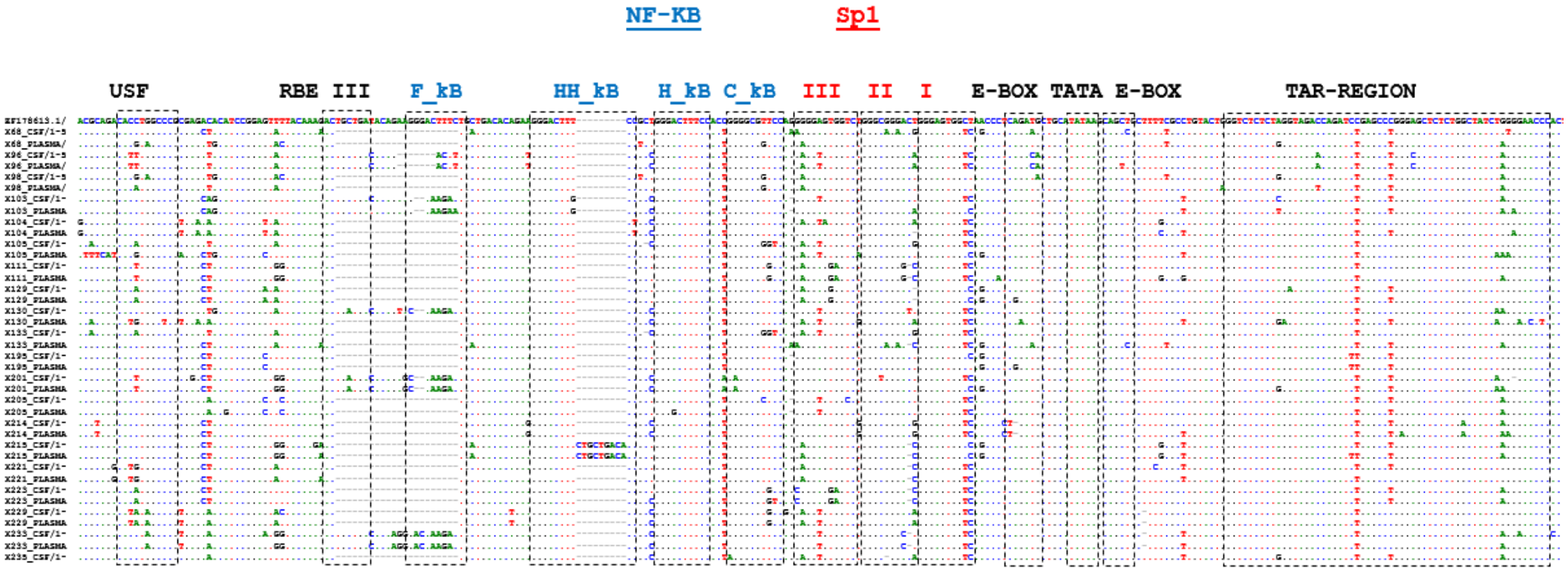


Figure 4.2. Multiple sequence alignment of CSF and plasma patient derived LTR element. The LTR element spans the core-modulatory, core-enhancer and core-promoter domain. The patient derived LTR elements was aligned against the Indian Subtype C reference sequence (EF178613.1). Several TFBS were mapped across the LTR element from left to right: USF, RBE III, 4th NF-κB (F_KB), two canonical NF-κB binding sites (HH_ and H_KB), 3rd NF-KB (C _KB), the three Sp1I, II and III, Tata box and E-box and TAR loop region. The RBE III, TATA Box, NF-κB I, NF-κB II, Sp1I, Sp1II and the E-Box TFBS were observed to be relatively well conserved with the exception of the USF, C NF-κB and mostly the Sp1III TFBS which exhibited a greater degree of genetic variability within the core promoter region.

4.3 CSF derived LTR have significantly high transcriptional activity compared to plasma derived LTR in Astrocyte cell lines (SVG).

The study by Gray *et al*. observed significantly reduced basal transcriptional activity of CNS derived LTR in both astrocytes and T cells compared to that of non-CNS derived LTR obtained from HIV-1 mono-infected individuals (Gray et al., 2013). However, the same study showed a heterogeneous Tat induced transcription activity in these HIV-1 mono-infected individuals where CNS derived LTR showed greater activation than non-CNS derived LTR in astrocytes, but this difference was not observed in T cells (Gray et al., 2013). The data from Gray *et al.* indicated that CSF and plasma derived LTR show differential transcriptional activities in Astrocyte and Jurkat cell lines (Gray et al., 2016), which could explain low viral load in the CNS of HIV-1 mono-infected individuals. However, it is unknown whether LTR genetic variation may be associated with higher transcriptional activity in the CNS of TBM co-infected individuals. Therefore, we hypothesized that CSF and plasma derived LTR elements from TBM co-infected individuals may result in differential transcriptional activity. We assessed basal and Tat induced LTR transcriptional activity from all the 17 CSF and 17 plasma derived LTRs of TBM co-infected individual's in Jurkat and SVG cell lines. All three CSF and two plasmas (X235 did not have a matching plasma derived LTR) derived LTR from the non-TBM (HIV-1 mono-infected) group were excluded in the transcriptional analysis due to the small sample size. Therefore, the total number of samples being analysed for the LTR transcriptional activity is 34 (n= 17 CSF and n=17 plasma) from the TBM group. The Jurkat cell line represents cells from the periphery while the SVG cell line represents cells from the CNS compartment.

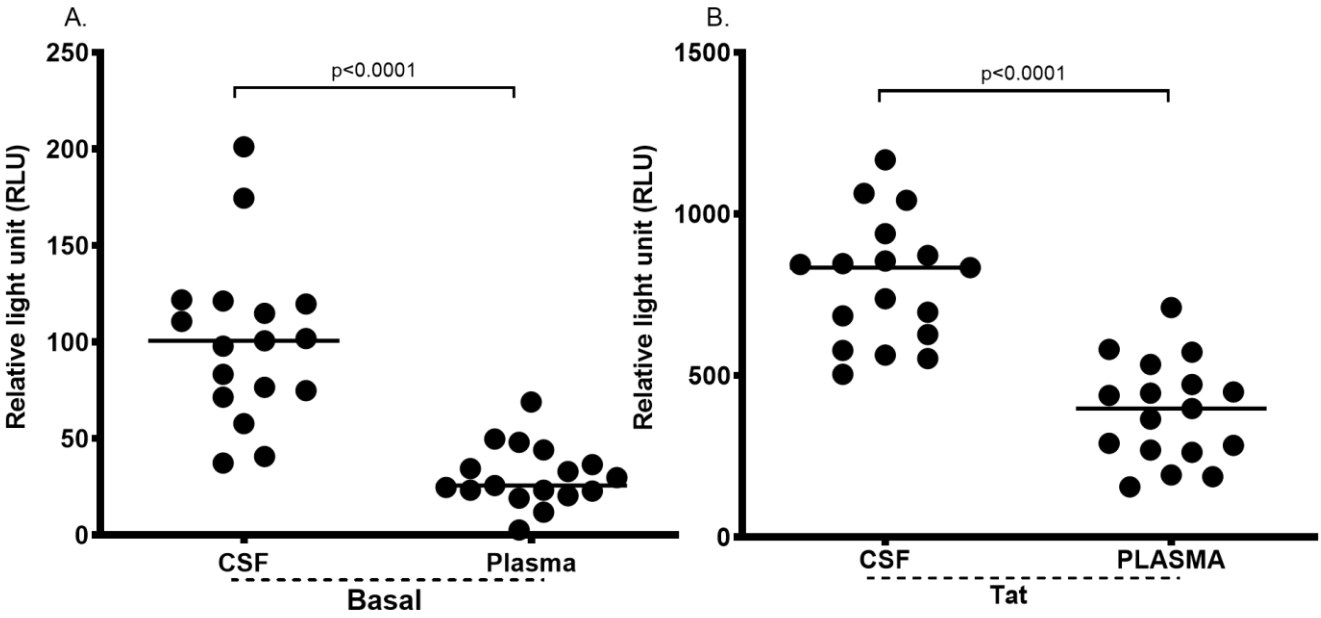


Figure 4.3 Transcriptional activity of CSF and plasma derived LTR elements from TBM co-infected individuals in Astrocyte cell line. A concentration of about 300 ng/µl of pGL3 plasmid cloned with patients derived LTR was transfected using PEI transfection reagent in SVG cell line to assess basal and Tat induced transcriptional activity. Panel A, Basal transfection was performed with 300 ng/µl of pGL3 and 100 ng/µl pTarget empty vector. Panel B, Tat induced transfection was performed with 300 ng/µl and 100 ng/µl pCTat vector containing a consensus *tat* gene. The transcriptional activity was measured as Relative light Unit (RLU) which is Average Luminescence/Wildtype Luminescence. CSF derived LTR results to significantly increased LTR transcriptional activity compared to plasma derived LTR in SVG cell line. The black lines between the dots represent the median. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data demonstrate that CSF derived LTR from TBM co-infected individuals exhibit significantly increased basal transcriptional activity compared to plasma derived LTR ($p<0.0001$) in SVG's (Figure 4.3A). Consistently, CSF derived LTR from TBM co-infected individuals exhibited significantly higher Tat induced transcriptional activity compared to plasma derived LTR ($p<0.0001$) in SVG's (Figure 4.3B). The Tat induced transcriptional activity was higher than basal, indicating that transcriptional is enhanced under Tat induced conditions. Our results suggest that CSF derived LTR from TBM co-infected individuals have significantly enhanced transcriptional activity compared to plasma derived LTR in SVG cell lines. However, it must be noted that these results are inconclusive since the appropriate normalization controls such as co-transfecting a renilla luciferase construct was not included to control for transfection efficiency in this experiment. While Tat mediated CSF derived LTR transcription activity was significantly higher than that of plasma derived LTR transcription activity, these data are also not conclusive since the cell numbers assayed were not controlled using assays such as the Bradford protein assay.

4.4 CSF derived LTR containing A at position 5 of the Sp1III transcription binding site have significantly higher transcriptional activity compared to a T at the same position in Astrocyte cell line (SVG).

Although previous studies reported that genetic variation within the LTR U3 domain translates to differential transcriptional activity (Bachu et al., 2012b, Jeeninga et al., 2000, Qu et al., 2016), this was only shown for plasma LTR sequences from HIV-1 mono infected individuals. Within the U3 domain, the Sp1III site showed genetic variation which have been reported previously. We therefore wanted to investigate the effect of Sp1III binding site genetic variation with the LTR transcriptional activity in the SVG cell line. Specifically, we hypothesized that LTR containing Sp1III A5T mutation might have differential transcription activity. The TBM derived CSF and plasma LTR containing a T and separately those containing an A at position 5 of the Sp1III in combination with other mutations, were selected and were analysed at basal and Tat induced conditions into SVG cell line. There were 9/17 CSF derived LTR containing an A and 8/17 containing a T at position 5 of Sp1III. On the other hand, 12/17 plasma derived LTR contained an A and 5/17 contained a T at position 5 of Sp1III.

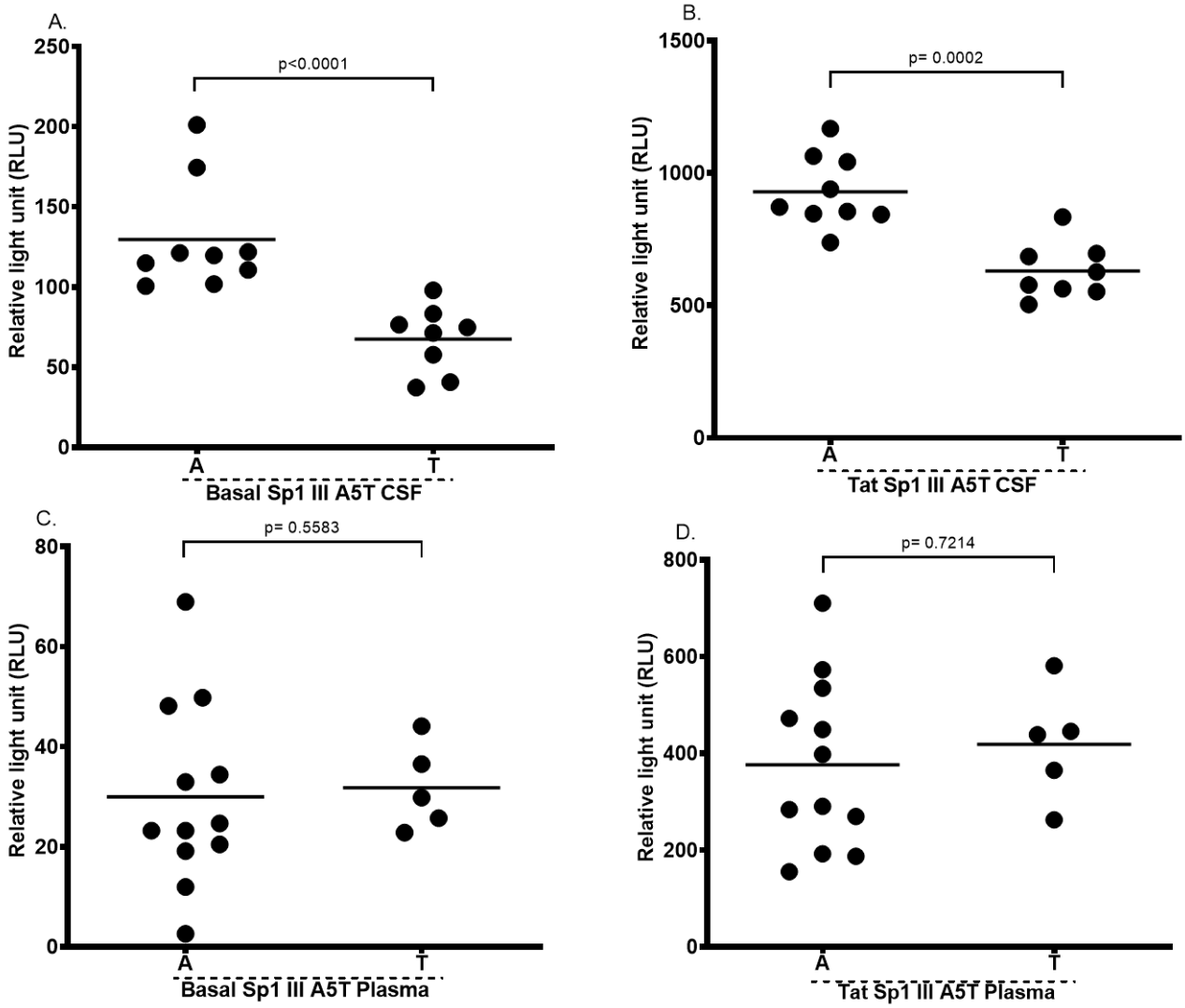


Figure 4.4 Transcriptional activity of CSF and plasma derived LTR element containing the A5T mutation in SVG cell line. Patient derived LTR was transfected using PEI transfection reagent in SVG cell line to assess basal and Tat induced transcriptional activity. Panel A and B is CSF derived LTR containing the A5T mutation in the Sp1III binding site whereas Panel C and D are plasma derived LTR with the same mutation at the same position. CSF derived LTR containing A in the Sp1III TFBS have significantly high transcriptional activity compared to LTR containing a T. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data show that CSF derived LTR containing an A at position 5 of the Sp1III binding site exhibits significantly higher basal transcriptional activity compared to a T at the same position ($p<0.0001$) in SVG's under basal conditions (Figure 4.4A). Consistently, the data from the current study demonstrate that, CSF derived LTR containing an A at position 5 of the Sp1III binding site have significantly high Tat induced transcriptional activity compared to T at the same position ($p=0.0002$) (Figure 4.4B). These results indicated that CSF derived LTR containing A at position 5 of the Sp1III

transcription binding site have significantly enhanced transcriptional activity compared to T at the same position in SVG cell line.

On the other hand, plasma derived LTR containing an A at position 5 of the Sp1III binding site showed no difference in LTR transcription activity compared to a T at the same position (p=0.5583) in SVG's under basal conditions (Figure 4.4C). This was also the case under Tat conditions (p=0.7214) (Figure 4.4D). However, a huge heterogenous spread of transcription activity in plasma derived LTR containing A and T was observed under basal and Tat induced transcription. Figure 4.4C&D clearly stipulates that the plasma derived LTR containing an A in the Sp1III at position 5 has no transcriptional difference compared to a T at the same position. This shows that the A5T mutation seen in plasma derived LTR has no notable effect on LTR transcriptional activity. Taken together our data therefore suggest that an A at position 5 of the Sp1III binding site enhances CSF derived LTR transcriptional activity compared to a T at the same position in SVG's cell lines. This clearly suggested that another mutation besides the A5T mutation existing within the LTR is associated with the enhanced transcriptional activity observed in CSF derived LTR since we did not see a similar trend with this mutation in plasma.

4.4.1 Transcription difference between the presence of a G or A in the Sp1III binding site at position 2 was not observed in CSF and plasma derived LTR Astrocyte cell lines (SVG).

Next, we investigated the effect of Sp1III G2A mutation on the LTR transcription activity. The TBM CSF and plasma derived LTR containing an A and separately those containing a G at position 2 of the Sp1III binding site in combination with other mutations were transfected under basal and Tat induced conditions in SVG cell line. There were 9/17 CSF derived LTR containing a G and 8/17 containing an A at position 2 of the Sp1III binding site. On the other hand, 10/17 plasma derived LTR contained a G and 7/17 contained an A at position 2 of Sp1III binding site.

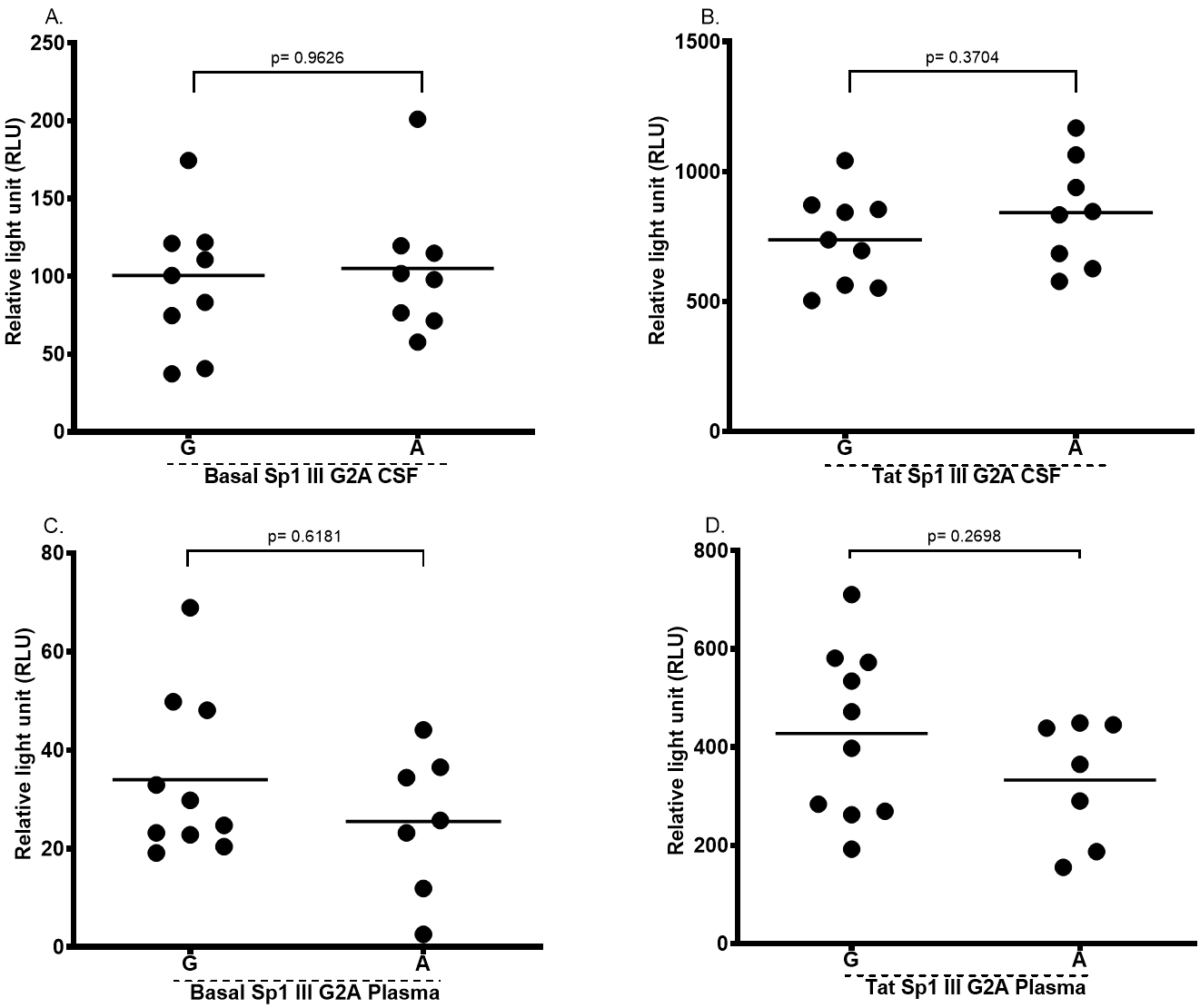


Figure 4.4.1 Transcriptional activity of CSF and plasma derived LTR element containing the G2A mutation in SVG cell line. Patient derived LTR was transfected using PEI transfection reagent in SVG cell line to assess basal and Tat induced transcriptional activity. Panel A and B is CSF derived LTR containing the G2A mutation at position 2 of the Sp1III binding site whereas Panel C and D are plasma derived LTR with the same mutation. LTR containing a G at position 2 of the Sp1III TFBS showed no differential LTR transcriptional activity compared to LTR containing an A at the same position. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data show that CSF derived LTR containing a G at position 2 of the Sp1III binding site has no LTR transcriptional difference (p=0.9626) compared to an A at the same position under basal conditions (Figure 4.4.1A) in SVG's. Under Tat induced conditions a similar case was observed with no significant difference (p=0.3704) in transcriptional activity between CSF derived LTR containing a G or an A at position 2 of Sp1III (Figure 4.4.1B). These data indicate that the presence of an A or a G occurring in the Sp1III binding site at position 2 has no effect in transcriptional activity of CSF derived LTR in SVG cell line.

Similarly, plasma derived LTR containing a G at position 2 of the Sp1III binding site showed no difference in LTR transcription activity compared to an A at the same position (p=0.6181) in SVG's under basal conditions (Figure 4.4.1C). A similar case was observed under Tat conditions (p=0.2698) (Figure 4.4.1D). This clearly stipulates that the G2A mutation has no notable effect on LTR transcriptional activity in plasma derived LTR. Taken together, these data therefore suggest that indeed the G2A mutation observed in CSF and plasma derived LTR has no effect in the transcriptional activity in SVG's cell line.

4.4.2 CSF and plasma derived LTR with Sp1III containing the A5T/G2A mutation showed no significant differential LTR transcriptional activity compared to non-mutants in Astrocyte cell lines (SVG).

Next, we investigated the effect of Sp1III A5T at position 5 occurring at the same time with the G2A mutation at position 2 on the LTR transcription activity (A5T/G2A) in SVG cell line. The CSF or plasma derived LTR containing the A5T/G2A double mutation in combination with other mutations along the LTR were selected separately from those with no mutation at position 2 and 5. The transcriptional activity was assessed under basal, and Tat induced conditions. The LTR which had one mutation occurring alone at position 2 or 5 was not included in this analysis which decreases the sample size being analysed. The LTR being analysed were only those with a double mutation versus no mutation occurring in the two positions. There were 4/17 CSF derived LTR containing the A5T/G2A double mutation and 7/17 non-mutants. On the hand there were 3/17 plasma derived LTR containing the A5T/G2A double mutation and 9/17 non-mutants.

Our data demonstrated that CSF derived LTR with the combination of the A5T/G2A mutation had no differential LTR transcriptional (p=0.1636) activity compared to non-mutants at the Sp1III position 5 and 2 under basal conditions (Figure 4.4.2A). Even under Tat conditions there was no transcriptional difference (p=0.1091) between CSF derived LTR with the A5T/G2A combination mutation compared to non-mutants (Figure 4.4.2B). Figure 4.4.2A&B indicated that the occurrence of the A5T/G2A mutation occurring at the same has no significant difference in transcriptional activity of CSF derived LTR compared to non-mutants in SVG cell line.

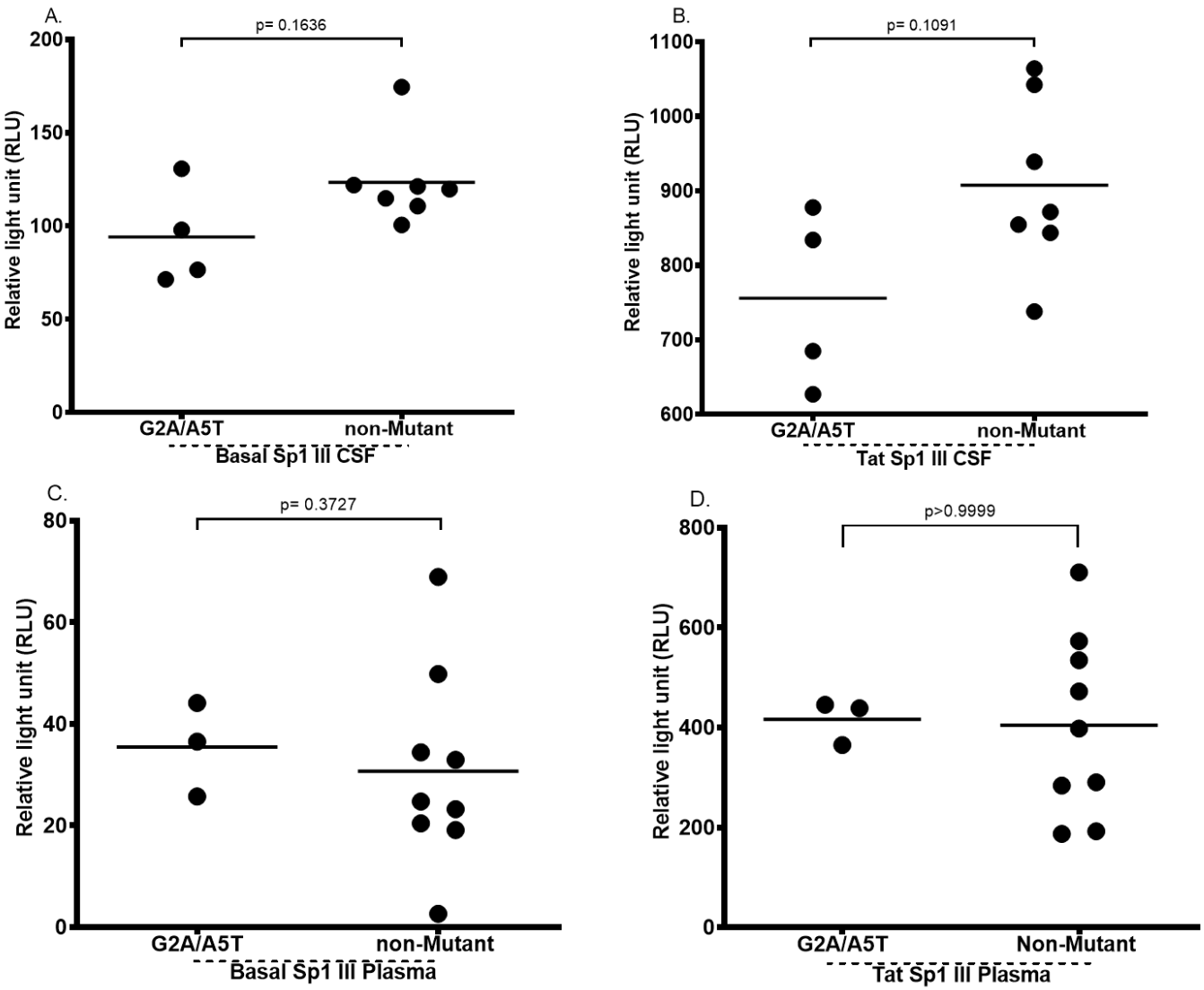


Figure 4.4.2 Transcriptional activity of CSF and plasma derived LTR containing G2A/A5T double mutation compared to non-mutants in SVG cell line. Patient derived LTR was transfected using PEI transfection reagent in SVG to assess basal and Tat induced transcriptional activity. Panel A and B is CSF derived LTR containing the G2A/A5T at position 2 and 5 of SP1III binding site vs non-mutants whereas panel C and D is plasma derived LTR with the same selection. There was no significant LTR transcriptional activity difference between LTR containing the G2A/A5T compared to non-mutants. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

The plasma derived LTR containing A5T/G2A mutation in the Sp1III binding site showed no difference in LTR transcription activity compared to non-mutants at position 5 and 2 ($p=0.3727$) in SVG's under basal conditions (Figure 4.4.2C). A similar case was observed under Tat conditions ($p>0.9999$) (Figure 4.4.2D). These data clearly stipulate that the presence of the A5T/G2A double mutation in the Sp1III binding site at position 5 and 2 does not have an effect on transcriptional activity in plasma derived LTR. Taken together these data therefore suggest that A5T/G2A mutation

occurring at the same time in Sp1III position 5 and 2 has no effect in the transcriptional activity of CSF and plasma derived LTR compared to non-mutants in SVG's cell line.

The following data sets are in Jurkat cell line. The same experimental and data analysis on LTR transcriptional activity performed on SVG cell line were also performed in Jurkat cell lines.

4.5 CSF derived LTR results to significantly increased LTR Tat induced transcriptional activity compared to plasma derived LTR in Jurkat cell line

Next, we assessed basal, and Tat induced LTR transcription activity from all the 17 matched CSF and plasma derived LTR in Jurkat cell lines.

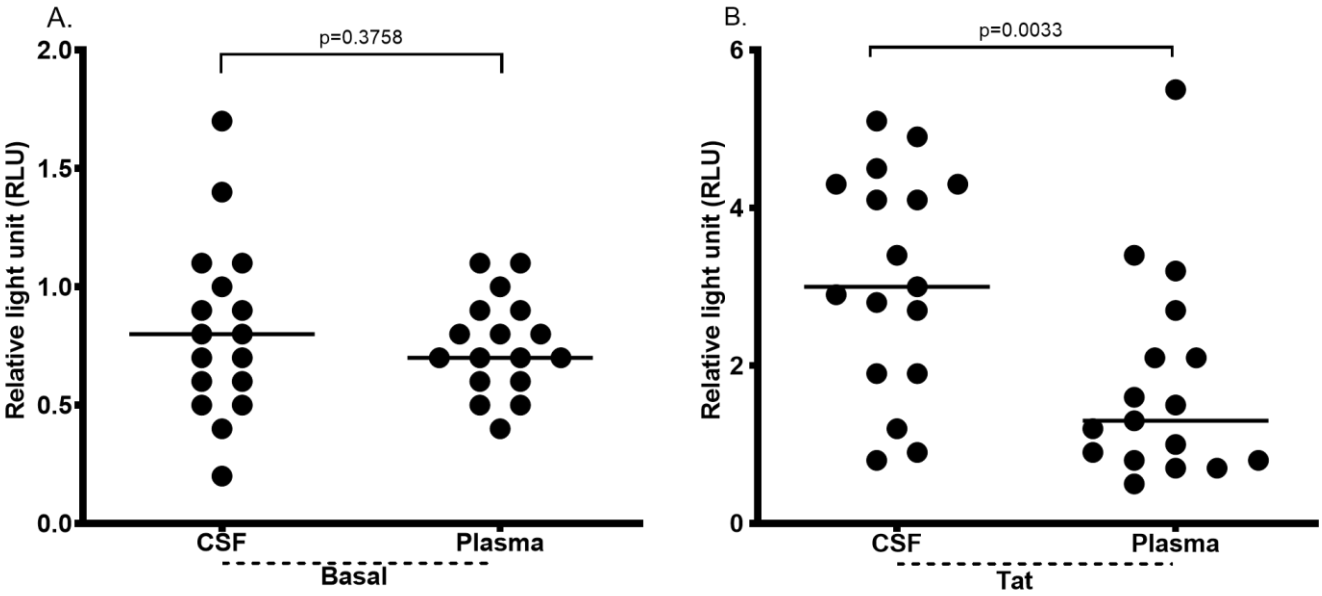


Figure 4.5 Transcriptional activity of CSF and plasma derived LTR element in Jurkat cell line. A concentration of 300 ng/µl of pGL3 plasmid cloned with patients derived LTR was transfected using PEI transfection reagent in Jurkat cell line to asses basal and Tat induced transcriptional activity. Panel A, basal transfection with 300 ng/µl of pGL3 and 100 ng/µl pTarget empty vector. Panel B, Tat induced conditions transfection was performed with 300 ng/ µl and 100 ng/ µl pCTat vector containing a consensus *tat* gene. The transcriptional activity was measured as Relative light Unit (RLU) which is Average Luminescence/ Wildtype Luminescence. CSF derived LTR results to significantly increased LTR Tat induced transcriptional activity compared to plasma derived LTR in Jurkat cell line. The black lines between the dots represents the median. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data demonstrate that CSF derived LTR from TBM co-infected individuals does not show any differential basal transcriptional activity compared to plasma derived LTR (p=0.3758) in Jurkat cell line (Figure 4.5A). Interestingly, CSF derived LTR from TBM co-infected individuals exhibit significantly higher Tat induced transcriptional activity compared to plasma derived LTR (p =0.0033) in Jurkat cell line (Figure 4.5B). Our data indicate CSF derived LTR is associated with significantly increase Tat induced LTR transcriptional activity compared to plasma derived LTR in Jurkat cell lines.

4.5.1 CSF derived LTR containing an A at position 5 of the Sp1III binding site have significantly higher transcriptional activity compared to a T at the same position in Jurkat cell line.

Next, we investigated the effect of Sp1III A5T mutation on the LTR transcription activity in Jurkat cell line. The CSF and plasma derived LTR containing an A or a T at position 5 of the Sp1III binding site in combination with other mutation were selected and analysed to assess basal and Tat induced conditions in Jurkat cell line.

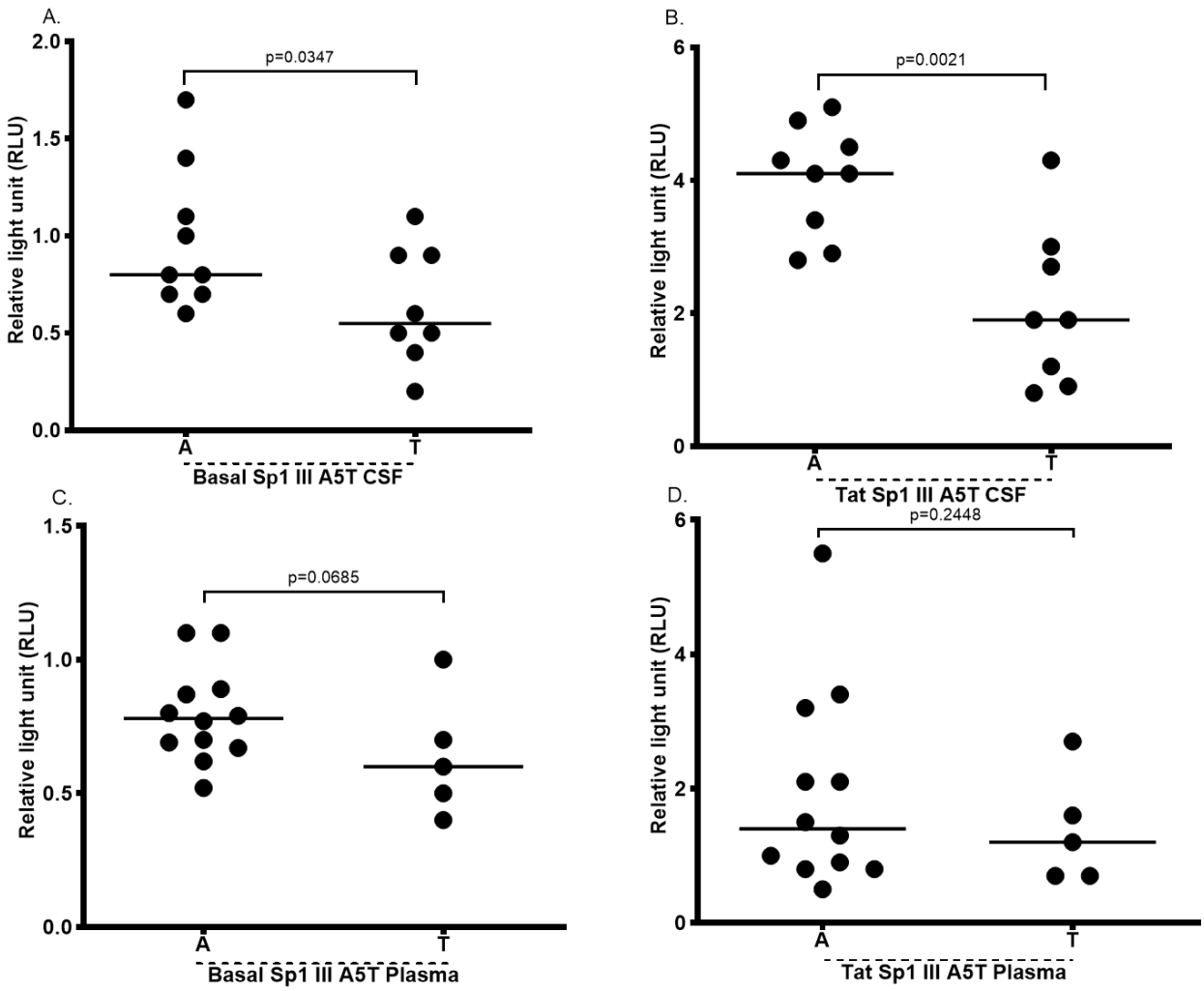


Figure 4.5.1 Transcriptional activity of CSF and plasma derived LTR element containing the A5T mutation in Jurkat cell line. Patient derived LTR was transfected using PEI transfection reagent in Jurkat cell line to assess basal and Tat induced conditions. Panel A and B is CSF derived LTR containing the A5T mutation in the Sp1III binding site whereas panel C and D are plasma derived LTR with the same mutation. CSF derived LTR containing an A in the Sp1III TFBS have significantly high transcriptional activity compared to a T at the same position. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data show that CSF derived LTR containing an A at position 5 of the Sp1III binding site exhibit significantly higher basal transcriptional activity compared to CSF derived LTR containing a T at the same position ($p<0.0347$) in Jurkat cell line (Figure 4.5.1A). Consistently, the data from the current study demonstrate that Tat induced transcription activity of CSF derived LTR containing an A at position 5 of the Sp1III is significantly higher compared to a T at the same position ($p=0.0021$) (Figure 4.5.1B). One CSF derived LTR from PID X133 (TBM) was an outlier with the highest Tat induced transcriptional activity of 4.3 RLU. Figure 4.5.1 A&B clearly show that, an A at position 5

of the Sp1III transcription binding site significantly enhances CSF derived LTR transcriptional activity compared to T at the same position in Jurkat cell line.

On the other hand, plasma derived LTR containing an A at position 5 of the Sp1III binding site showed a trend towards higher LTR transcription activity compared to a T at the same position ($p=0.0685$) in Jurkat cells under basal conditions (Figure 4.5.1C). Under Tat induced conditions plasma derived LTR containing an A at position 5 of the Sp1III binding site showed no differential transcription activity compared to a T at the same position ($p=0.2448$) in Jurkat cells (Figure 4.5.1D). Figure 4.5.1C&D clearly stipulates that the A5T mutation has no effect on transcriptional activity in plasma derived LTR. Taken together these data therefore suggest that an A at position 5 of the Sp1III binding site enhances CSF derived LTR transcriptional activity compared to a T at the same position in Jurkat cell lines, which is not seen in plasma derived LTR. These data suggest that the enhanced transcriptional activity observed in CSF derived LTR is not only associated with the A5T mutation however, there may be other mutations within the LTR which could be a contributing factor.

4.5.2 The presence of the G2A mutation in the Sp1III has no differential transcriptional activity in CSF and plasma derived LTR in Jurkat cell line.

Next, we investigated the effect of Sp1III G2A mutation on the LTR transcription activity. To this end CSF or plasma derived LTR containing an A or G at position 2 of the Sp1III binding site in combination with other mutation were selected and analysed under basal and Tat induced conditions in Jurkat cell line.

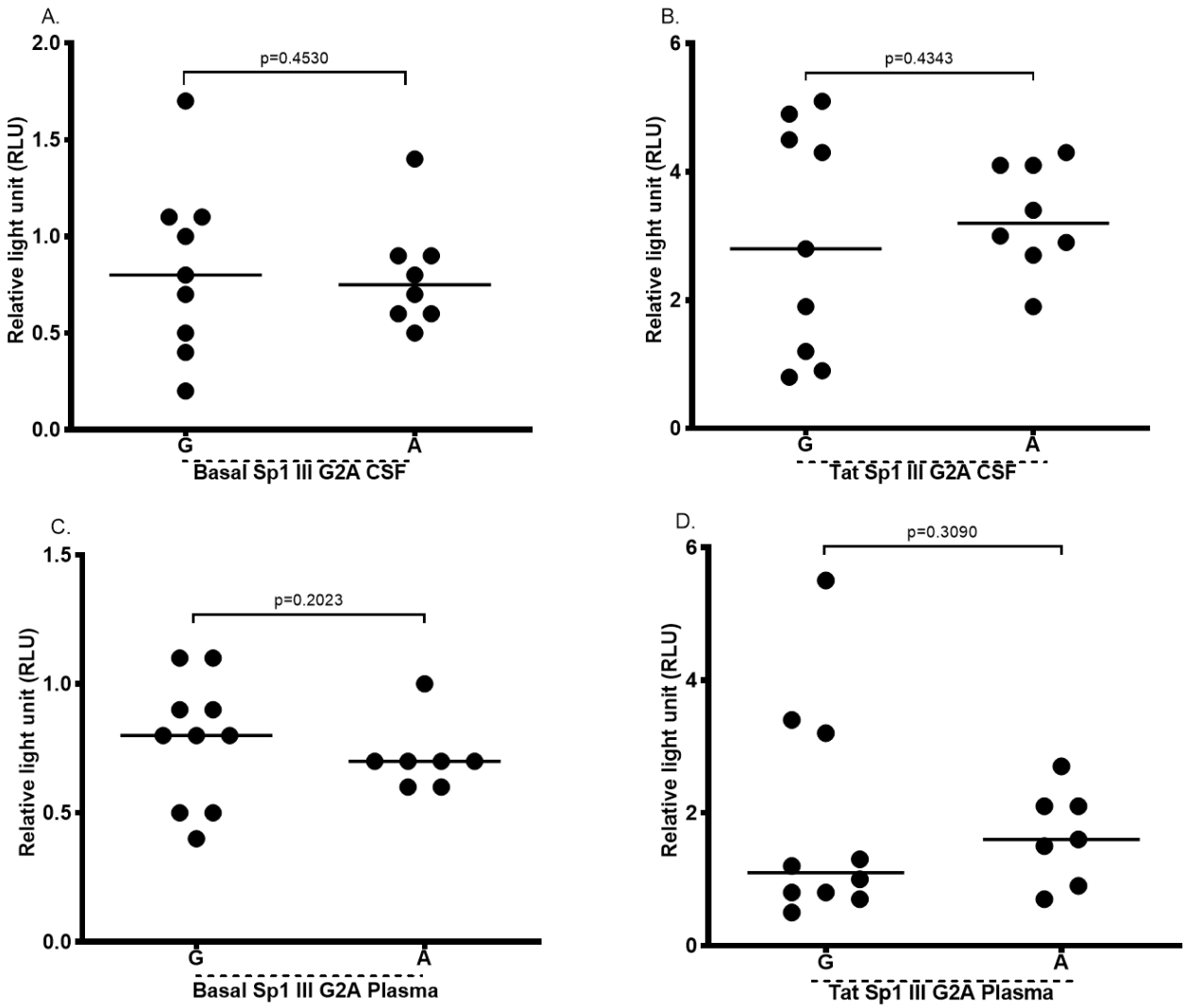


Figure 4.5.2 Transcriptional activity of CSF and plasma derived LTR element containing the G2A mutation in Jurkat cell line. Patient derived LTR was transfected using PEI transfection reagent in Jurkat cell line to assess basal and Tat induced transcriptional activity. Panel A and B is CSF derived LTR containing the G2A in the Sp1III binding site mutation whereas panel C and D are plasma derived LTR with the same mutation. LTR containing a G in the Sp1III TFBS showed no differential LTR transcriptional activity compared to LTR containing an A. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data show that CSF derived LTR containing a G at position 2 of the Sp1III binding site has no transcriptional activity difference (p=0.4530) when compared to an A at the same position under basal conditions (Figure 4.5.2A). Under Tat induced conditions a similar case was observed with no significant difference (p=0.4343) in transcriptional activity between CSF derived LTR containing a G and A in the Sp1II (Figure 4.5.2B). Figure 4.5.2A & B therefore indicates that the presence of the G2A mutation occurring in the Sp1III binding site at position 2 does not affect transcriptional activity of CSF derived LTR in Jurkat cell line.

The plasma derived LTR containing a G at position 2 of the Sp1III binding site showed no difference in transcription activity compared to an A at the same position (p=0.2023) in Jurkat cells under basal conditions (Figure 4.5.2C). A similar case was observed under Tat conditions (p=0.3090) (Figure 4.5.2D). Taken together these data indicate that the G2A mutation has no effect in the transcriptional activity of CSF and plasma derived LTR in Jurkat cell line.

4.5.3 In the presence of Tat the occurrence of the A5T/G2A double mutation in the Sp1III binding site reduces CSF derived LTR transcriptional activity compared to non-mutants in Jurkat cell line

Next, we investigated the effect of Sp1III A5T occurring at the same time with the G2A mutation on the LTR transcription activity (A5T/G2A). To this end CSF or plasma derived LTR containing the A5T/G2A at position 5 and 2 against non-mutants were selected and analysed under basal conditions in Jurkat cell line.

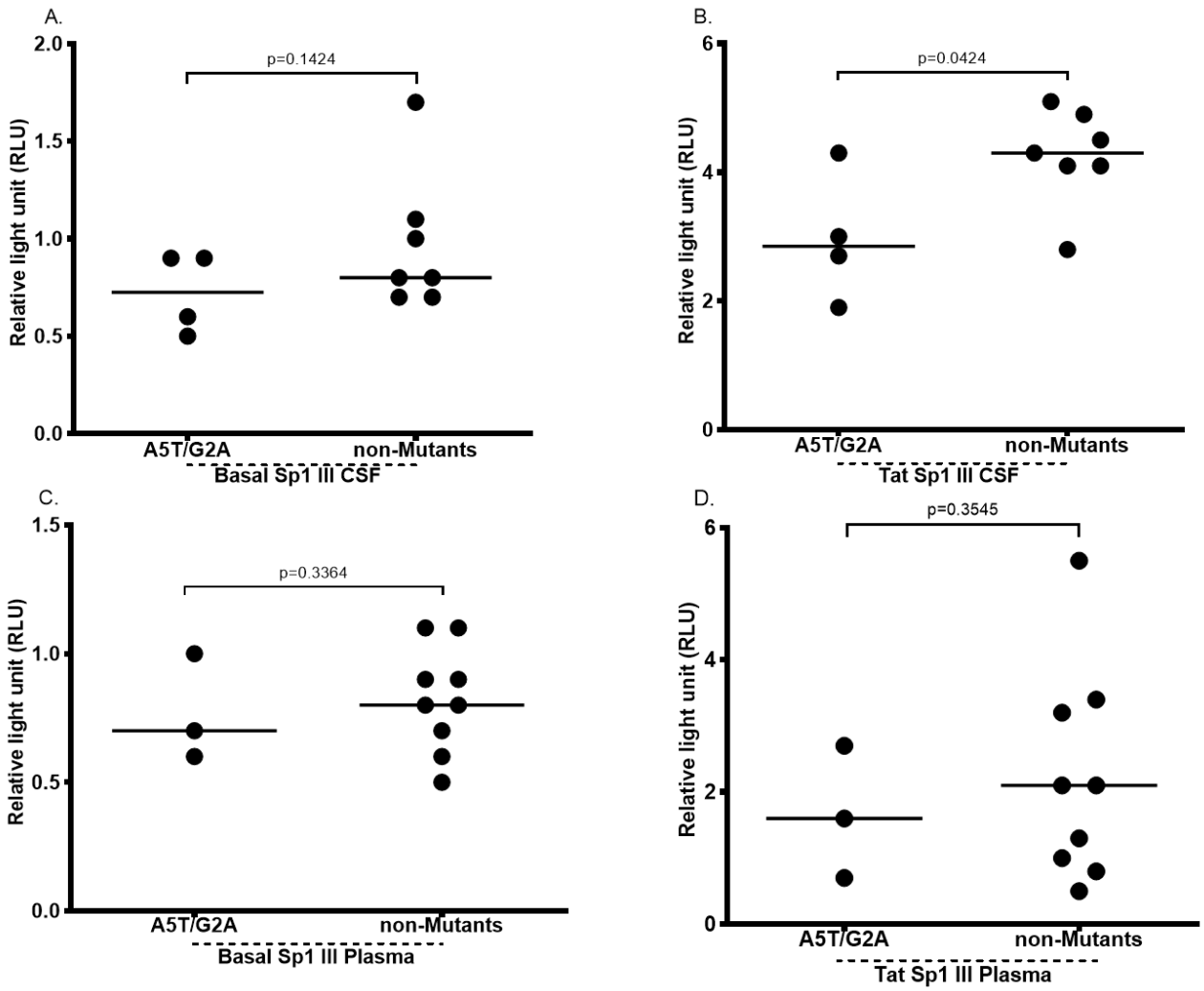


Figure 4.5.3 Transcriptional activity of CSF and plasma derived LTR containing G2A/A5T double mutation compared to non-mutants in Jurkat cell line. Patient derived LTR was transfected using PEI transfection reagent in Jurkat to assess basal and Tat induced transcriptional activity. Panel A and B is CSF derived LTR containing the G2A/A5T vs non-mutants whereas panel C and D are plasma derived LTR. CSF derived LTR with the combination of the G2A/A5T mutation had significantly reduced LTR transcriptional activity under Tat induced conditions compared to non-mutants. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Our data demonstrated that CSF derived LTR with the combination of the A5T/G2A mutation at the Sp1III binding site at position 5 and 2 had no differential LTR transcriptional (p=0.1424) activity compared to non-mutant's under basal conditions (Figure 4.5.3A). Interestingly, under Tat conditions, there was a significant difference observed in transcriptional difference (p=0.0424) where non-mutant CSF derived LTR showed higher transcriptional activity compared to the double A5T/G2A mutation (Figure 4.5.3B). Figure 4.5.3A & B indicated that, in the presence of Tat the

existence of the A5T/G2A double mutation in the Sp1III binding site reduces CSF derived LTR transcriptional activity compared to non-mutants in Jurkat cell line.

The plasma derived LTR containing A5T/G2A mutation in the Sp1III binding site was observed to have no difference in transcription activity compared to LTR with no mutation at position 5 and 2 (p=0.3364) in Jurkat cells under basal conditions (Figure 4.5.3C). This was also the case under Tat conditions (p=0.3545) (Figure 4.5.3D). Figure 4.5.3C&D clearly indicates that the presence of the A5T/G2A double mutation occurring together in the Sp1III binding site at position 5 and 2 does not have an effect on transcriptional activity in plasma derived LTR. Taken together, these data suggest that A5T/G2A mutation in the presence of Tat the CSF derived LTR transcription activity is reduced, however this is not the case in plasma derived LTR in Jurkat the cell line.

4.6. There is no significant association found between the CSF and plasma derived LTR with the HIV-1 viral load

Next, we looked at the impact of LTR genetic variation on viral load. Previous studies have been able to show that genetic variation present within the LTR region may results in differential viral replication capacities (Bachu et al., 2012). We hypothesized that LTR with enhanced transcriptional activity will be associated with increased viral load. We first looked at the viral load between the CSF and plasma compartment. From table 1 TBM co-infected individuals clearly showed higher median viral load in the CSF (630 291 copies/ml) compared to the plasma compartment (390 650 copies/ml). However, there is no significant (p=0.9188) difference in viral load between the CSF and plasma compartment in Figure 4.6. There were only 7 individuals that were observed to have higher viral load in the CSF compared to the plasma compartment.

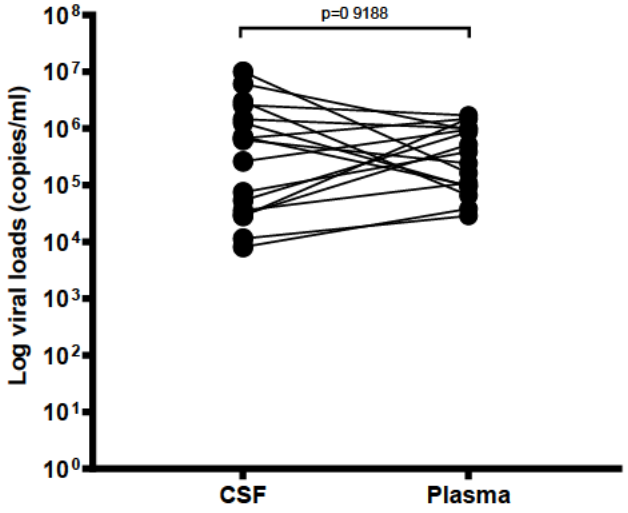


Figure 4.6 Comparison of HIV-1 viral load in CSF and plasma compartment. The viral load values were taken from the HIV Pathogenesis Programme (HPP) database from the TBM cohort. There is no significant difference in viral load between the CSF and plasma compartment. The three non-TBM individuals were excluded. A value of $p < 0.05$ is considered to be significant by Mann Whitney test.

4.6.1 CSF derived LTR containing an A at position 5 of the Sp1III binding site exhibits significantly increased viral load compared to a T

Consequently, we compared the CSF and plasma viral load of all the individuals that had LTR sequences exhibiting A at position 5 of the Sp1III binding site in combination with other mutations to the CSF of individuals that exhibited a T at same position. Our data show that CSF derived LTR containing A at position 5 of the Sp1III binding has significantly higher viral load ($p=0.0079$) compared to a T at the same position (Figure 4.6.1A).

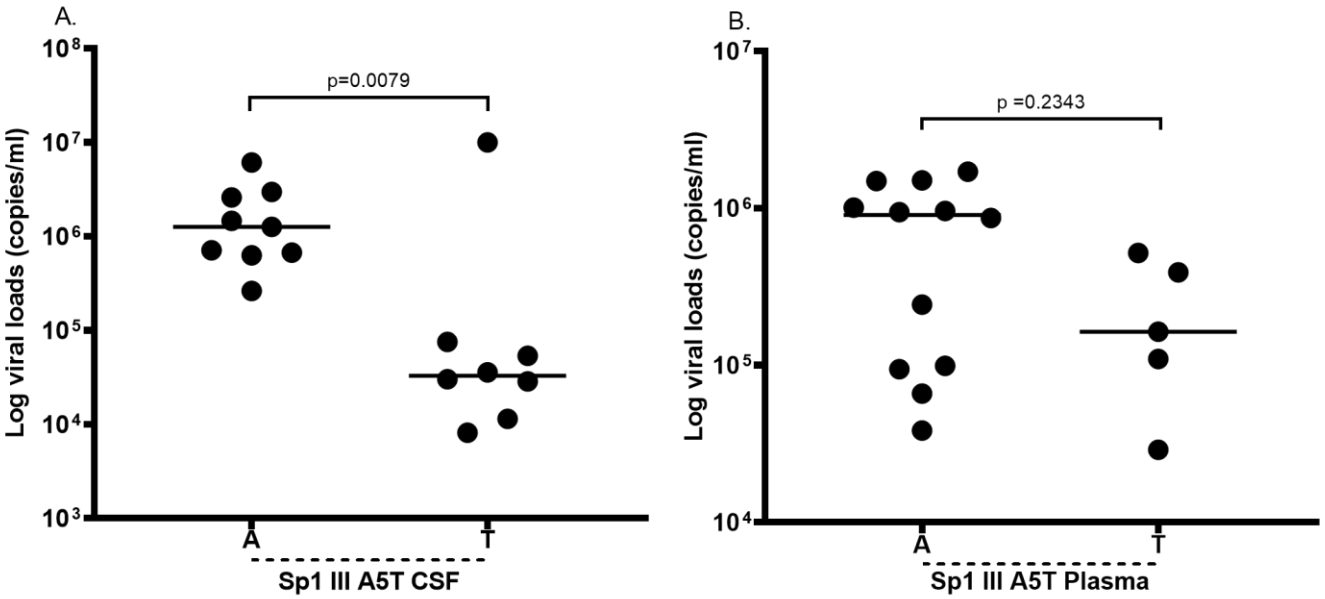


Figure 4.6.1 HIV-1 viral load of CSF and plasma derived LTR containing the A5T mutation. Panel A represents the HIV-1 viral load in the CSF compartment whereas panel B represents HIV-1 viral load in the plasma compartment. LTR that were assessed are those with the A5T mutation. The viral load values were from the HIV Pathogenesis Programme (HPP) database from the TBM cohort. CSF derived LTR containing an A in the Sp1III transcription binding site have significantly high HIV-1 viral load compared to T mutants. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Plasma derived LTR containing an A at position 5 of the Sp1III binding site showed no difference in viral load (p=0.2342) compared to those containing a T at the same position (Figure 4.6.1B). Taken together, these results suggest that CSF derived LTR containing an A at position 5 of the Sp1III are associated with increased viral load compared to those containing a T at this position. However, the increased CSF viral load may not be due to the presence of the A5T mutation alone as other mutations within the LTR may also be a contributing factor since this mutation did not show the same effect in plasma.

4.6.2 CSF derived LTR containing G or A at position 2 in the Sp1III binding site has no association with viral load.

Next, we wanted to compare the viral load of all the CSF and plasma derived LTR exhibiting the G at position 2 of the Sp1III binding site in combination with other mutations to the CSF and plasma derived LTR exhibiting an A at the same position.

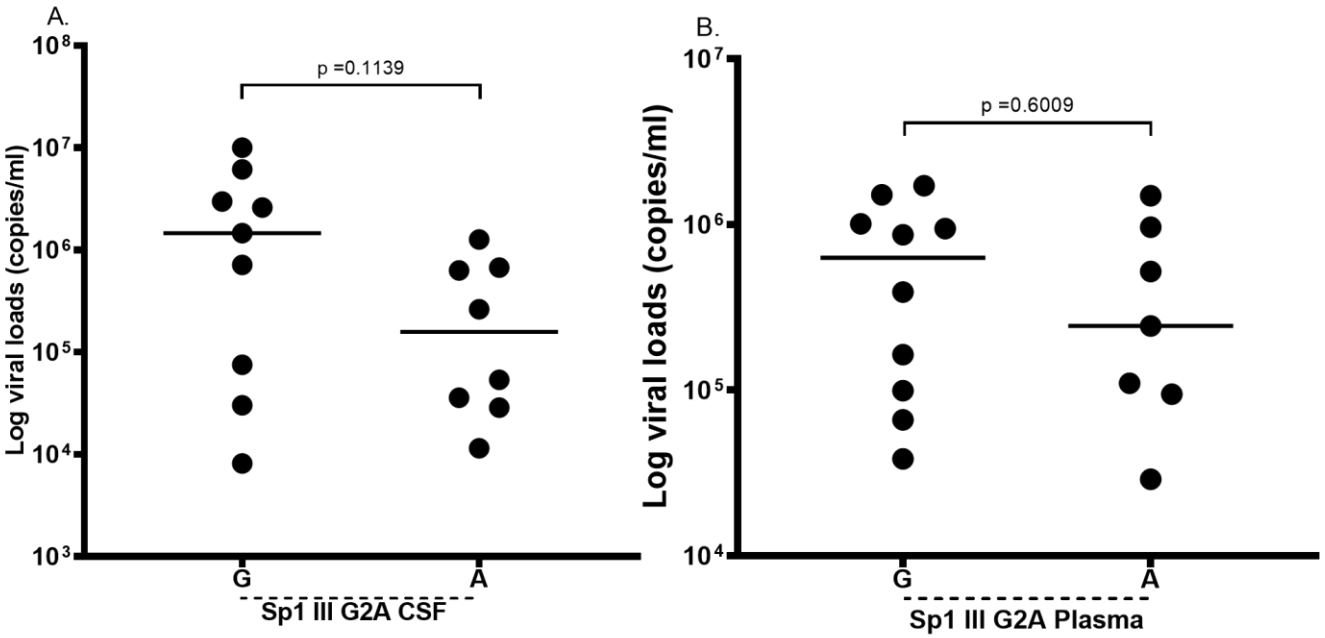


Figure 4.6.2. HIV-1 viral load of CSF and plasma derived LTR containing the G2A mutation. Panel A represents the HIV-1 viral load in the CSF compartment whereas panel B represents HIV-1 viral load in the plasma compartment. LTR containing the G2A mutation were selected and analysed. The viral load values were taken from the HIV Pathogenesis Programme (HPP) database from the TBM cohort. CSF derived LTR containing a G at position 2 in the Sp1III transcription binding site showed no significant difference (p=0.1139) in HIV-1 viral load compared to an A. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

The data show that CSF derived LTR containing a G at position 2 of the Sp1III binding site showed no significant difference in viral load (p=0.1139) compared to CSF derived LTR containing an A at the same position (Figure 4.6.2A). Plasma derived LTR containing a G at position 2 of the Sp1III binding site also showed no difference in viral loads (p=0.6009) compared to an A at the same position (Figure 4.6.2B). Taken together Figure 4.6.2A&B therefore indicate that the presence of an G or A nucleotide in the Sp1III binding site of plasma derived LTR has no association with viral load.

4.6.3 CSF derived LTR containing the A5T/G2A double mutation is associated with significantly reduced viral load compared to non-mutants.

Next, we compared the viral load of all the individuals that were infected with viruses exhibiting the A5T and G2A mutation simultaneously at position 5 and 2 (A5T/G2A) of the Sp1III binding site to the viral load of all the individuals that were infected with viruses that did not contain the

A5T/G2A mutation. The data shows that CSF derived LTR containing A5T/G2A mutation at position 5 and 2 of the Sp1III binding has significantly reduced viral loads (p=0.0061) compared to no mutation at position 5 and 2 (Figure 4.6.3A). However, this may be due to the small sample number of individuals within the A5T/G2A mutation group.

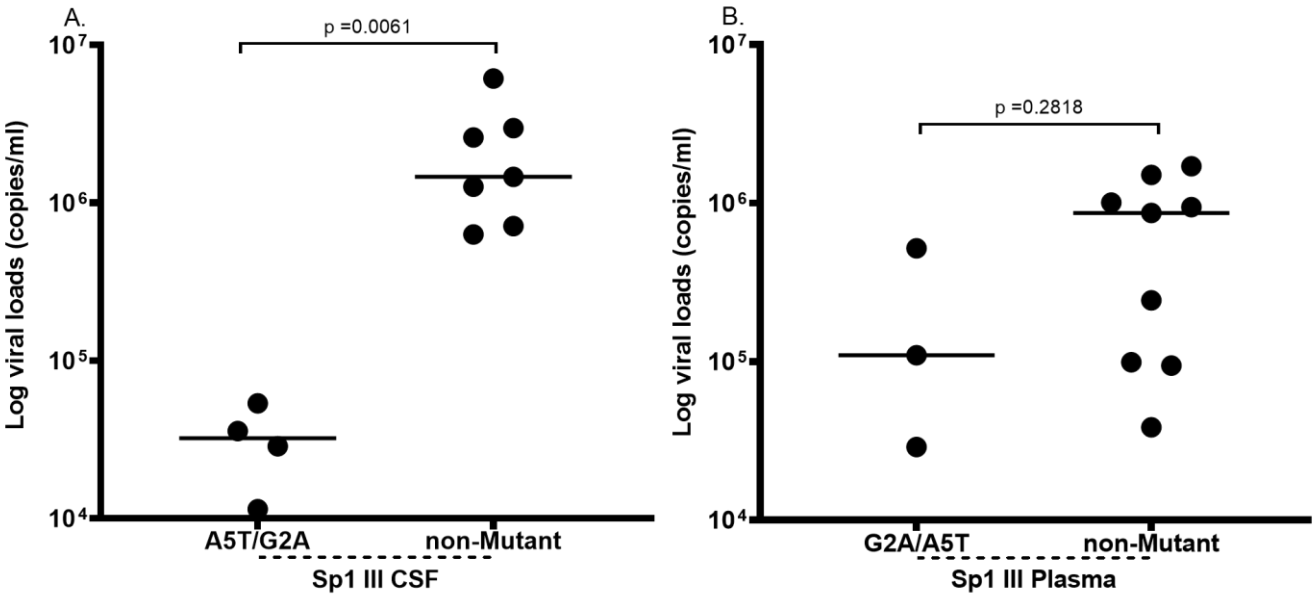


Figure 4.6.3 HIV-1 viral load of CSF and plasma derived LTR containing the A5T/G2A combination mutation compared to non-mutants. Panel A represents the HIV-1 viral load in the CSF compartment whereas panel B represents HIV-1 viral load in the plasma compartment. LTR containing the A5T/G2A double mutation were selected against non-mutants. The viral load values were from the HIV Pathogenesis Programme (HPP) database from the TBM cohort. CSF derived LTR with the A5T/G2A have significantly lower HIV-1 viral load compared to non-mutants. A value of $p<0.05$ is considered to be significant by Mann Whitney test.

Plasma derived LTR containing the A5T/G2A at position 5 and 2 of the Sp1III binding showed no significant difference in viral load (p=0.2818) compared to non-mutants (Figure 4.6.3B). Taken together, these results suggest that CSF derived LTR exhibiting the A5T/G2A mutation in combination is associated with reduced viral load compared to non-mutants.

CHAPTER 5: DISCUSSION

In this study we hypothesized that the central nervous system (CNS) derived long terminal repeat (LTR) genetic variation may be the driver of higher viral load in the CNS compared to plasma compartment of Tuberculous meningitis (TBM) co-infected individuals. The data presented in this study support the notion of LTR genetic variation driving higher viral load in the CNS compared to plasma compartment of TBM co-infected individuals.

Our study shows that matched cerebrospinal fluid (CSF) and plasma derived LTR from each patient clustered together demonstrating that the CSF and plasma LTR were isolated from the same participant and that there was no cross contamination or intermingling. Interestingly, 63% of the sequences showed branch length differences between CSF and plasma derived LTR, suggesting genetic variation between CSF and plasma derived LTR sequences. However, these genetic variations were not observed in the generated multiple sequence alignment. A previous study by Ohagen *et al*. showed that the genetic evolution of human immunodeficiency virus type 1 (HIV-1) within the brain is genetically distinct from that in lymphoid tissues and other organs in mono-infected individuals (Ohagen et al., 2003). A subsequent study consistently reported that the CNS compartment is characterized by the compartmentalization of unique viral variants demonstrated at the level of *env, gag* and *pol* sequences (Gianella et al., 2016). Collectively, the data from these previous studies could possibly support genetic variation between CSF and plasma derived LTR.

Previous studies have identified that HIV-1 ribonucleic acid (RNA) is marginally lower in the CNS than in plasma of antiretroviral treatment naïve individuals (Valcour et al., 2012). Paradoxically, there is a higher HIV-1 viral load in the CSF than plasma of treatment naïve people living with HIV (PLHIV) who are co-infected with TBM (Morris et al., 1994, Seipone et al., 2018). However, the mechanisms responsible for higher viral load in CSF of TBM co-infected individuals is still unknown. The HIV-1 replication dynamics in the CNS of TBM co-infected individuals and the factors which may mediate viral replication in this compartment remains elusive, especially for subtype C which dominates in sub-Saharan Africa. In an attempt to decipher the mechanisms underlining the phenomenon of increased viral replication in the CNS of TBM co-infected individuals, we investigated the effect of the HIV-1 5’ LTR genetic variation and functional differences on viral replication in this compartment.

While previous studies showed that inter- and intra- subtype LTR genetic variation exist within the LTR region (Jeeninga et al., 2000, Boullosa et al., 2014, Bachu et al., 2012a), our multiple sequence alignment showed no specific genetic variation between the CSF and plasma derived LTR. This suggest that the CNS and peripheral of TBM co-infected study participants was infected with the same HIV-1 LTR variant. This may be due to the disruption of the blood brain barrier (BBB) which results to pleocytosis that has been previously suggested to take place in TBM co-infected individuals, facilitating the trafficking of viruses between the two compartments (Seipone et al., 2018). Furthermore, the data from this study demonstrate that the transcription factor binding site (TFBS) including the USF, RBE III, Sp1 (Sp1I, Sp1II and Sp1III) sites, TATA Box, E- Box, H NF-κB and HH NF-κB and C- κB and TAR region aligned correspondently with the TFBS of the Indian subtype C reference sequence (EF 17 8613.1). This proved that viruses isolated from the TBM co-infected individuals were indeed of subtype C HIV-1 viruses.

Previous studies have characterized subtype C viruses with the presence of three and even four NF-κB binding site circulating in India, Mozambique and South Africa (Bachu et al., 2012b, Boullosa et al., 2014, Obasa et al., 2019). Indeed, our study showed that the CSF and plasma derived LTR from TBM co-infected individuals were subtype C viruses since the presence of 3 NF-κB binding sites specific to subtype C was observed. However, none of the LTR examined in this study exhibited the 4th NF-κB binding site. The NF-κB (H NF-κB and HH NF-κB) and the RBE III TFBS have been reported previously to be highly conserved (Mbondji-Wonje et al., 2018); however, this was not observed in our sequences. The data showed an insertion at the end of the canonical sequence of HH NF-κB “CTGCTGACA” in one participant.

The Sp1I binding site in our data showed nucleotide polymorphism at position 9 and 10 from Cytosine (C) to Thymine (T) and T to C. These mutations are African specific subtype C LTR sequences (Jeeninga et al., 2000). The Sp1II did show a mutation at position 10 from T to C. However, this mutation has not been shown by previous studies instead a mutation at position 5 from C to T has been identified (Nonnemacher et al., 2004). Considerable genetic variation was observed in the Sp1III binding site. The Sp1III binding site show notable genetic variation at position 2 Guanine(G) to Adenine(A) (G2A), and position 5 A to T (A5T). Previous studies have also investigated variation within the Sp1 binding sites and they were able to show that there is an occurrence of genetic change in position 5 of the Sp1III binding site from a C to T in HIV-1 subtype B mono-infected individuals (Qu et al., 2016, Nonnemacher et al., 2004).

The TAR element was variable exhibiting subtype C specific polymorphism at position 6 from G to A (Jeeninga et al., 2000). Within the TAR element nucleotide changes at position 24 and 30 from a C to T were observed. A nucleotide change from G to A at position 50 was also seen within the TAR region. The TAR element is known to be highly conserved (Mbondji-Wonje et al., 2018). Lastly, the genetic variation that has been reported previously to exist within the TATA box of CRF01_AE (TATAA to TAAAA) (Montano et al., 1998) (Jeeninga et al., 2000), was not observed in the subtype C LTR sequences of this study. Even with the E-Box, the genetic mutation previously observed on subtype G strains 5' E box from T to A at position 5, which occurs together with mutation on 3' E box from T to C at the same position (Baar et al., 2000), was not present in the LTR sequences analysed in this study.

The study was able to demonstrate that CSF derived LTR from TBM co-infected individuals exhibit significantly increased basal ($p < 0.0001$) and Tat ($p < 0.0001$) induced transcriptional activity compared to plasma derived LTR in SVG cell line, respectively. Gray *et al*. observed significantly reduced basal transcriptional activity from CNS derived LTR in both SVG's and T cells compared to that of non-CNS derived LTR obtained from HIV-1 mono-infected individuals (Gray et al., 2013). Taken together, these results clearly suggest that TBM co-infection may trigger enhanced transcriptional activity in CSF derived LTR compared to plasma derived LTR, which is contrarily to mono-infected individuals. The study has a limitation worth mentioning such that the lack of crucial normalization controls for transfection efficiency by co-transfecting a renilla luciferase construct and for cell numbers assayed by measuring protein content via Bradford in the different experiments, makes these results to be inconclusive.

Consistent with previous studies (Mbondji-Wonje et al., 2018, McAllister et al., 2000, Nonnemacher et al., 2004, Zhang et al., 1997), our data from this study demonstrate that the Sp1I and Sp1II sequences are relatively conserved, while the Sp1III sequence was the most variable. Specifically, our data show that position 2 and 5 is where the Sp1III binding site exhibited variability. These mutations were nucleotide changes at position 2 from a G to an A (G2A), and position 5 from an A to T (A5T). We therefore wanted to see how CSF and plasma derived LTR transcriptional activity differ according to the G2A/A5T mutation. Our study revealed that CSF derived LTR containing an A at position 5 of the Sp1III binding site have significantly high

transcriptional activity at basal and Tat induced transcriptional activity compared to CSF derived LTR containing a T at the same position in Astrocyte (SVG's) cell line.

On the other hand, plasma derived LTR showed no transcriptional response differences with this mutation at basal and Tat induced transcription activity. These results suggested that an A at position 5 of the Sp1III binding site enhances CSF derived LTR transcriptional activity compared to a T at the same position in SVG's cell lines. Previous studies were able to demonstrate that single nucleotide changes of the Sp1III binding site can impair basal gene transcription (McAllister et al., 2000). The presence of a T at position 5 of the Sp1III binding site have been observed previously in mono-infected individuals to cause reduced to no binding affinity of Sp1 transcription factors to the Sp1III binding site. The reduced binding affinity was suggested to have been the cause of decreased LTR transcriptional activity (Nonnemacher et al., 2004, Qu et al., 2016).

Furthermore, our data demonstrate that both CSF and plasma derived LTR containing a G at position 2 of the Sp1III binding site has no significant difference in transcriptional activity compared to LTR containing an A at this position in SVG cell lines. From this, we can learn that the G2A mutation observed in CSF and plasma derived LTR from TBM co-infected individuals has no effect in the transcriptional activity in SVG's cell line. Limited studies have documented the G2A mutation and its effect on LTR transcriptional activity. Again, our study was able to demonstrate that CSF and plasma derived LTR with the combination of the G2A/A5T mutation have no difference in transcriptional activity compared to the LTR with no mutation occurring at the Sp1III binding site at position 2 and 5 in SVG's.

From these data, we learn that CSF derived LTR has enhanced transcriptional activity and that an A at position 5 of the Sp1III binding site enhances CSF derived LTR transcriptional activity compared to a T at the same position in SVG cell line. However, this is not seen with plasma derived LTR. This shows that the enhanced transcriptional activity observed in CSF derived LTR is not only associated with the A5T mutation but there may also be other mutations within the LTR which could be a contributing factor.

Our data demonstrate that CSF derived LTR from TBM co-infected individuals does not show any differential basal transcriptional activity compared to plasma derived LTR ($p=0.3758$) in Jurkat cell line. Interestingly, CSF derived LTR from TBM co-infected individuals exhibit significantly higher Tat induced transcriptional activity compared to plasma derived LTR ($p=0.0033$). These data suggest that TBM co-infection is associated with significantly increased Tat induced CSF derived LTR transcriptional activity compared to Tat induced plasma derived LTR transcriptional activity in Jurkat cell lines. Conversely, Gray et al. observed significantly reduced basal transcriptional activity from CNS derived LTR in Jurkat cells compared to that of non-CNS derived LTR obtained from HIV-1 subtype B mono-infected individuals (Gray et al., 2013). Taken together, these data suggest that TBM co-infection may trigger increased transcriptional activity in CSF derived LTR compared to plasma derived LTR in Jurkat cells.

Next, we looked at the effect of the A5T mutation occurring at position 5 of the Sp1III binding site on CSF and plasma derived LTR transcriptional activity in Jurkat cell line. Our data show that CSF derived LTR containing an A at position 5 of the Sp1III binding site exhibit significantly higher transcriptional activity compared to CSF derived LTR containing a T at the same position in Jurkat cell line under basal ($p<0.0347$) and Tat ($p=0.0021$) induced conditions, respectively. Plasma derived LTR containing an A at position 5 of the Sp1III binding site showed a trend towards higher LTR basal transcription activity compared to a T at the same position ($p=0.0685$) in Jurkat's. This was not the case under Tat induced conditions, plasma derived LTR containing an A at position 5 of the Sp1III binding site showed no differential transcription activity compared to a T at the same position ($p=0.2448$) in Jurkat's. These data shows that an A at position 5 of the Sp1III binding site enhances CSF derived LTR transcriptional activity compared to a T at the same position in Jurkat cell lines, which is not seen in plasma derived LTR. This indicate that the enhanced LTR activity observed is not due to the A5T mutation alone however, there may be other mutation within the LTR causing this effect.

Furthermore, our data show CSF and plasma derived LTR containing a G at position 2 of the Sp1III binding site has no basal or Tat induced transcriptional activity difference when compared to an A at the same position in Jurkat cell line. When we investigated the A5T/G2A combination mutation, our data demonstrated that CSF derived LTR containing the A5T/G2A mutation at the Sp1III binding site at position 5 and 2 had no differential LTR basal transcriptional activity compared to

non-mutants. Interestingly, under Tat conditions, non-mutant CSF derived LTR showed significantly ($p=0.0424$) higher transcriptional activity compared to double A5T/G2A mutation in Jurkat's. On the other hand, the presence of the A5T/G2A double mutation in the Sp1III binding site at position 5 and 2 does not have an effect on transcriptional activity in plasma derived LTR when compared to non-mutants. These data may suggest that when the A5T/G2A mutation occurs at the same time, it reduces the LTR transcriptional activity in Jurkat cell line. Furthermore, the A5T mutation when assessed on its own resulted in significant increase in CSF LTR transcriptional activity, however when it is occurring at the same time with the G2A mutation the high transcriptional activity observed is reduced. This may suggest that the G2A mutation dampens that the A5T effect on LTR transcription.

In summary, CSF derived LTR from TBM individuals exhibited significantly higher transcriptional activity compared to the plasma derived LTR in both SVG's and Jurkat cell lines. It was observed that transcriptional activity is higher in SVG's compared to Jurkat cell line. These findings may suggest that the increased LTR transcriptional activity observed from the CSF derived LTR in SVG and Jurkat cell line is cell dependent, and the CSF derived LTR demonstrated increased LTR transcriptional activity compared to plasma derived LTR. The cause of this effect requires further investigation. The A5T genetic variation may explain the transcriptional difference observed, since this specific mutation observed in the Sp1III binding site demonstrated enhance CSF derived LTR transcriptional activity in SVG's and Jurkat cell lines. The A5T mutation in plasma derived LTR did not show a similar trend in both cell lines. These data therefore indicate that there has to be additional variation within the LTR that determines the transcriptional difference. Furthermore, the combination of the A5T/G2A mutation was observed to significantly reduce CSF derived LTR transcriptional activity compared to non-mutants in Jurkat cell line, such may indicate that when the A5T/G2A mutation occurs at the same time, it reduces the CSF LTR transcriptional activity.

Next, we wanted to associate the observed transcriptional activity of CSF and plasma derived LTR from TBM co-infected individuals with viral load. We hypothesized that LTR with high transcriptional activity may result into increased viral load.

In Table 1, it was clearly demonstrated that the CSF compartment of TBM individuals is associated with higher median viral load (630 291 copies/ml) compared to the plasma compartment (390 650 copies/ml). However, we were not able to see a significant difference such that only 7/17 TBM co-infected individuals were observed to have higher viral load in the CSF compared to the plasma compartment. The remaining individuals had higher plasma viral load compared to the CSF. This may indicate these individuals had not permeabilized. In this case we failed to significantly link our findings with previous studies showing that the CSF of TBM co-infected individuals is associated with significantly higher levels of HIV-1 viral load (Morris et al., 1998, Seipone et al., 2018). Seipone *et al.* wanted to see if there were virological and host immunological biomarkers that could distinguish TBM from other meningitis however, none of the evaluated soluble immunological biomarkers could reliably distinguish between TBM from other HIV-associated meningitides (Seipone et al., 2018).

We therefore examined genetic variation with the LTR as a biomarker of disease progression in HIV/TBM co-infected individuals. Our data show that CSF derived LTR containing A at position 5 of the Sp1III binding site are associated with significantly higher viral load ($p = 0.0079$) compared to a T at the same. However, this was not the case with plasma derived LTR. Taken together, these results indicate that CSF derived LTR containing the A5T mutation are associated with increased viral load compared to plasma derived LTR. From the above analysis, the CSF derived LTR containing the A5T mutation were shown to have significantly enhanced transcriptional activity and, in this case, they are further associated with increased viral load which supports our hypothesis. These findings suggest that the Sp1III A5T mutation may be the cause of high viral replication and a predictor of disease outcome in the CNS of TBM co-infected individuals compared to the plasma compartment.

Interestingly, in our study we observed that there is significantly lower HIV-1 viral load which resulted from CSF ($p=0.0061$) derived LTR containing the A5T/G2A combination mutation compared to LTR with no mutations at these specific positions. From the above analysis, it was observed that CSF LTR containing the A5T/G2A mutation resulted to reduced Tat transcriptional activity compared to non-mutants in Jurkat's. Such results clearly demonstrate that the occurrence

of the A5T/G2A combination mutation within the Sp1III binding site reduces transcriptional activity and is further associated with reduced viral load.

In summary, our data showed that CSF derived LTR showed enhanced transcriptional activity in SVG's and Jurkat cell lines and were further shown to be associated with increased HIV-1 viral load compared to plasma derived LTR. Likewise, CSF derived LTR containing the A5T mutation in the Sp1III binding site showed increased transcriptional activity and was further associated with increased HIV-1 viral load. The CSF derived LTR containing the A5T/G2A double mutation that showed reduced transcriptional activity also resulted in lower HIV-1 viral load compared to non-mutants.

The limitation to this study was the small sample size. Another major limitation of the study was the inconclusive transcriptional activity results due to the lack of appropriate normalization controls such as the renilla luciferase construct to control for transfection efficiency and the Bradford assay to control protein content in the different experiments. However, our study shows the importance of LTR diversity in study participants co-infected with TBM. These experiments will be repeated using the crucial normalization transfection controls and perform Site Directed Mutagenesis of the Sp1III A5T mutation for future manuscript publications. This will help us see if the transcriptional activity that we're seeing in these individuals is due to the A5T mutation alone, or rather other mutation within the LTR region are a contributing factor.

Conclusion

The data from our study show that there is inter-patient LTR genetic variation with intra-host evolution between the CSF and plasma compartment in 63% of the participants. The TFBS were relatively conserved between the two compartments suggesting that these compartments were infected by the same virus. CSF derived LTR exhibited higher LTR transcription activity compared to plasma derived LTR. Specifically, our data showed that the Sp1III A5T mutation in combination with other variants was associated significantly with higher transcription activity and viral loads. Future studies should investigate the effect of the single mutation A5T alone or in combination with other mutations on its ability to mediate LTR transcription activity.

CHAPTER 6: REFERENCE

ALDRICH, C. & HEMELAAR, J. 2012. Global HIV-1 diversity surveillance. *Trends in molecular medicine,* 18**,** 691-694.

ARHEL, N. 2010. Revisiting HIV-1 uncoating. *Retrovirology,* 7**,** 1-10.

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., MUKTHEY, A. B., MURALI, R. V., CHEEDARLA, N., MAHADEVAN, A., SHANKAR, S. K., SATISH, K. S., KUNDU, T. K. & RANGA, U. 2012a. Sequence insertions in the HIV type 1 subtype C viral promoter predominantly generate an additional NF-κB binding site. *AIDS research and human retroviruses,* 28**,** 1362-1368.

BACHU, M., YALLA, S., ASOKAN, M., VERMA, A., NEOGI, U., SHARMA, S., MURALI, R. V., MUKTHEY, A. B., BHATT, R. & CHATTERJEE, S. 2012b. 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.

BACHU, M., YALLA, S., ASOKAN, M., VERMA, A., NEOGI, U., SHARMA, S., MURALI, R. V., MUKTHEY, A. B., BHATT, R., CHATTERJEE, S., RAJAN, R. E., CHEEDARLA, N., YADAVALLI, V. S., MAHADEVAN, A., SHANKAR, S. K., RAJAGOPALAN, N., SHET, A., SARAVANAN, S., BALAKRISHNAN, P., SOLOMON, S., VAJPAYEE, M., SATISH, K. S., KUNDU, T. K., JEANG, K.-T. & RANGA, U. 2012c. Multiple NF-κB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. *The Journal of biological chemistry,* 287**,** 44714-44735.

BANKS, W. A., FREED, E. O., WOLF, K. M., ROBINSON, S. M., FRANKO, M. & KUMAR, V. B. 2001. Transport of human immunodeficiency virus type 1 pseudoviruses across the blood-brain barrier: role of envelope proteins and adsorptive endocytosis. *Journal of virology,* 75**,** 4681-4691.

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.

BLOOD, G. A. C. 2016. Human immunodeficiency virus (HIV). *Transfusion Medicine and Hemotherapy,* 43**,** 203.

BOSSO, M., STÜRZEL, C. M., KMIEC, D., BADARINARAYAN, S. S., BRAUN, E., ITO, J., SATO, K., HAHN, B. H., SPARRER, K. M. & SAUTER, D. 2021. An additional NF-κB site allows HIV-1 subtype C to evade restriction by nuclear PYHIN proteins. *Cell reports,* 36**,** 109735.

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.

BRUCHFELD, J., CORREIA-NEVES, M. & KÄLLENIUS, G. 2015. Tuberculosis and HIV coinfection. *Cold Spring Harbor perspectives in medicine,* 5**,** a017871.

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. *Journal of Experimental Medicine,* 188**,** 2113-2125.

BURDICK, R. C., LI, C., MUNSHI, M., RAWSON, J. M., NAGASHIMA, K., HU, W.-S. & PATHAK, V. K. 2020. HIV-1 uncoats in the nucleus near sites of integration. *Proceedings of the National Academy of Sciences,* 117**,** 5486-5493.

CANTÓ-NOGUÉS, C., SÁNCHEZ-RAMÓN, S., ÁLVAREZ, S., LACRUZ, C. & MUÑÓZ-FERNÁNDEZ, M. Á. 2005. HIV-1 infection of neurons might account for progressive HIV-1-associated encephalopathy in children. *Journal of molecular neuroscience,* 27**,** 79-89.

CAPOFERRI, A. A., LAMERS, S. L., GRABOWSKI, M. K., ROSE, R., WAWER, M. J., SERWADDA, D., GRAY, R. H., QUINN, T. C., KIGOZI, G. & KAGAAYI, J. 2020. Recombination Analysis of Near Full-Length HIV-1 Sequences and the Identification of a Potential New Circulating Recombinant Form from Rakai, Uganda. *AIDS research and human retroviruses,* 36**,** 467-474.

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.

CHAVALI, S. S., BONN-BREACH, R. & WEDEKIND, J. E. 2019. Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. *Journal of Biological Chemistry,* 294**,** 9326-9341.

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.

CHRIST, F. & DEBYSER, Z. 2013. The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy. *Virology,* 435**,** 102-109.

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.

CHURCHILL, M. J., WESSELINGH, S. L., COWLEY, D., PARDO, C. A., MCARTHUR, J. C., BREW, B. J. & GORRY, P. R. 2009. Extensive astrocyte infection is prominent in human immunodeficiency virus–associated dementia. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society,* 66**,** 253-258.

CLAVEL, F., GUYADER, M., GUÉTARD, D., SALLÉ, M., MONTAGNIER, L. & ALIZON, M. 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature,* 324**,** 691-695.

COFFIN, J., HAASE, A., LEVY, J. A., MONTAGNIER, L., OROSZLAN, S., TEICH, N., TEMIN, H., TOYOSHIMA, K., VARMUS, H. & VOGT, P. 1986. Human immunodeficiency viruses. *Science,* 232**,** 697-697.

COFFIN, J. M., HUGHES, S. H. & VARMUS, H. E. 1997. Retroviruses.

COLIN, L. & VAN LINT, C. 2009. Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. *Retrovirology,* 6**,** 1-29.

COLIN, L., VERDIN, E. & LINT, C. V. 2014. HIV-1 chromatin, transcription, and the regulatory protein Tat. *Human Retroviruses.* Springer.

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.

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.

DEBAISIEUX, S., RAYNE, F., YEZID, H. & BEAUMELLE, B. 2012. The ins and outs of HIV-1 Tat. *Traffic,* 13**,** 355-363.

DHARAN, A., BACHMANN, N., TALLEY, S., ZWIKELMAIER, V. & CAMPBELL, E. M. 2020. Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nature Microbiology,* 5**,** 1088-1095.

DUTILLEUL, A., RODARI, A. & VAN LINT, C. 2020. Depicting HIV-1 transcriptional mechanisms: a summary of what we know. *Viruses,* 12**,** 1385.

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., POSADA, D., PEETERS, M., PYBUS, O. G. & LEMEY, P. 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science (New York, N.Y.),* 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.

FAUQUET, C. M. 1999. TAXONOMY, CLASSIFICATION AND NOMENCLATURE OF VIRUSES. *Encyclopedia of Virology***,** 1730-1756.

FERGUSON, M. R., ROJO, D. R., VON LINDERN, J. J. & O'BRIEN, W. A. 2002. HIV-1 replication cycle. *Clinics in laboratory medicine,* 22**,** 611-635.

FREED, E. O. 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology,* 251**,** 1-15.

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

GALLAGHER, J. 2014. Aids: Origin of pandemic'was 1920s Kinshasa'. *BBC News,* 2.

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

GALLO, R. C., SALAHUDDIN, S. Z., POPOVIC, M., SHEARER, G. M., KAPLAN, M., HAYNES, B. F., PALKER, T. J., REDFIELD, R., OLESKE, J. & SAFAI, B. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *science,* 224**,** 500-503.

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

GHISLAIN, M. R., MUSHEBENGE, G.-A. A. & MAGULA, N. 2021. Cause of hospitalization and death in the antiretroviral era in Sub-Saharan Africa published 2008–2018: A systematic review. *Medicine,* 100.

GOMEZ, C. & HOPE, T. J. 2005. The ins and outs of HIV replication. *Cellular microbiology,* 7**,** 621-626.

GRAY, L., COWLEY, D., WELSH, C., LU, H., BREW, B. J., LEWIN, S., WESSELINGH, S., GORRY, P. & CHURCHILL, M. 2016. CNS-specific regulatory elements in brain-derived HIV-1 strains affect responses to latency-reversing agents with implications for cure strategies. *Molecular psychiatry,* 21**,** 574-584.

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.

GRILL, M. F. & PRICE, R. W. 2014. Central nervous system HIV-1 infection. *Handbook of clinical neurology,* 123**,** 487-505.

GROEN, J. N. & MORRIS, K. V. 2013. Chromatin, non-coding RNAs, and the expression of HIV. *Viruses,* 5**,** 1633-1645.
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.
GUZMAN, P. R. 2015. *Cellular host factors involved in the translation of the HIV-1 genomic RNA.* Ecole normale supérieure de lyon-ENS LYON.
HAENEL, F. & GARBOW, N. 2014. Cell counting and confluency analysis as quality controls in cell-based assays. *Multimode Detection***,** 1-5.
HEMELAAR, J., ELANGOVAN, R., YUN, J., DICKSON-TETTEH, L., FLEMINGER, I., KIRTLEY, S., WILLIAMS, B., GOUWS-WILLIAMS, E., GHYS, P. D. & ALASH'LE G, A. 2019. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *The Lancet infectious diseases,* 19**,** 143-155.
HILL, M., TACHEDJIAN, G. & MAK, J. 2005. The packaging and maturation of the HIV-1 Pol proteins. *Current HIV research,* 3**,** 73-85.
HU, W.-S. & HUGHES, S. H. 2012. HIV-1 reverse transcription. *Cold Spring Harbor perspectives in medicine,* 2**,** a006882.
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**,** 1-12.
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.
JOSEPH, S. B., ARRILDT, K. T., STURDEVANT, C. B. & SWANSTROM, R. 2015. HIV-1 target cells in the CNS. *Journal of neurovirology,* 21**,** 276-289.
KARRIS, M. A. & SMITH, D. M. 2011. Tissue-specific HIV-1 infection: why it matters. *Future virology,* 6**,** 869-882.
KIERNAN, R. E., VANHULLE, C., SCHILTZ, L., ADAM, E., XIAO, H., MAUDOUX, F., CALOMME, C., BURNY, A., NAKATANI, Y. & JEANG, K.-T. 1999. HIV-1 tat transcriptional activity is regulated by acetylation. *The EMBO journal,* 18**,** 6106-6118.
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.
KONSMAN, J. P., DRUKARCH, B. & VAN DAM, A.-M. 2007. (Peri) vascular production and action of pro-inflammatory cytokines in brain pathology. *Clinical science,* 112**,** 1-25.
KREBS, F. C., HOGAN, T. H., QUITERIO, S., GARTNER, S. & WIGDAHL, B. 2001. Lentiviral LTR-directed expression, sequence variation, and disease pathogenesis. *HIV sequence compendium,* 2001**,** 29-70.
LAFORTUNE, L., NALBANTOGLU, J. & ANTEL, J. P. 1996. Expression of tumor necrosis factor α (TNFα) and interleukin 6 (IL-6) mRNA in adult human astrocytes: comparison with adult microglia and fetal astrocytes. *Journal of Neuropathology & Experimental Neurology,* 55**,** 515-521.
LI, G.-H., ANDERSON, C., JAEGER, L., DO, T., MAJOR, E. O. & NATH, A. 2015. Cell-to-cell contact facilitates HIV transmission from lymphocytes to astrocytes via CXCR4. *AIDS (London, England),* 29**,** 755.
LIDDELOW, S. & BARRES, B. 2015. SnapShot: astrocytes in health and disease. *Cell,* 162**,** 1170-1170. e1.

LIU, X., KRAUS, W. L. & BAI, X. 2015. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends in biochemical sciences,* 40**,** 516-525.

MAJOR, E. O., MILLER, A. E., MOURRAIN, P., TRAUB, R. G., DE WIDT, E. & SEVER, J. 1985. Establishment of a line of human fetal glial cells that supports JC virus multiplication. *Proceedings of the National Academy of Sciences,* 82**,** 1257-1261.

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.

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.

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

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

MORRIS, L., SILBER, E., SONNENBERG, P., EINTRACHT, S., NYOKA, S., LYONS, S. F., SAFFER, D., KOORNHOF, H. & MARTIN, D. J. 1998. High human immunodeficiency virus type 1 RNA load in the cerebrospinal fluid from patients with lymphocytic meningitis. *Journal of infectious diseases,* 177**,** 473-476.

MÜLLER, T. G., ZILA, V., PETERS, K., SCHIFFERDECKER, S., STANIC, M., LUCIC, B., LAKETA, V., LUSIC, M., MÜLLER, B. & KRÄUSSLICH, H.-G. 2021. HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *Elife,* 10**,** e64776.

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.

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. *Journal of neuroimmunology,* 157**,** 39-47.

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.

OHAGEN, A., DEVITT, A., KUNSTMAN, K. J., GORRY, P. R., ROSE, P. P., KORBER, B., TAYLOR, J., LEVY, R., MURPHY, R. L. & WOLINSKY, S. M. 2003. Genetic and functional analysis of full-length human immunodeficiency virus type 1 env genes derived from brain and blood of patients with AIDS. *Journal of virology,* 77**,** 12336-12345.

OOMS, M., ABBINK, T. E., PHAM, C. & BERKHOUT, B. 2007. Circularization of the HIV-1 RNA genome. *Nucleic acids research,* 35**,** 5253-5261.

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.

PERKINS, N. D., EDWARDS, N. L., DUCKETT, C. S., AGRANOFF, A. B., SCHMID, R. M. & NABEL, G. J. 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. *The EMBO journal,* 12**,** 3551-3558.

POESCHLA, E. M. 2008. Integrase, LEDGF/p75 and HIV replication. *Cellular and molecular life sciences,* 65**,** 1403-1424.

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**,** 1-11.

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.

RAMRUTHAN, J. 2018. *Genetic and functional diversity of central nervous system (CNS) derived Human Immunodeficiency Virus type 1 (HIV-1) tat from Tuberculous Meningitis (TBM) patients.*

RIO-HORTEGA, P. 1939. The microglia. *The Lancet,* 233**,** 1023-1026.

ROGERS, L., OBASA, A. E., JACOBS, G. B., SARAFIANOS, S. G., SÖNNERBORG, A., NEOGI, U. & SINGH, K. 2018. Structural implications of genotypic variations in HIV-1 integrase from diverse subtypes. *Frontiers in microbiology***,** 1754.

ROJAS-CELIS, V., VALIENTE-ECHEVERRÍA, F., SOTO-RIFO, R. & TORO-ASCUY, D. 2019. New challenges of HIV-1 infection: how HIV-1 attacks and resides in the central nervous system. *Cells,* 8**,** 1245.

ROTTMAN, J. B., GANLEY, K. P., WILLIAMS, K., WU, L., MACKAY, C. R. & RINGLER, D. J. 1997. Cellular localization of the chemokine receptor CCR5. Correlation to cellular targets of HIV-1 infection. *The American journal of pathology,* 151**,** 1341.

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.

SAMPATHKUMAR, R., SHADABI, E. & LUO, M. 2012. Interplay between HIV-1 and host genetic variation: A snapshot into its impact on AIDS and therapy response. *Advances in virology,* 2012.

SEIPONE, I. D., SINGH, R., PATEL, V. B., SINGH, A., GORDON, M. L., MUEMA, D. M., DHEDA, K. & NDUNG'U, T. 2018. Tuberculous meningitis is associated with higher cerebrospinal HIV-1 viral loads compared to other HIV-1-associated meningitides. *PloS one,* 13.

SELYUTINA, A., PERSAUD, M., LEE, K., KEWALRAMANI, V. & DIAZ-GRIFFERO, F. 2020. Nuclear import of the HIV-1 core precedes reverse transcription and uncoating. *Cell reports,* 32**,** 108201.

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

SIKORSKI, T. W. & BURATOWSKI, S. 2009. The basal initiation machinery: beyond the general transcription factors. *Current opinion in cell biology,* 21**,** 344-351.

SPUDICH, S. & GONZÁLEZ-SCARANO, F. 2012. HIV-1-related central nervous system disease: current issues in pathogenesis, diagnosis, and treatment. *Cold Spring Harbor perspectives in medicine,* 2**,** a007120.

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

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

SUÑÉ, C. & GARCÍA-BLANCO, M. A. 1995. Sp1 transcription factor is required for in vitro basal and Tat-activated transcription from the human immunodeficiency virus type 1 long terminal repeat. *Journal of virology,* 69**,** 6572-6576.

SUZUKI, K., ISHIDA, T., YAMAGISHI, M., AHLENSTIEL, C., SWAMINATHAN, S., MARKS, K., MURRAY, D., MCCARTNEY, E. M., BEARD, M. R. & ALEXANDER, M. 2011. Transcriptional gene silencing of HIV-1 through promoter targeted RNA is highly specific. *RNA biology,* 8**,** 1035-1046.

TABASI, M., NOMBELA, I., JANSSENS, J., LAHOUSSE, A. P., CHRIST, F. & DEBYSER, Z. 2021. Role of Transportin-SR2 in HIV-1 Nuclear Import. *Viruses,* 13**,** 829.

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.

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

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.

UNAIDS 2020. 2020 Global AIDS Update: Seizing the moment. *Geneva: Joint United Nations Programme on HIV/AIDS (UNAIDS*.

VAILLANT, A. A. J. & NAIK, R. 2021. HIV-1 associated opportunistic infections. *StatPearls [Internet]*. StatPearls Publishing.

VALCOUR, V., CHALERMCHAI, T., SAILASUTA, N., MAROVICH, M., LERDLUM, S., SUTTICHOM, D., SUWANWELA, N. C., JAGODZINSKI, L., MICHAEL, N., SPUDICH, S., VAN GRIENSVEN, F., DE SOUZA, M., KIM, J., ANANWORANICH, J. & GROUP, R. S. S. 2012. Central nervous system viral invasion and inflammation during acute HIV infection. *The Journal of infectious diseases,* 206**,** 275-282.

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.

WENTE, S. R. & ROUT, M. P. 2010. The nuclear pore complex and nuclear transport. *Cold Spring Harbor perspectives in biology,* 2**,** a000562-a000562.

WILLIAMS, K. C. & BURDO, T. H. 2009. HIV and SIV infection: the role of cellular restriction and immune responses in viral replication and pathogenesis. *Apmis,* 117**,** 400-412.

WISE, J. 2014. HIV pandemic originated in Kinshasa around 1920, say scientists. *Bmj*, 349.

ZHANG, X.-J., WANG, R.-R., CHEN, H., LUO, R.-H., YANG, L.-M., LIU, J.-P., SUN, H.-D., ZHANG, H.-B., XIAO, W.-L. & ZHENG, Y.-T. 2018. SJP-L-5 inhibits HIV-1 polypurine tract primed plus-strand DNA elongation, indicating viral DNA synthesis initiation at multiple sites under drug pressure. *Scientific reports,* 8**,** 1-11.

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.