UNIVERSITY OF GHANA
COLLEGE OF BASIC AND APPLIED SCIENCES

CORRELATES OF VIRUS RELEASE IN HIV-1

ABENA ADOMAH KISSI-TWUM
(10308200)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE AWARD OF MASTER OF PHILOSOPHY IN MOLECULAR CELL BIOLOGY OF INFECTIOUS DISEASES

DEPARTMENT OF BIOCHEMISTRY, CELL AND MOLECULAR BIOLOGY

MARCH 2019

DECLARATION

I, Abena Adomah Kissi-Twum declare that the experimental work described in this project was performed by me at the Telesnitsky lab of the Microbiology and Immunology Department of the University of Michigan Medical School, under the supervision of Prof. Alice Telesnitsky (PhD.) and Dr. Osbourne Quaye.

..... March 28, 2019

Abena Adomah Kissi-Twum — Date

Student

..... March 28, 2019

Prof Alice Telesnitsky (PhD.) — Date

..... 28-05-19

Dr. Osbourne Quaye — Date

DEDICATION

This work is dedicated to my family; my husband, nuclear and extended, by and large, local and foreign. Thank you for being an incredible support system

ACKNOWLEDGEMENT

All praise and thanks to God who has brought me thus far.

I am very grateful to Prof Alice Telesnitsky for accepting to mentor me, her scientific guidance, words of encouragement, dedication and supervision of this work. I owe a debt of gratitude to Dr. Quaye for the collaborative opportunity to work with Prof Telesnitsky and for his patience throughout my time away. To the director, Prof Gordon Awandare (PhD) and staff of WACCBIP thank you for granting me an MPhil fellowship which paid my fees and funded my travel.

My sincere gratitude goes to all past and present members and collaborators of the R33 group of the Telesnitsky lab especially, Edmond Atindaana and David Read for carrying out the foundational work. I am thankful to Prof. Jeffrey Kidd (PhD.) and Ms. Sarah Emery of the Kidd lab in the Human Genetics department who provided the platforms for high throughput sequencing and bioinformatics data analysis and Dr. Kalyani Pyaram of the Chang lab in the Microbiology and Immunology department for her assistance with flow cytometry data acquisition. This work could not have been completed without the invaluable contribution and assistance of members of the Telesnitsky lab; Edmond Atindaana, Dr. Siarhei Kharytonchyk and Dr. Jackie Esquiaqui who were my sounding boards, Mr. Patrick Rimple who carried out poly A purification of RNA and related assays, Ms. Cleo Burnett who helped me with protocols and procedures and Mr. Jake Pitcher who offered assistance as and when needed. Thank you.

Ms. Precious Cramer! Where do I begin from? Mr. George Mensah - friend who loves at all times!

Last but definitely not the least, I thank all faculty, staff, students and colleagues of the 2016 intake group for MPhil. MCBI, for the words of encouragement, support, advice, all the memories we shared together and the many times you were my ‘legs’ in Ghana while I carried out my work in Michigan. Thank you all for enriching my scientific educational experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
AP-1	Activator protein 1
APOBEC3G	Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3G
ART	Antiretroviral therapy
CA	Capsid protein
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
CTL	Cytotoxic T-lymphocyte
CTP	Cytidine triphosphate
CXCR4	C-X-C chemokine receptor type 4
DIS	Dimerization initiation sequence
DNA	Deoxyribonucleic Acid
DNMT1	DNA methyltransferase-1
DSIF	DRB Sensitivity Inducing Factor
Env	HIV-1 envelope protein
FBS	Fetal Bovine Serum
G	Guanosine
Gag	HIV-1 group-specific antigen protein
GFP	Green fluorescent protein
GPP	Gag Pol Puromycin construct of HIV-1
gRNA	Genomic RNA
HAT	Histone acetyl transferase
HDAC	Histone deacetylase

HEK293T	Human Embryonic Kidney cell line clone 293T
HIV-1	Human Immunodeficiency Virus type-1
IN	HIV-1 Integrase
IRES	Internal ribosome entry site
KFS	Kpn-1 frameshift construct of HIV-1
LTR	Long Terminal Repeat
MA	Matrix protein
mRNA	Messenger RNA
NC	Nucleocapsid
Nef	Negative factor of HIV-1
NELF	Negative elongation factor
NFKB	Nuclear factor KB
NL4-3	HIV-1 NL4-3 Infectious Molecular Clone
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NRTI	Nucleoside-analog reverse transcriptase inhibitors
ORF	Open reading frame
P-TEFb	Positive transcription elongation factor
PBAF	Polybromo- and BAF containing complex
PBMC	Peripheral blood mononuclear cell
PBS	Primer binding site
PCR	Polymerase Chain Reaction
PHA	Phytohemagglutinin
PI	Protease inhibitor

Pol	HIV-1 Polymerase protein
Ppt	Polypurine tract
PR	HIV-1 protease
qPCR	Quantitative polymerase chain reaction
QVOA	Quantitative viral outgrowth assay
R	Repeat region of HIV-1 LTR
Rev	Really exasperating viral protein
RNA	Ribonucleic Acid
RNAP II	RNA polymerase II
RPA	RNase protection assay
RPMI	Rosewell Park Memorial Institute
RRE	Rev Response Element
RT	Reverse Transcriptase
RT PCR	Reverse Transcriptase Polymerase Chain Reaction
SD1	First 5' splice site located after 5' LTR
SIV	Simian Immunodeficiency Virus
SP-1	Specificity protein 1
SU	Surface protein of HIV-1 Env
SWI/SNF	Switch/Sucrose Non-Fermentable complex
TAR	Tat response element
Tat	Transactivator of transcription of HIV-1
TE	Tris EDTA buffer
TILDA	Tat/rev Induced Limiting Dilution Assay

TM	Transmembrane protein of HIV-1 Env
tRNA	Transfer RNA
U3	Unique to the 3' region
U5	Unique to the 5' region
UNAIDS	United Nations Program on HIV and AIDS
Vif	HIV-1 Virus infectivity factor protein
Vpr	Viral protein R of HIV-1
Vpu	Viral protein U of HIV-1
WHO	World Health Organization

ABSTRACT

Integration of HIV-1 within the host genome may result in latent infection, where a replication competent provirus remains in a state of transcriptional silence. The latent reservoir remains the major barrier to cure efforts as to achieve cure, the reservoir should be depleted of all replication competent proviruses. Various assays to measure the latent reservoir rely on virus release and vary widely in their estimations. Quantitative viral outgrowth assay (QVOA) which is the gold standard relies on reactivation of latent proviruses to release virions and thus falls short when proviruses are non-inducible. The correlates of virus release per cell were examined in this study.

The TZ-5 zipcoded library which is a Jurkat cell line with one uniquely tagged HIV-1 integrant per cell was used in this study. The library consisted of infected cells expressing HIV-1 *gag* and *pol* with an inactivated *env* gene and *vpr*-. GFP is expressed as a Nef spliced product to indicate HIV-1 expression. Intracellular Gag staining coupled with the relative amounts of p24 detected in culture supernatant showed GFP expression is an accurate proxy for HIV-1 expression.

High-throughput sequencing was carried out on PCR amplified zipcodes. The topmost 74 clones which represents ≥ 97% of the raw sequence reads was analysed in this study. Although cell proliferation is thought to drive HIV-1 persistence, clonal abundance was not found to correlate with high virus release. Rather a higher GFP+ expression as determined by the green proportion correlates with high virus release per cell. Spliced and unspliced mRNA products are needed to productively assemble virions in HIV-1 although Gag which is an unspliced product is the driving force. Clones with higher amounts of unspliced product to spliced had higher green proportions and subsequently overlapped with higher amounts of virus released. Taken together, higher green proportions and unspliced RNA may predict a higher amount of virus released per cell. As green proportion is a measure of overall HIV-1 expression by a clone, incorporating the extent of HIV-

1 expression in QVOA measurements will calibrate HIV-1 expression among inducible and non-inducible cells in the latent pool and ultimately the decay rate of such clones.

The frequency of formation of host-virus read-through and read-in chimeras and packaging into virions was examined as these unusual species can be helpful in integration site determination and ultimately determine genetic correlates of chimeric RNA generation. RNase protection assays showed the presence of chimeric read-through and read-in RNAs both in cells and virus. The genomic context of integration may contribute to our observed findings as more read-in RNA was observed to be produced and packaged into virions than read-through RNA.

CHAPTER ONE

INTRODUCTION

1.1 Background

The Human Immunodeficiency virus (HIV), the etiological agent of Acquired Immunodeficiency Syndrome AIDS has claimed about 35 million lives with about 60 million people infected globally. Despite the many advances made, HIV and eventually AIDS continues to be an infection of global public health concern. The World Health Organisation (WHO) estimates about 37 million people were living with HIV in the year 2017 and 1.8 million of those were new infections. The Africa region leads with the number of infected individuals at a staggering 25.7 million representing approximately 70% of the global burden. The Joint United Nations Programme on HIV/AIDS (UNAIDS factsheet, 2018) reports three out of four new infections in the Sub-Saharan region of Africa occurs amongst adolescent females between the ages of 15 to 19 years. The main route of contracting the infection remains exposure of broken skin layer to bodily fluids (mainly blood or semen) of an infected individual.

There are two types of HIV namely HIV-1 and HIV-2. HIV-1 is the most prevalent type accounting for the majority of recorded morbidity and mortality globally. However, HIV-1 infection is no longer a death sentence due to the introduction of antiretroviral therapy (ART). Most infected individuals on ART can carry on their day-to-day activities with no significant setbacks due to the infection. The WHO estimates in the year 2017, 59% of people living with HIV-1 were receiving ART. Viral suppression has been achieved in some individuals on combination ART. Current data from the UNAIDS (2018) estimates 47% of all people living with HIV-1 are virally suppressed. The cure of Timothy Ray Brown otherwise known as the Berlin patient and very recently Adam Catillejo (the London patient) demonstrates that achieving a cure for HIV-1 is possible although

the methods used in the treatment of these two patients may not be suitable for most infected individuals due to the associated risk (Brown, 2015, Gupta et al., 2019).

Integration of the virus into the host genome is known to be a non-random process with preference for actively transcribed regions (Schroder et al., 2002). Integration within the host genome fundamentally poses a challenge to virus eradication as virus can effectively be eradicated only by indefinitely silencing infections of the virus or excising it out of the genome. CD4+ T-cells which are primarily the targets of HIV-1 and are central to host immunity act as part of their functions as effector cells and memory cells of the immune system. Infected effector cells in circulation can be targeted by ARTs while memory cells on the other hand go into a resting state thus allowing the infection to evade and persist. HIV-1 infection of cells which go into long term memory poses a major challenge to cure efforts. In spite of the preference for integration in transcription hotspots, not all HIV-1 infections are productive. Some infections are latent. HIV-1 cure efforts have been severely hampered due to the limited understanding of the latent reservoir.

Upon activation of resting CD4+ T-cells from the latent reservoir, there is expansion to produce more infected daughter cells thus the need for lifelong ART even in virally suppressed individuals. Studies conducted on CD4+ T-cells from patients who were undergoing ART treatment reveal about 40% of proviruses are integrated in clonally expanded cells with oncogenes and genes for cell proliferation favoured as sites of integration (Maldarelli et al., 2014, Wagner et al., 2014). This suggest that clones which persist, expand and may eventually be part of the latent pool are proviruses integrated into proliferation genes. Cell proliferation therefore contributes to persistence of infection. However, Wang et al., (2018) in investigating replication competent latently infected CD4+ T-cells ex-*vivo* show that these cells can undergo homeostatic proliferation without virus release in response to T- cell receptor agonist or cytokines.

Earlier work carried out by Telesnitsky and associates (Read, Atindaana et al., 2019) using the zipcoded library (which allows individual tagging and tracking of provirus to virus) in order to analyse differences in HIV-1 expression due to integration site observed differences in clonal HIV-1 expression both in Jurkat and infected primary cells. Using known repressive and enhancer chromatin markers for Jurkat cells, there was no clear correlation identified for clones which expressed HIV-1 or did not. Moreover, Chen et al., (2017) using Barcoded HIV-1 ensembles which allows proviral tracking to map integration sites and transcriptional activity showed that sites within the chromosomes which were integration hotspots were not necessary expression hotspots. HIV-1 gene expression was dependent on the presence of promoters and enhancers at such sites. Integration of provirus near active transcription units however has been found to impact HIV-1 gene expression as RNA polymerase II from the preceding transcription transcribes through the HIV-1 promoter (Lenasi et al., 2008).

Pollack et al., (2017) observe the expression of HIV-1 without virus release by showing in defective proviruses with intact LTRs, HIV-1 can be transcribed and translated. In clones with mutations in the major 5' splice donor D1 of HIV-1, Gag is released into the supernatant indicating the presence of HIV-1 Gag may not be a conclusive measure of full-length HIV-1 transcription or virus release by extension. Although Gag is known to drive virion assembly, assembly of HIV-1 virions requires both spliced and unspliced RNA products generated from transcription of HIV-1 from the 5' LTR which serves as the viral promoter. One primary mRNA transcript is made in HIV-1 from which over 40 splicing products are generated outlined in figure 1.11. Figure 1.11A shows multiple 5' (donor- D) splice sites and 3' (acceptor-A) splice sites allowing for the generation of different mRNA from the primary mRNA transcript.

Each mRNA generated usually expresses one gene product as outlined in figure 1.11 B. There are three sizes for which HIV-1 mRNA can be classified (Emery et al, 2017). There is the unspliced RNA consisting of the approximately 9.3 kb primary transcript (represented by the solid purple bar in figure 1.11 D) which is either packaged into virions as genomic RNA or from which the Gag and Gag-Pol precursor proteins are generated. The second-class size is the 4kb products also known as the incompletely spliced RNA. The 4kb mRNA group includes Vif, Vpr, the single exon form of Tat, Vpu, and Env proteins. These gene products make use of the D1 5' splice site in conjunction with any of the centrally located 3' splice sites (including but not limited to A1, A2 and A3 3' splice sites). The major difference between the 4kb and the 1.8kb spliced products is the retention of the second intron in the 4kb spliced products as shown in figure 1.11 D. To be considered a fully spliced HIV-1 mRNA transcript, both introns of HIV-1 must be spliced out as shown in figure 1.11 C. The 1.8kb class of fully spliced HIV-1 expresses the Rev, Tat, Vpr and Nef proteins.

Virion assembly requires proteins translated from various mRNA transcripts across the various classes. Primarily, there is the Gag and Gag-Pol polyprotein from the unspliced mRNA, the Env protein from the singly spliced RNA pool, two copies of unspliced HIV-1 genomic RNA which gets packaged into virions. If the equilibrium of spliced to unspliced RNA in a cell is shifted in either direction, availability of gene products to assemble a virus will be affected. Moreover, GFP is expressed as a Nef splicing product in the cell implying that any mutations which affect Nef splicing will affect the expression of GFP and thus reporting for HIV-1 expression within the cell. Also, large internal deletions, splice site mutations and mutations within the HIV-1 genome not affecting Nef splicing may also lead to GFP expression without virus release.

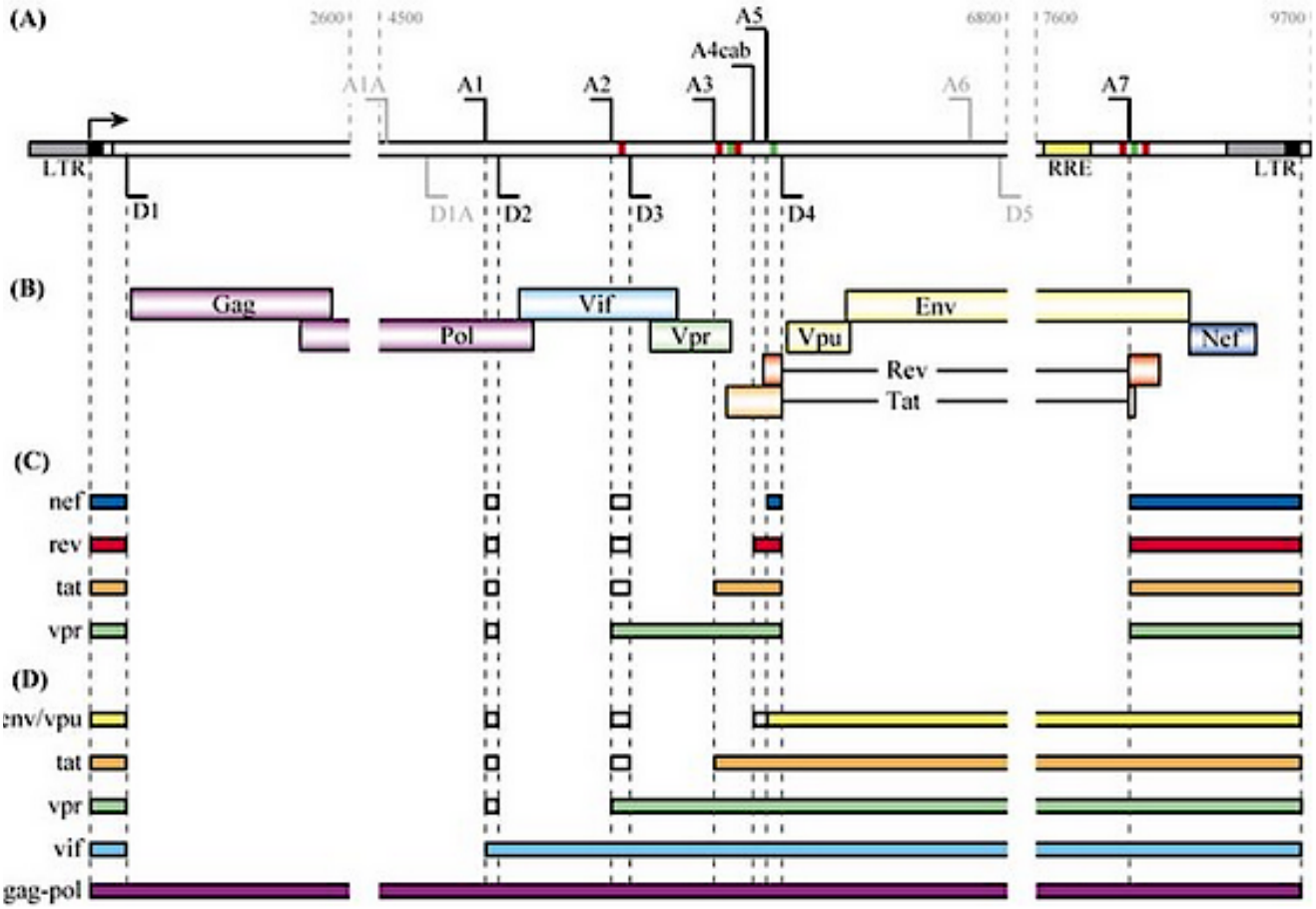

Figure 1.11 the organisation of HIV-1 proviral genome adapted from Saliou et al., (2009). A) Shows the transcription start site and the five HIV-1 5' splice sites (D1 to D5) and the 3' splice acceptor sites (A1 to A7). B) Shows HIV-1's nine open reading frames (ORFs) which code for Gag, Gag-Pol, Vif, Vpr, Vpu, Tat, Rev, Env and Nef. C) Shows fully spliced products of the 1.8 kb class. These products have both introns of HIV-1 completely spliced out. D) Incompletely spliced mRNAs of HIV-1. These retain the second intron of HIV-1. The Gag-Pol mRNA retains both introns of HIV-1.

Kharytonchyk et al., (2016) have shown that the structure of the 5' leader sequence in HIV-1 RNA determines the fate of the RNA either as genomic RNA which gets packaged into virions or as mRNA which is used in translation. This observation therefore suggests that changes which affect the RNA folding structure of the 5' leader sequence may interfere with either RNA packaging or translation and therefore the RNA may be fated for degradation. As observed both in CD4+ T-cells and in HIV-1 infected tissue culture models, host-provirus RNA chimeras have been known to be produced in HIV-1 due to the semi-random nature of provirus integration. The amount of virus released amongst clones which generate high amounts of chimeric RNA may be affected if chimeric RNA generated in cells are fated for degradation. Integration of the provirus within a gene or in close proximity to a strong promoter may cause the generation of host-provirus read-in chimeras. Improperly assembled poly adenylating factors may lead to the generation of host provirus-read-through chimeric RNA. RNA polymerase II has also been shown to by-pass the polyadenylation site (West et al., 2004) up to a 1000 bp before it falls off the DNA (Proudfoot et al., 2002). This has been proposed to occur because of improperly assembled factors for polyadenylation which inefficiently recognize the poly A signal and thus the assembly of the 3' end is not productive (Kaneko and Manley, 2005).

Figure 1.12 B shows HIV-1 provirus has two identical LTRs made up of U3-R-U5 sequences at the ends of the genome, HIV-1 transcription is expected to begin from the R region of the 5' LTR and terminate with the recognition of the polyadenylation signal in the R region of the 3' LTR as shown in figure 1.12 B. By this convention, all HIV-1 RNA should begin with R sequences and end with R sequences (figure 1.12 A). The U3 region of the 5' LTR and the U5 region of the 3' LTR are not expected to form a part of RNA species generated in cells.

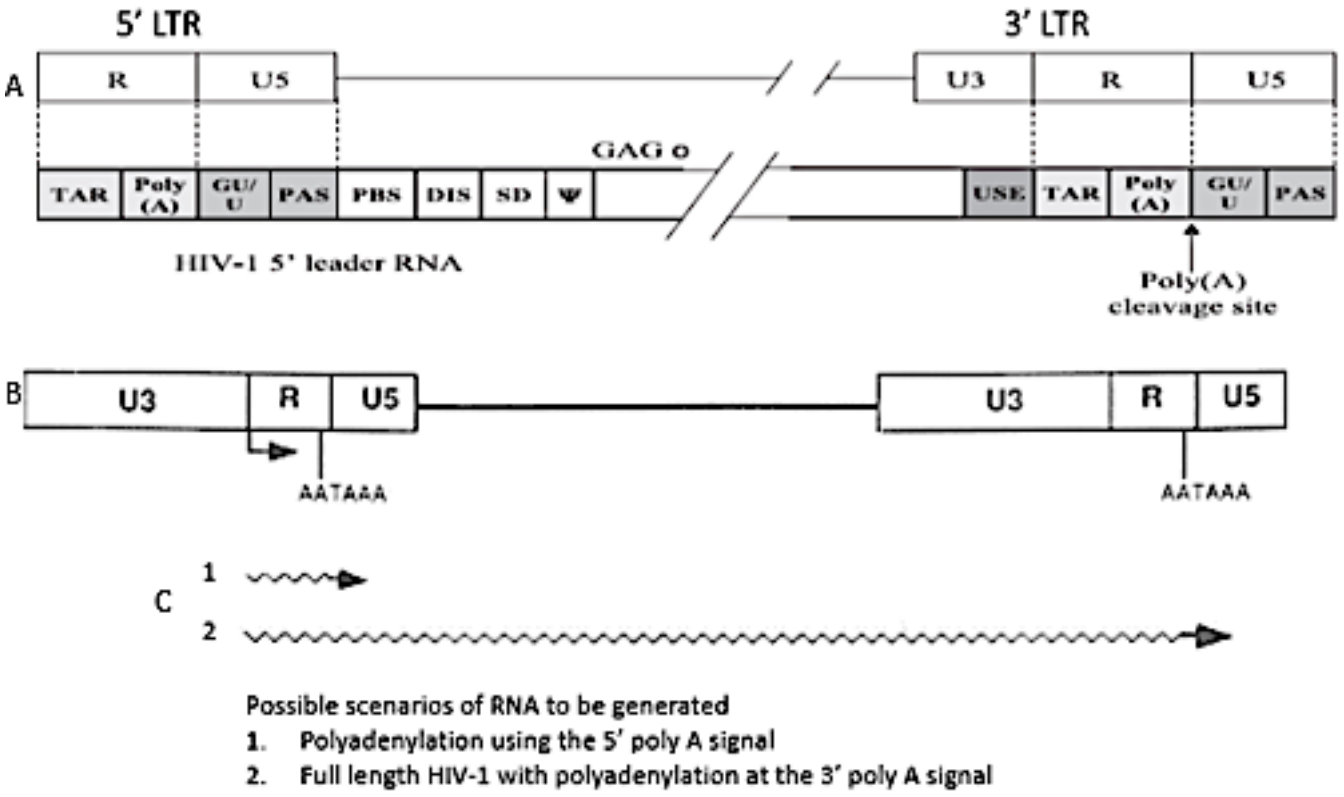

Figure 1.12 The organization of HIV-1 LTRs. A) Schematic showing the organization of the 5' and 3' LTRs. Each LTR has the Tat response element TAR, the poly adenylation signal and transcription start site indicating either of the two LTRs can be used in initiating transcription. Figure B shows transcription initiation from the 5' LTR and C shows the possible outcomes of transcription should polyadenylation occur at either LTR. The figure is adapted from Gee et al., (2006), B and C Cherrrington and Ganem (1992).

However, due to the non-random nature of HIV-1 provirus integration with preference for introns of actively transcribed genes, the provirus may integrate in close proximity to upstream transcription machinery leading to transcriptional interference. Transcriptional interference is said to occur when one transcriptional process has a *cis* effect on a second process of transcription (Greger et al., 1998). Integration of the provirus within a gene or in close proximity to a strong promoter may cause the generation of host-provirus read-through or read-in chimeras. Various scenarios have been proposed as the possible mechanism for generation of provirus-host RNA chimeras. Sequences in the U3 region of the HIV-1 genome have been implicated in upregulating polyadenylation at both homologous and heterologous poly A signals (Cherrington and Ganem, 1992). If transcription from an upstream promoter is not terminated and reads into HIV-1 or polyadenylation at the 3'LTR does not occur, the 5' LTR U3 or 3' LTR U5 sequences may form part of RNA generated within cells. Upstream transcription into HIV-1 may also be polyadenylated at the 5' LTR R region (as shown in figure 1.12 C) which has a polyadenylation signal although it is usually not used during transcription of HIV-1 genes due to the presence of a silencer sequence. An and Telesnitsky (2004) show 7% read-through RNA is packaged into virions. Viral RNA with host sequences provides an opportunity to detect sites of proviral integration using retained host sequences and also to be able to quantify how often chimeric RNA is generated. Determination of integration sites variation amongst clones which generate different chimeric RNA will also allow identification of genomic context within which chimeric RNA is generated and packaged. How much read-through and read-in chimeric RNA is made in cells and how much is packaged into virions is examined here.

1.2 Rationale

Although research over the last 30 years has broadened our understanding of HIV-1 infection, it would be better to understand properties of cells which persist in patients. HIV-1 persistence has been attributed to proliferation/clonal abundance of infected cells. The contribution of polyclonal or oligoclonal persistent clones to the virus released is of interest as it will inform which clones are important to clearance of the latent reservoir in patients. To develop assays which correctly estimate the pool of latently infected cells, it is fundamental to identify factors which explain differential virus release per cell. Using the zipcoded cell culture system, we can determine the intrinsic outcomes of HIV-1 infection in a cell culture system with regards to virus release. The fundamental question to be addressed is whether all GFP+ cells express HIV-1 and if these cells expressing HIV-1 contribute to the virus pool in similar proportions. Specifically, what factors correlate with high virus release? Does transcriptional state and rapid cell proliferation determine high virus release or the amount of unspliced RNA generated per cell determine which clones release the most virus. It is necessary to examine the effect the semi-random nature of HIV-1 integration may have on HIV-1 gene expression. Are provirus-host chimeric read-through and read-in RNA generated in cells and ultimately if these host-virus RNA are fated into virus as different conformations may be assumed by RNA with host sequences.

1.3 Aim

This study sought to identify the correlates of virus release from HIV-1 infected cells

1.4 Specific Objectives

The specific objectives were to

1. Determine if only cells expressing GFP released virus and if the extent of virus released varied among clones.
2. Determine the effect and relationship of splicing on virus released.
3. Determine if host-provirus RNA chimeras were packaged into virions.

1.5 Significance

Accurately estimating the latent viral reservoir in HIV-1 is fundamental to evaluate the efficiency of ART in reducing the size of the latent reservoir. The Quantitative Viral Outgrowth Assay (QVOA) which is the current gold standard in estimating the latent reservoir relies on the ability to reactivate latently infected cells to produce virus. Establishing factors which can account for differential virus release among clones may help in determining whether virus rebound is consistent among reactivated clones and by extension the decay rate of the latent reservoir. The distinct genomic context of integrants could be ascertained for proviruses which generate chimeric RNA but differ in amounts of chimeric RNA packaged.

CHAPTER TWO

LITERATURE REVIEW

2.1 HIV- An introduction

HIV is the causative agent of Acquired Immune Deficiency Syndrome (AIDS) which until recently was akin to a death sentence. As one of the most successful pathologies known to man, HIV has infected about 60 million people globally (Merson et al. 2008) with about 35 million mortalities (WHO, 2018). In spite of over 30 years of advances against the pandemic, about 940,000 deaths resulted from HIV infections in the year 2017 making it still a very relevant public health challenge. The WHO estimates about 37 million individuals live with HIV with 1.8 million new infections in the year 2017 alone. About 0.8% of adults (15–49-year-olds) worldwide live with HIV with the African region severely affected with 4.1% of adults living with HIV (WHO, GHO data 2018).

There are two types of HIV which have been known to cause AIDS named HIV-1 and HIV-2. While transmission routes, the intracellular mechanism of replication and susceptibility to opportunistic infections remain similar for both types of HIV, progression to the active disease state of AIDS is generally thought to be much slower in HIV-2 (Nyamweya et al., 2013). HIV-1 on the other hand is thought to be the more infectious virus due to high viral loads observed in HIV-1 patients as compared to HIV-2 infected individuals (Rowland-Jones and Whittle, 2007). The success of HIV-1 as a human pathogen has been attributed to its relatively recent introduction into humans from simian immunodeficiency viruses (SIVs) which naturally infects non-human primates mostly without any pathologies (Sharp and Hahn, 2012). Identification of similarities between the prevalent circulating SIVsmm strain found in wild and captive sooty Mangabey communities and the prevalent HIV-2 strain harboured by persons in Sierra Leone in West Africa lent credence to zoonotic source of HIV (Chen et al., 1996). Based on phylogenetic analysis, HIV-

1 strains have been classified into four groups namely M, N, O and P and HIV-2 groups A through H. Group M viruses of HIV-1 are responsible for pandemic spread (Shaw and Hunter, 2012), while groups A and B of HIV-2 have spread within humans although it is mostly confined to West Africa. Each phylogenetic related group is thought to represent an independent introduction of SIV into humans.

The major routes of transmission of HIV-1 in humans remains the introduction of the virus at mucosal surfaces, mother to baby transmission and through breakages in the skin barrier (Shaw and Hunter, 2012). Various groups which have been identified as being at an increased risk of acquiring an HIV-1 infection irrespective of the epidemic type or local context include men who have sex with men, people who inject drugs, individuals in prisons and closely interactive quarters, transgender individuals, sex workers and patrons of sex work (WHO, 2018).

HIV-1 infects primarily CD4+ T-cells of the immune system making the individual susceptible to a barrage of opportunistic infections. The current gold standard in diagnosing HIV-1 is the Enzyme-linked Immunosorbent assays (ELISAs) which detects HIV-1 antibodies from an individual's serum (Wu and Zaman, 2012). For infants, polymerase chain reaction (PCR) diagnosis is preferred in order not to incorrectly detect maternal antibodies persistent in the bloodstream. Development of rapid diagnostic tests (RDTs) which detect antibodies against HIV-1 has made diagnosing HIV-1 much easier and faster even in resource constrained areas and is often used as a first line in diagnosis (WHO, 2018). This has allowed for early treatment and care of infected individuals and to some extent reduced transmission. CD4+ T-cells count, and viral load have been used as markers of infection progression and/or control.

There are estimated about 10,000 to 100,000 HIV-1 particles per a millilitre of plasma in an HIV-1 patient who is not on treatment (Arts & Hazuda, 2012). These viruses turn over at a rate of

approximately 10^{10} particles per day (Perelson et al., 1996). While HIV-1 remains incurable, the introduction of combination antiretroviral therapy (ART) has helped achieve undetectable levels of viremia in people and reduced transmission rates. Estimates by the WHO (2018) indicate 21.7 million persons living with HIV-1 were receiving ART globally in the year 2017. There are 6 different classes of antiretroviral drugs which are used in suppressing HIV-1 infection. These include non-nucleoside reverse transcriptase inhibitors (NNRTIs), nucleoside-analog reverse transcriptase inhibitors (NRTIs), HIV-1 viral Integrase inhibitors, Protease Inhibitors (PI), Fusion inhibitors and Co-receptor antagonists. Current ART combinations employ three antiretroviral agents designated against at least two independent molecular targets (Arts and Hazuda., 2012). This approach seeks to avert the evolution of drug resistance in the virus. The success of combination ART has been credited with the approximately 40% global reduction in new HIV-1 infections and deaths (WHO, 2018).

2.2 Biology of HIV-1

HIV-1 belongs to the family of viruses referred to as retroviruses due to their ability to reverse transcribe their RNA genomes into a DNA copy which is inserted into the host genome. Within the retrovirus family, HIV-1 is a member of the genus lentiviruses- viruses which cause the disease state (AIDS in the case of HIV-1) after a relatively long period of infection. A mature HIV-1 particle has a roughly spherical lipid bilayer envelope which has a diameter of about 145 nm (Briggs et al, 2003). The gp120 surface proteins (SU) project from the envelope and are anchored in trimers by the gp41 transmembrane protein (TM). There is a conical core structure the capsid, the ends of which are in close proximity to the viral membrane (Briggs et al, 2003). The capsid membrane is formed by the matrix protein MA (p17) and the inner capsid formed by the capsid protein CA (p24). HIV-1 has a nucleocapsid (NC) which is composed mainly of basic residues and zinc fingers (Carter and Saunders, 2007). Within the core structure is HIV-1's enzymes

proteases (PR), ribonuclease, reverse transcriptase (RT) and integrase (IN) and the genomic material bound to the nucleocapsid. The genomic material is made up of two positive sense single stranded RNAs. The two copies of RNA are held in a dimer (Muriaux et al., 1996).

HIV-1's genome is approximately 9.7 kb long and contains nine open reading frames which encode fifteen different viral proteins (Fevrier et al., 2011) starting from one transcript. Two major classes of proteins are produced by HIV-1 the structural polyproteins group-specific antigen protein (Gag), Polymerase protein (Pol), HIV-1 envelope protein (Env) and regulatory Transactivating protein (Tat) and the Really exasperating viral protein (Rev). Accessory proteins Virus infectivity factor protein (Vif), protein R (Vpr), Negative factor (Nef) and protein U (Vpu) are also synthesized. The Gag polyprotein is cleaved to produce MA, CA, NC and protein p6, while Pol produces the enzymes PR, RT and IN. Cleavage of the Env polyprotein gives rise to the envelope glycoproteins SU and TM.

Upon introduction of the virus at mucosal surfaces, the HIV-1 envelope glycoprotein projection gp120 interacts with mainly CD4 receptors on target cells such as macrophages and T helper lymphocytes to establish an infection. Either one of the chemokine receptors CXCR4 or CCR5 is often recruited as a co-receptor for HIV-1 which determines the tropism of the virus as either X4 or R5 for a specific cell type. Some viruses are dual tropic meaning these viruses have the ability to initiate infection with either receptor. The cure of both the Berlin and London patients exploited the dependence of the virus on the CCR5 co-receptor to establish infection (Gupta et al., 2019, Brown, 2015). Donor stem cells with the double null CCR5 Δ32/Δ32 genotype which is not permissive of HIV-1 infection were used. The new CD4+ T-cells produced in both patients were resistant to HIV-1 with the Berlin patient recording undetectable viremia for almost a decade. This concept was further validated when virus rebounded in the two Boston patients who received bone

marrow transplants without the HIV-1 resistant phenotype (Henrich et al., 2014). Minor co-receptors used by the virus include CXCR6 (Woodham et al., 2016).

Upon attachment of the receptor moieties, there is a conformational change in gp120 which triggers a change in fusion protein gp41. Exposure of the N-terminus of the envelope glycoprotein gp41 on the viral membrane leads to the formation of a channel which is inserted into the plasma membrane fusing the viral envelope to the cell membrane. The cofactors CD26 and α4β7 catalyse the fusion of the membranes (Woodham et al., 2016). Viruses which bud off with ICAM-1incorporated into the envelope interact with LFA-1 to enhance viral infectivity (Woodham et al., 2016).

The viral core is released into the cells cytoplasm and migrate or are transported to the nucleus. Reverse transcription is activated within the cytoplasm. HIV-1s reverse transcriptase initiates the process of reverse transcription from the genomic RNA summarised in figure 2.2. Briefly, reverse transcription starts with the binding of a transfer RNA (tRNA) lysine primer which binds its complementary sequences at the primer binding site on the genomic RNA (gRNA) in figure 2.2 A. This process initiates the minus strand DNA synthesis until the minus strand strong stop DNA is generated at the 5' end of the gRNA (shown in figure 2.2 B). Degradation of the gRNA by RNase H in the RNA-DNA heteroduplex leaves the minus strand DNA which undergoes a template switch by binding to complementary sequences in the 3' R region (figure 2.2 C). The minus strand is continuously synthesized from the 3' end. Synthesis of the plus strand begins from the polypurine tracts (ppt) which are oligo ribonucleotides which resist RNase H degradation (Onafuwa-Nuga & Telesnitsky, 2009).

Plus strand synthesis continues terminating on reaching the first modified nucleotide of the tRNA lysine is reached generating the plus strand strong stop DNA shown in figure 2.2 D. Figure 2.2 E

shows RNase H a secondary enzymatic activity of RT, degrades the tRNA primer then initiates a plus strand strong stop strand transfer to the complementary sequences on the minus strand (Onafuwa-Nuga and Telesnitsky, 2009). DNA synthesis is then completed generating the terminally redundant long terminal repeats (LTRs) and a flap of DNA at the central polypurine tract (figure 2.2 F). The linear proviral DNA in association with the viral enzyme Integrase is transported into the nucleus through nuclear pores. HIV-1 integrates into the host genome preferentially into gene rich chromosomes (Chen et al., 2017) and the host repair machinery remove flaps and gaps in the provirus. A persistent infection is established as the provirus is replicated along with the host genome during cell division. It is estimated that approximately 24 hours after infection there is the release of progeny viruses (Moudgil and Daar, 1993).

Components of HIV-1 cytotoxic to cells and the continuous budding out of new virion using the plasma membrane as an envelope lyse infected cells within 2-4 days of infection releasing antigens and proteins. Antigen presenting cells and T-cells are subsequently infected perpetuating the infection. Cytotoxic T lymphocytes attack and lyse virus infected cells through Fas and Fas ligand interactions.

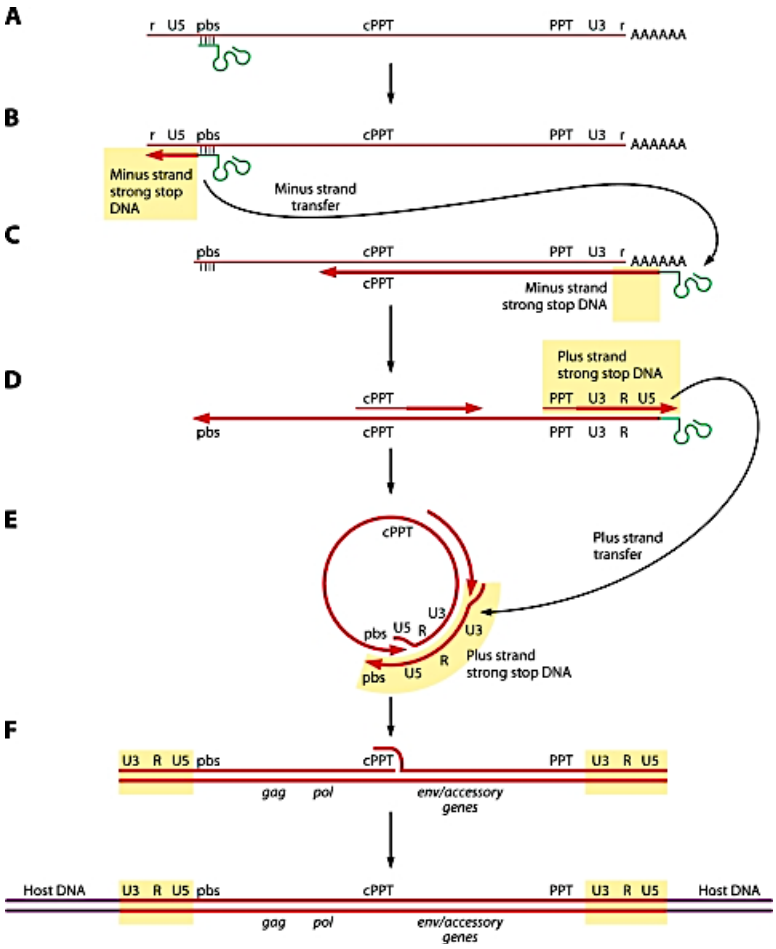

Figure 2.2 Illustration of the process of reverse transcription in HIV-1. Adapted from Onafuwa-Nuga and Telesnitsky (2009). A shows binding of a tRNA-Lysine to the primer binding site of HIV-1 initiating transcription. In B reverse transcription has proceeded until the R region generating the minus strand strong stop. Complementarity between the R regions enables the transfer of minus strand strong stop DNA to the 3' end continuing the process of reverse transcription (C). Synthesis of the minus strand continues with degradation of the RNA template occurring simultaneously (D) leaving the polypurine tracts which primes plus strand synthesis until the plus strand strong stop DNA is generated. Generation of complementary sequences allows the nascent DNA to re-circularize on itself allowing for synthesis of double stranded full-length genome. The flap at the polypurine tracts is removed and the proviral genome integrates into the host genome.

2.3 HIV-1 Transcription and Gene Expression

HIV-1 undergoes two distinct phases of transcription facilitated by interactions between *cis* acting elements within the promoter region of HIV-1 and cellular factors and the successive build-up of sufficient amounts of the viral transactivation factor Tat (Wu, 2004). The LTR is organized into four functional units; the Tat response element TAR which binds Tat, promoter sequences, enhancer sequences and modulatory sequences (Schiralli Lester and Henderson, 2012). The chromatin environment at the site of integration determines the basal transcription activity for the HIV-1 promoter. Host transcription factors such as nuclear factor KB (NFKB), specificity protein 1 (Sp1) and activator protein 1 (AP-1) are needed for basal transcription of HIV-1.

For transcription to be initiated, cellular transcriptional factors are recruited to the LTR forming an initiation complex at the transcription start site as shown in figure 2.3. The 5' LTR promoter is organized into nucleosomes 0, -1 and -2. Downstream of the transcription start site is nuc-1 which is disassembled upon activation of transcription (Verdin et al., 1993). The modification of nucleosomes favours RNA polymerase II (RNAP II) transcriptional elongation (figure 2.3). Transcription proceeds with RNAP II until DRB Sensitivity Inducing Factor (DSIF) and Negative elongation factor (NELF) bring it to a halt just downstream of the transcriptional start site. Transcription of the nascent RNA is further blocked by the host factor transcription factor Pcf1 which terminates transcription prematurely and the further recruitment of Histone deacetylases (HDAC) which enforces a closed chromatin conformation. The binding of Tat to TAR hairpin structure is thought to facilitate the recruitment of Positive transcription elongation factor (P-TEFb) to the LTR through interactions with cyclin T1 (CycT1). Active cyclin dependent kinase 9 (CDK9) is brought into close proximity to the prematurely terminated transcription complex through Tat-P-TEFb interactions (Wei et al, 1998). Recruitment of P-TEFb to the paused elongation complexes leads to phosphorylation of serine 5 of the c-terminal domain of RNAP II

and also phosphorylation of factor DSIF. The phosphorylation of DSIF which leads to the dissociation of NELF from the transcription complex and the further recruitment by Tat of factors Polybromo- and BAF containing complex (PBAF), Switch/Sucrose Non-Fermentable complex (SWI/SNF) and Histone acetyl transferases (HATs) which acetylate the nucleosomes to enable transcriptional elongation to proceed (Schiralli Lester and Henderson, 2012), (Henderson et al, 2004).

Transcription of HIV-1 genes takes place in the nucleus of the cell. The Unique to the 3' region (U3-R) junction in HIV-1 has three Guanosine (G) residues which Kharytonchyk et al (2016) have demonstrated is selectively used in transcription initiation depending on the type of mRNA to be generated. While mRNA packaged into virions preferentially start with one G with a methyl Guanosine cap, transcripts with two and three G residues (with a cap) are enriched on polysomes for translation. As HIV-1 genome is packaged as a dimer, Kharytonchyk et al (2016) showed that the one G form of mRNA allowed dimerization of the two genomes to be packaged into a virion.

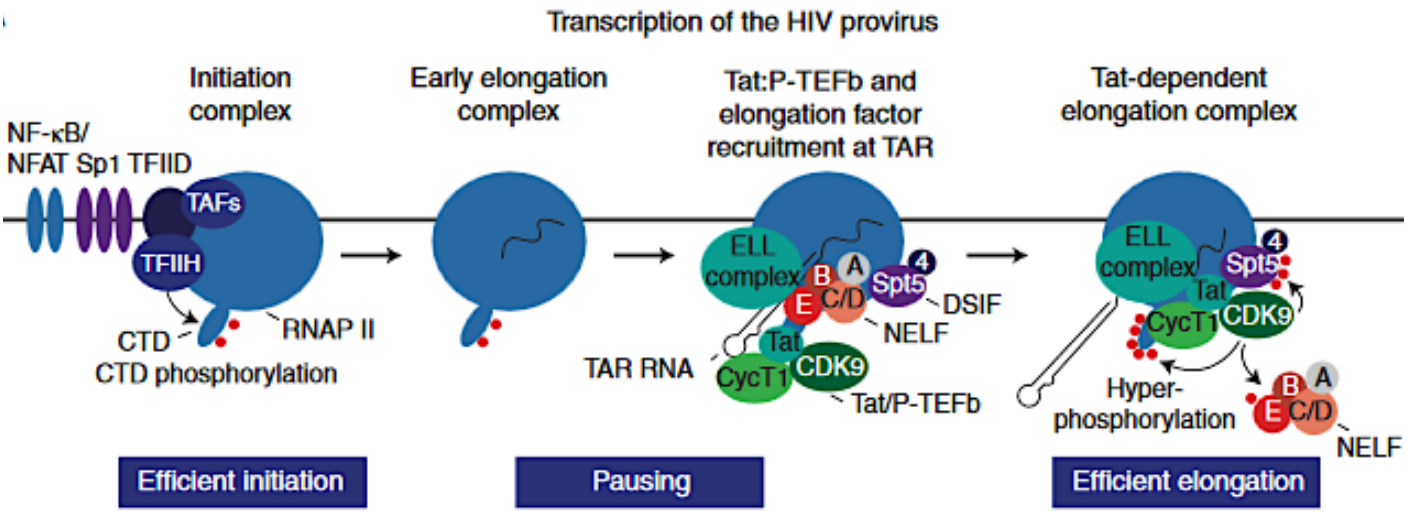

Figure credit: Karn and Stoltzfus (2012)

Figure 2.3 The initiation of HIV-1 proviral transcription (as detailed in section 2.3**).**

2.4 HIV-1 mRNA transport out of nucleus, translation and virion assembly

Different messenger RNA (mRNA) products are generated from the one pre mRNA product generated from the HIV-1 provirus. There are the completely spliced mRNA products, incompletely spliced product and the unspliced mRNA. These are 1.8 kb, 4 kb and 9.3 kb in size respectively. Completely spliced HIV-1 mRNA which encode Rev, Nef, Vpr and Tat are exported to the cytoplasm using endogenous cellular pathways. However, cells degrade incompletely spliced RNA and unspliced cellular RNA in the nucleus. HIV-1 therefore, has to bypass cellular surveillance to enable the export of unspliced and incompletely spliced viral transcripts out of the nucleus. HIV-1 generates incompletely spliced mRNA such as Env, Nef, Vif, Vpu and the single exon form of Tat as well as full length unspliced mRNA which are crucial for virion assembly and cellular function modulation in infection. To enable cellular transport of this class of mRNA, the already exported and translated Rev protein binds to the Rev Response Element (RRE) which is encoded in the incompletely spliced and unspliced mRNA. Conformational changes due to the Rev-RRE interaction along with other stabilising recruited factors enable export of unspliced and incompletely spliced mRNA to the cytoplasm. RNA is exported out of the nucleus into the cytoplasm through nuclear pore complexes of the cell (Karn and Stoltzfus, 2012).

Translation of HIV-1 mRNA products takes place in the cellular cytoplasm. HIV-1 mRNA translation is said to be initiated either through the RNA methyl Guanosine Cap dependent means or through RNA motifs known as the internal ribosomal entry site (IRES) (Chamond et al., 2010). In the cap-initiated translation process, the Guanosine cap recruits the 43S ribosomal subunit as part of the initiation complex which scans the mRNA 5' to 3' until a translation initiation codon is encountered (Ohlmann et al., 2014). For translation to be initiated by the IRES, there is the recruitment of the initiation complex straight to the RNA at the start codon independent of the RNA cap structure (Chamond et al., 2010).

The Gag precursor protein (p55) is expressed from unspliced viral mRNA. The N-terminus of Gag p55 is myristoylated during translation allowing association of the protein with the cell membrane. Gag-Pol precursor protein is also generated from unspliced RNA by a frameshift resulting in the generation of 20-fold less Gag-Pol protein as compared to Gag. Shehu-Xhilaga et al., (2001) have showed that generation of only the Gag-Pol precursor protein can cause an 8-fold reduction in virus protein and up to a 100-fold less packaging of genomic RNA into virions as compared to wildtype virus. They also demonstrated that altering Gag to Gag-Pol ratios resulted in virions with impaired ability to establish new infections (Shehu-Xhilaga et al., 2001). Cleavage of Gag from Pol during virus maturation gives rise to the virally encapsidated enzymes.

HIV-1 Gag protein is known to drive the assembly of virions (Dilley et al., 2017) and thus virus cannot be produced in the absence of Gag. Ganser-Pornillos et al., (2008) hypothesize that some level of Gag-Gag contact is required for viral particle assembly thus ensuring all the needed essentials such as RNA packaging and plasma membrane binding are fulfilled. Gag-RNA (genomic RNA in most cases) interactions is observed to be the definitive signal triggering particle formation (Dilley et al., 2017). Gag anchors into the cell membrane, recruits two copies of HIV-1 genomic RNA which along with many other cellular and viral factors triggers virion budding from the cell. The envelop protein Env is expressed from a singly spliced 4kb mRNA. Env has about 72 trimeric projections from the lipid bilayer surface (Blood, 2016). HIV-1 infectivity is thought to be enhanced by glycosylation of asparagine residues of Env in the cell Golgi. Mature HIV-1 virion is roughly the shape of a football ball (soccer) with a diameter of 145 nm (Briggs et al., 2003).

2.5 HIV-1 cure nemesis – the latent viral reservoir

Establishment of infection by integrating in the genome of the host implies until a cell dies due to infection or replication ceases, HIV-1 theoretically has the ability to produce virus. However, it is observed in patients and laboratory models that HIV-1 goes into a state of quiescence referred to

as the latent infection state where cellular replication occurs, but virus production is absent. This state allows the virus to evade both chemotherapies aimed at HIV-1 and immune surveillance and clearance since HIV-1 antigens are not generated/detected. Spontaneous reactivation from latently infected cells implies ART cannot be interrupted since virus production will be re-initiated. A classic example of treatment interruption leading to viral rebound is the Mississippi baby. This infant did not present with detectable virus for almost 27.6 months when ART use was discontinued sparking hopes of HIV-1 cure. Sequence similarity of 98.6% to maternal HIV-1 sequences detected in plasma confirmed the detection of reactivated viruses and not the introduction of a new infection to the baby (Luzuriaga et al., 2015).

SIV studies in rhesus macaques gave the indication of early establishment of the latent reservoir post infection and most importantly the size difference and ultimately length of time for virological rebound after ART interruption (Whitney et al., 2014). Animals with treatment initiated 3 days post infection had a smaller viral reservoir and thus a longer rebound period as compared to others with much delayed treatment initiation. Treatment was initiated just 30 hours after birth until 18 months of age for the Mississippi baby, yet still the long-term interruption of HIV-1 ART led to virus rebound. Although early initiation of ART leads to a smaller latency reservoir with less virus divergence (Jain et al., 2013, Josefsson et al., 2013), virus rebound observed in the baby shed more light on the very early establishment of the latent reservoir aided by the body's routine establishment of immunological memory. When a CD4+ T-cell encounters its cognate antigen, the cell is activated and expands to generate many daughter cells with the specificity for the cognate antigen. In the course of an infection, most activated cells die off leaving a few some of which go into a state of rest known as long lived memory. Memory cells allow the immune system to mount a quick and robust immune response when the same antigen is encountered again.

Barton and colleagues in their 2017 review define the latent reservoir in people living with HIV-1 as cells harbouring HIV-1 which has the ability to replicate upon ART treatment cessation or interruption (Barton et al., 2017). The identification of greater than 99% of infected cells as phylogenic related members of clonal populations validate cell proliferation as a driving force in maintaining the persistence latent reservoir (Reeves et al., 2018). Hiener et al., (2017) provide evidence for memory cells serving as reservoirs for HIV-1 in infected patients. Using a near full length sequencing technique (FLIP), they were able to locate proviruses integrated in naïve, effector memory, transitional memory and central memory CD4+ T-cells. Of little surprise, effector memory cells; cells tasked with mounting swift effector functions (Sallusto et al., 2004) contained the highest number of near full-length proviruses in all 6 study participants.

Aside memory cells of the adaptive immune system, most cells associated with immunologic response such as dendritic cells and macrophages can act as reservoirs. Effector T-cells, T-helper cells and various hematopoietic precursor cells have also been cited as reservoirs for HIV-1 in patients on ART (Sebastian and Collins, 2014, McNamara and Collins, 2011). Tissue reservoirs for the virus include the genital tract, bone marrow, immunological privileged sites like the brain, central nervous system where astrocytes and various macrophage cell types can serve as reservoirs, the lymphoid tissues including the gut associated lymphoid tissues and lymph nodes (Barton et al., 2017, Fletcher et al., 2014). Low therapeutic levels of ART permissive of viral replication has been reported in lymphatic tissues thus validating the lymphatic tissues as a good location to establish latent reservoirs (Fletcher et al., 2014).

2.6 The fluidity of the viral landscape

HIV-1 cure strategies have been focused on the ability to induce the latent proviral population in the so called ‘Shock and Kill’ approaches (Archin et al., 2012) or orchestrating a permanent block which prevents proviruses in the latent state from ever reactivating in the ‘Block and Lock’

approach (Besnard et al., 2016). These strategies aim to achieve a cure where no provirus will rebound in the absence of ART. However, Ho and colleagues (2013) observed in patient samples intact proviruses which were replication competent but incapable of induction by Phytohemagglutinin (PHA) a known resting CD4+ T-cell activator (Crabtree, 1989). This observation intimates achieving cure devoid of replication competent proviruses may be improbable. Moreover, a longitudinal study assessing the decay rate of the latent reservoir amongst four patients on ART identified that as expected the frequency of intact proviruses declined over time however overall proviral DNA levels remained unchanged (Pinzone et al., 2019). While clonal expansion was observed among all clones, clones with relatively limited expression did not decrease in size over time.

However, there is overwhelming evidence of the accumulation of defective proviruses which form the bulk of the latent pool (Ho et al., 2013, Bruner et al., 2016, Pollack et al 2017). Heiner et al., (2017) estimate only about 5% of proviruses sequenced from 6 patients were genetically intact although this percentage varied from 1 to 10% between individual patients. Defective proviruses were thought to accumulate during the chronic stage of HIV-1 infection without ART treatment. However, Bruner et al., (2016) showed that among patients who undergo ART treatment during acute infection stage, the latent reservoir has significant amounts of hypermutated provirus. Ho et al., (2013) observed from 8 patient samples that about 88.3% of latent proviruses which were non inducible incorporated hypermutation defects introduced by the Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3G (APOBEC3G) which deaminates Cytidines to Uracil during reverse transcription. Hypermutations were detected in various start and stop codons and particularly in the codons for the amino acid tryptophan.

Aside hypermutations introduced by APOBEC3G during reverse transcription, HIV-1 RT is very error-prone with an error rate of about 1 out of every 1700 nucleotides (Roberts et al., 1988). Roberts and colleagues report error rates of 1 per 70 nucleotides in some mutational hotspots. With a relatively small genome of approximately 9.7 kb and a high error prone RT with no proofreading activity, mutations accumulate quickly in the viral genome driving virus evolution. Rapid accumulation of mutations in the viral genome complicates containment of infection by giving rise to many quasi species in a patient. Specific mutations in the sequence of RT are thought to favour certain types of mutations. Abram et al., (2014) using a plasmid which expresses Gag, Pol and HIV-1 wildtype RT demonstrate that known RT mutations which confer resistance to ART slightly increased the errors accumulated during reverse transcription. The mutations were also observed to be localized to hotspots as observed by Roberts and colleagues (1988).

Although some of the mutations lead to replication incompetent proviruses, HIV-1 exploits mutational escape to continually evade the immune system (Geels et al., 2003, Bronke et al., 2013). Pollack et al., (2017) provide evidence that subsets of defective proviruses are targets for cytotoxic T- lymphocytes (CTLs) thus driving the production of CTLs which target non replication competent HIV-1 cells.

Packaging of two genomic RNA gives room for genetic recombination during reverse transcription by switching between templates (Onafuwa-Nuga and Telesnitsky, 2009). Ho et al., (2013) detected large internal deletions from the proviruses they sequenced which were confirmed not to be artifacts of PCR amplification. Template switching between sequences was hypothesized as a possible mechanism. Other defects detected among non-inducible latent proviruses by sequencing included mutations to the D1 5' splice site and packaging signal (psi) mutations, nonsense mutations, frameshift insertions or deletions in at least one gene open reading frame (ORF) (Ho et

al., 2013). In proviruses with significant mutations in only the Psi element or D1, the mutation was sufficient to render the entire provirus defective.

2.7 Determining treatment outcomes- the need for accurate quantification of virus reservoir

Due to the persistent latent reservoir, HIV-1 treatment outcomes have to be monitored over a long period of time. The low frequency of latently infected cells often demands the collection of large amounts of patient blood in culture-based methods in order to detect latently infected cells. PCR detects about 300 proviruses in a million resting CD4+ T-cells while QVOA detects about 1 in 1 million cells (Eriksson et al., 2013). Accurate quantification of the reservoir is needed in order to evaluate the success or failure of therapeutic interventions aimed at the latent reservoir. Bruner et al., (2015) note in their review the many different approaches ongoing clinical trials use in estimating reactivation of the latent reservoir. Among the methods frequently used are the Viral Outgrowth assay which has been modified into the current gold standard quantitative viral outgrowth assay (QVOA), Tat/Rev Induced Limiting Dilution Assay (TILDA) and various PCR based techniques. TILDA and QVOA involve reactivation of latent cells and thus fall short of quantifying non-inducible latent cells.

A literature search reveals fluorescence based quantitative PCR (qPCR) and RT-PCR techniques as the two most commonly used PCR techniques in estimating the latent reservoir. PCR based methods in estimating the latent reservoir are useful because of amplification of detected viral sequences. This implies rare clones can be detected and measured. PCR methods also have the advantage of detecting both inducible and non-inducible proviruses from the latent reservoir. Some fraction of plasma viruses is not infectious, and this may be a reflection of the fraction of defective proviruses. The presence of many defective proviruses in the latent reservoir (Pollack et al., 2017, Ho et al., 2013) however poses a challenge to PCR based methods as PCR does not distinguish between defective and intact proviruses which primers can bind to. Since defective proviruses are

not replication competent, PCR based methods may overestimate the actual latent reservoir by several folds. While qPCR is often used in measuring proviral DNA, RT-PCR is used in measuring released virus RNA from HIV-1 infection. Pollack and colleagues (2017) indicate the detection of supernatant RNA often by RT-PCR more accurately reflects virus release from cells. While PCR based methods may not accurately determine the size of the latent reservoir, it is a quick and relatively easy method adaptable to different laboratories.

The Tat/Rev Induced limiting Dilution Assay (TILDA) is often used to detect HIV-1 RNA in cells usually after latency reversal. Procopio et al., (2015) report on their developed TILDA method as follows; briefly, the process for TILDA involves isolation of PBMC from about 10 ml of patient blood sample. Limiting dilutions of the isolated peripheral blood mononuclear cells (PBMCs) are carried out with the aim of generating a monoclonal cell population. Monoclonal cultures are then stimulated with PMA-Ionomycin and then used in a nested PCR reaction. TILDA does not require nucleic acid extraction from cells and has a turnover time of two days. Multiply spliced RNA (Tat and Rev) is detected from induced culture. However, TILDA also has the disadvantage of detecting defective proviruses which are not replication competent but may generate some viral proteins (Pollack et al., 2017, Ho et al., 2013).

The current gold standard of latent reservoir quantification is the Quantitative Viral Outgrowth Assay (QVOA). This assay involves the purification of resting CD4+ T-cells which forms the bulk of the latent reservoir from about 200 ml of patient blood sample (Archin et al., 2009). Limiting dilution of isolated CD4+ T-cells yields monoclonal populations which are then activated with PHA to activate T-cells to undergo mitosis. Irradiated lymphoblast from three non-infected people is added to the activated latent cells to expand the virus released. After about 2-3 weeks of culture,

enough p24 is detected from virus via ELISA assays (Crooks et al., 2015). Although QVOA is able to detect residual virus in all infected people, it is very laborious and time consuming.

2.8 Zipcoded HIV-1 model

As HIV-1 is primarily a human disease, it has been a challenge to develop study models which recapitulate the disease as in humans. A cell culture model which is easily adaptable to different laboratories and relatively quick and inexpensive and can inform us of possible outcomes of infection is needed in basic science HIV-1 research. A model which couples HIV-1 gene expression with unique genetic markers within the provirus and the virion enables identification of individual proviruses and the virus released thus enabling the study of gene expression and outcomes of infection on a cell basis.

Figure 2.8 shows a schematic of the zipcoded HIV-1 model described henceforth. The zipcoded HIV-1 model has a 20 bp sequence inserted in the LTR. The zipcoded viral libraries are based on HIV-1 NL4-3 and express HIV-1 genes *gag* and *pol*. The *vpr* gene is disrupted to prevent cell cycle arrest in infected cells. The envelope gene *env* is disrupted to prevent reinfection of cells in culture and allow the correlation of individual virus release to an individual provirus/cell infection. GFP is expressed in the *nef* open reading frame and is used as a proxy for HIV-1 gene expression. The GFP reporter also allows for cell sorting using a flow cytometer to differentiate cells expressing HIV-1 from transcriptionally silent cells. Puromycin selection of infected cells compared to a mock culture ensures surviving cells in culture are all productively infected.

The TZ-5 zipcode library is a Jurkat cell line infected with 50 ul of virus media generated from the GPV^- construct at a multiplicity of infection (MOI) of 0.001 and was used in this study. The low MOI ensures each cell is singly infected also allowing for determination of integration sites for each provirus within the genome.

Work done by Read, Atindaana et al, (2019) characterising the zipcoded library indicate distinct cell clones within the library support alternative waves of transcription activation and silencing. Cells heritably maintained the proportion of proviruses which expressed GFP or were transcriptionally silent within each cell clone. Characteristics of the zipcoded library makes it suited to study gene expression dynamics of an HIV-1 infection.

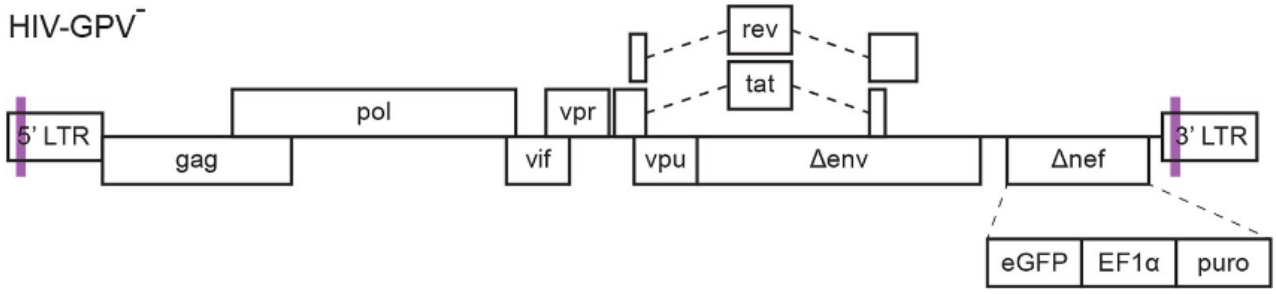

Figure 2.8: The schematic of the proviral construct used in the zipcoded library adapted from Read, Atindaana et al., (2019). The HIV-GPV$^-$ construct has the HIV-1 genes *gag* and *pol* with *env* disrupted to prevent cell to cell infection. The gene *vpr* was also disrupted. The 20 bp individualized and unique zipcode was inserted in the U3 region of the LTR. The construct has a puromycin selection marker which is driven by the EF1α promoter.

CHAPTER THREE

MATERIALS AND METHODS

3.1 MATERIALS

Rosewell Park Memorial Institute (RPMI) 1640 culture medium (Gibco, USA), Foetal Bovine Serum (FBS) (Hyclone, USA), Penicillin Streptomycin Glutamate (PSG) (Gibco, USA), Phosphate Buffered Saline (PBS) (Gibco, USA), Trizol (Invitrogen, USA), Flexi Go taq (Promega, USA), Superscript First-Strand Synthesis System for RT-PCR (Invitrogen, USA), Recombinant RNase (Promega, USA), RQ1 RNase-Free DNase (Promega, USA), NEB Phusion High-Fidelity DNA Polymerase system (USA), Phycoerythrin (KC-57 RD1 Beckman Coulter, USA), Propidium iodide (Thermo Fisher Scientific, USA), Qiagen DNeasy Blood and Tissue kit (USA), Zymo DNA Clean and Concentrator (USA), SyBr Green (Promega, USA).

3.2 METHODS

3.2.1 Study design

This research was part of an ongoing NIH funded project (Grant Number R33 AI116190 from the National Institute of Allergy and Infectious Disease) to study T-cell development biology, human genomics and molecular virology approaches to develop novel high throughput methods for studying classes of latently infected HIV-1 cells and HIV-1 reactivation signatures. This project was a cross sectional study which formed part of the proof-of-concept phase of the broader study.

3.2.2 Cell culture

Jurkat cells infected with 50 ul of pseudotyped zipcoded virus at a multiplicity of infection (MOI) of 0.001 was used in this study. Puromycin selection was used in enriching for infected cells. The pool of infected cells was referred to as the TZ-5 library. Cells were maintained in RPMI 1640 supplemented with 10% heat inactivated Fetal Bovine Serum (FBS), 1x Penicillin Streptomycin glutamate in a 5% CO_2 incubator at 37 °C.

3.2.3 Intracellular Gag staining

Intracellular Gag staining was carried out to ascertain if GFP expression was specific to HIV-1 gene expression. Gag monoclonal antibody conjugated to Phycoerythrin (KC-57 RD1 Beckman Coulter) specific for HIV-1 proteins 55, 39, 33 and 24 kD of the core antigen was used in staining permeabilized cells. A starting total cell concentration of $1x10^5$ of infected cells from the TZ-5 library and uninfected Jurkat cells were used. Cells were washed once with FACS buffer and fixed with 100 ul of BD cytofix for 10 minutes at room temperature in the dark. Cells were then washed twice with FACS buffer and then once with BD perm/wash buffer. Staining for intracellular Gag was carried out at 1:100 and 1:200 dilution of antibody in 1x BD perm buffer. The cells were incubated in the dark at room temperature for 15 minutes and then washed twice to remove excess antibodies. Pelleted cells were resuspended in 200 ul of FACS buffer for acquisition. Acquisition was carried out on the FITC channel for GFP and PE channel for Gag.

3.2.4 Cell sorting

FACS analysis was carried out to sort cells into GFP+ and GFP- populations. Briefly, cells were washed twice with PBS with 10% FBS and stained with propidium iodide to differentiate live and dead cells. Unstained uninfected cells and unstained infected cells were included as controls. The procedure was carried out on ice. Prepared cell samples were sent to the University of Michigan Flow Cytometry core to be analysed. GFP- and GFP+ cells were split 2 ways (as replicate 1 and 2) and each put into a 10 ml culture after sorting alongside an uninfected cell pool.

3.2.5 Reverse Transcription PCR assay

In the HIV-1 virion the enzyme reverse transcriptase (RT) is packaged. RT reverse transcribes HIV-1's positive stranded genome into DNA which gets integrated into the host genome. To assay for virus released in culture, an aliquot of viral media was tested for reverse transcriptase activity by cDNA synthesis of the template MS2 RNA supplied as described by Vermiere et al (2012). Briefly, centrifuged filtered media was collected from uninfected cells, GFP- or GFP+ cells. The

media was lysed at a 1:1 ratio with 2x lysis buffer for 15 minutes. An aliquot of 2 ul of lysed sample was used in the SyBr green based RT-PCR reaction. Exogenous RNA in the form of MS2 RNA was used as a template for viral reverse transcriptase in the assay. cDNA was generated from the MS2 RNA by reverse transcriptase released from lysed virus. SyBr Green dye binds to cDNA amplifying the signal with each cycle of PCR. A standard curve was generated using standardized (viral media with p24 concentration determined by ELISA) viral media from HIV-1 NL4-3 with Kpn I frame shift (KFS) and was used to estimate the relative p24 concentration in each sample since the capsid (p24) to RT ratio is presumed to be the same in every virion.

Virus released into culture supernatant was assayed for by relative p24 concentration by RT- PCR. Virus from filtered culture supernatant was lysed releasing the RT enzyme to convert the substrate MS2 RNA supplied into cDNA. By comparison to a standard curve generated, the relative p24 concentration of each culture was determined since p24 to RT ratio is the same for all viruses. For each of the four samples, two independent experiments were carried out with two replicates per sample per experiment. Each 10ml culture plate had approximately 2 million cells. The culture media was changed 24 hours before qPCR was performed. Jurkat cells divide approximately every 20 hours thus a snapshot of virus release by the cells within a 24-hour period was captured.

3.2.6 Nucleic acid extraction and cDNA synthesis

For each culture of infected cells processed, the culture was centrifuged at 1000 rpm to pellet cells. The supernatant which contained the viral media was decanted off the cell pellet and filtered with a 0.2-micron filter for further processing. Cell pellets were transferred into fresh Eppendorf tubes, centrifuged again and any remaining media pipetted off. Cells were either used in DNA or RNA extraction.

3.2.6.1 DNA extraction

To detect proviral DNA, genomic DNA was extracted from infected Jurkat cells using the Qiagen DNeasy Blood and Tissue kit and manufacturers protocol followed. Extracted DNA was checked for purity and integrity by Qubit analysis and manufacturers protocol followed.

3.2.6.2 RNA extraction from cells

Similar to cellular RNA, HIV-1 RNA was translated in the cell's cytoplasm. Unspliced RNA was packaged in a dimer as genomic RNA into virus. To analyse viral RNA within the cells, the Trizol (Invitrogen) method was used in extracting RNA and manufacturers protocol followed. Extracted RNA was subjected to DNase treatment and DNA free RNA was extracted using Phenol Chloroform Isoamyl Alcohol. Purified RNA was stored at -80 °C until used. RNA was checked for DNA contaminants by setting up a cDNA reaction without the addition of reverse transcriptase and sample was passed when no PCR bands were visualised for no RT controls. Extracted samples were quantified by Nanodrop and quality controlled by Qubit analysis for purity and integrity.

3.2.6.3 RNA extraction from viral media

To extract RNA from the viral media, culture supernatant was gently filtered through a 0.2-micron membrane to ensure no cells were being transferred. Filtered supernatant was centrifuged at 25,000 rpm for 2 hours at 4 °C on a 20% sucrose cushion. The supernatant was discarded after centrifuging. Centrifuge tubes were inverted to drain off any excess liquids. Trizol (Invitrogen) RNA extraction was carried out on virus pellet (not visible) according to manufacturer's protocol. Extracted RNA was subjected to DNase treatment and purified using Phenol Chloroform Isoamyl Alcohol. Purified RNA was stored at -80 °C until used.

3.2.6.4 cDNA synthesis

cDNA was synthesized from extracted RNA using the Superscript first strand synthesis system without RNase H activity to ensure synthesis of full-length cDNA. The procedure was carried out as outlined by the manufacturer's protocol. A gene specific primer (reverse zipcode primer) was

used to prime cDNA synthesis. cDNA was synthesized at 42 °C for 50 minutes and the reaction terminated at 70 °C for 15 minutes. RNase H degradation of RNA was carried out afterwards at 37 °C for 20 minutes. A sample control was set up to detect genomic DNA carried over from the extraction step.

3.2.7 Zipcode PCR

Zipcodes were amplified from cDNA and genomic DNA using the NEB Phusion High-Fidelity polymerase system. The PCR was set up following manufacturers recommendations. PCR reaction was setup on ice and transferred to a heated thermocycler at 98 °C. Denaturation was carried out at 98 °C for 30 seconds, followed by annealing at 59 °C for 10 secs and extension for 20 secs at 72 °C for 25 cycles. Tubes were chilled on ice immediately after PCR for 5 minutes. A 2% agarose gel was used to confirm the presence of the desired PCR product. The PCR product was cleaned out using Zymo DNA Clean and Concentrator kit.

3.2.8 High-throughput sequencing and data analysis

Purified and cleaned PCR products were submitted to the Kidd lab for further processing and analysis. Briefly, samples were prepared for Illumina MiSeq Next generation sequencing which allows for amplification, base calling and alignment after passing quality control tests. Using a custom python programming language-based analysis script (https://github.com/KiddLab/hiv-zipcode-tools), various zipcodes were clustered into zipcode families by their K-means. By collating the zipcodes in descending order according to read numbers, the various zipcode outputs were compared and put into clusters of zipcode families with a cut-off of 5.

3.2.9 Capture Assay

The high affinity specific binding interactions of Biotin and Streptavidin was exploited in the capture assay. A biotinylated probe was designed to bind 30 base pairs complementary sequences in HIV-1 Pol gene which is found in genomic RNA and unspliced RNA but not in spliced RNA since the Pol gene is excluded from spliced transcripts. This probe was used in the capture assay.

Briefly, RNA and probe denaturation were carried out at 65 °C for 15 minutes after which the reaction was put on ice for 3 minutes to anneal probe to RNA. The mixture was immediately added to an equal volume of streptavidin coated magnetic beads and incubated at room temperature for 10 minutes. By use of a magnetic holder, the flow through supernatant (containing spliced and all other RNA) was collected. The beads were washed four times and on the final wash the supernatant was collected and saved as the final wash sample. Unspliced RNA was eluted from the beads by incubating the samples at 95 °C for 5 minutes. The elution step was repeated and the two elutes were combined. The flow through, final wash and eluted RNA samples were precipitated with 0.25 ug/ul of glycogen in ethanol and incubated at -80 °C for 60 minutes. RNA was pelleted by centrifuging at 21,000 g for 80 minutes and the dried pellet dissolved in Tris EDTA buffer (TE).

3.2.10 RNase Protection Assay for Capture Assay

The RNase protection assay exploited the principle that double stranded RNA was protected from degradation from RNases while single stranded RNA was degraded upon digestion with RNases. RNA from the capture assay and total cell RNA were hybridized with the single stranded antisense probe radiolabelled with α-^{32}P Cytidine triphosphate (CTP) (probe referred to as pSKh 65 for capture assay RNA). The full-length probe (128 bp) protected all unspliced RNA through complementary binding with the antisense RNA probe. All HIV-1 RNA which was fated to be spliced was spliced at the major splice donor site D1 5' splice site and thus had 72 bases only protected.

To detect successfully captured unspliced RNA, a mixture of molar excess of probe and RNA was denatured at 95 °C for 5 minutes after which the sample was immediately transferred to a 65 °C water bath to hybridize overnight. On cooling to room temperature, the samples were digested with RNase T1 for 30 minutes. The action of RNase was stopped by the addition of SDS and

proteinase K which degrade proteins and nucleases releasing the RNA into solution. Phenol-Chloroform-Isoamyl Alcohol extraction was then carried out to purify the RNA. The RNA in solution was precipitated in the -80 °C for 30 minutes with Ethanol and yeast tRNA used as a carrier. Centrifugation was carried out at 21,000 g for 80 minutes to pellet RNA. Dried RNA pellets were resuspended in loading dye, denatured at 95 °C for 5 minutes and immediately loaded onto an 8% acrylamide gel. The gel was run for ninety (90) minutes at 500 V after which the gel was transferred to a Whatman filter paper, covered with Saran wrap and dried at 78 °C for 50 minutes. The dried gel was exposed to a Kodak phosphor screen in a cassette overnight. The exposed screen was scanned for radioactive phosphor signal and the resulting data analysed by comparing the band sizes to the RNA ladder loaded on the gel.

3.2.11 Read-through and read-in probes

Two probes (figure 3.2.11) were synthesized to detect read-through and read-in RNA. The antisense sequence of the sequence to be protected was cloned downstream of the T7 promoter in the multiple cloning sites of the pBluescript KS (-) plasmid. A 157 bp Lac Z fragment was downstream of the inserted sequence 18 bp before the Sal I cut site to distinguish full length bound and unbound probe. The primer binding site downstream of the 5′ LTR allowed to distinguish between read-in transcription, read-through transcription and normal HIV-1 transcription.

The read-through probe (figure 3.2.11 A) which was 321 bp long (TL# 2336, AKT_141-1, 3′ LTR probe) was designed to protect 81 bp of U5, 100 bp of R and 140 bp of U3 in the 3′ LTR region of HIV-1. The PBS probe (figure 3.2.11 B) which was 342 bp long (TL# 2337, AKT pg 173, read-in probe) protected 61 bp of the Primer binding site (PBS) at the 5′ LTR, 81 bp of U5, 100 bp of R and 100 bp of U3. Briefly, to generate the designed probes, overlap PCR was used. A first PCR was carried out using primer CX34 and the reverse primer which overlaps with T7 promoter in the

A B

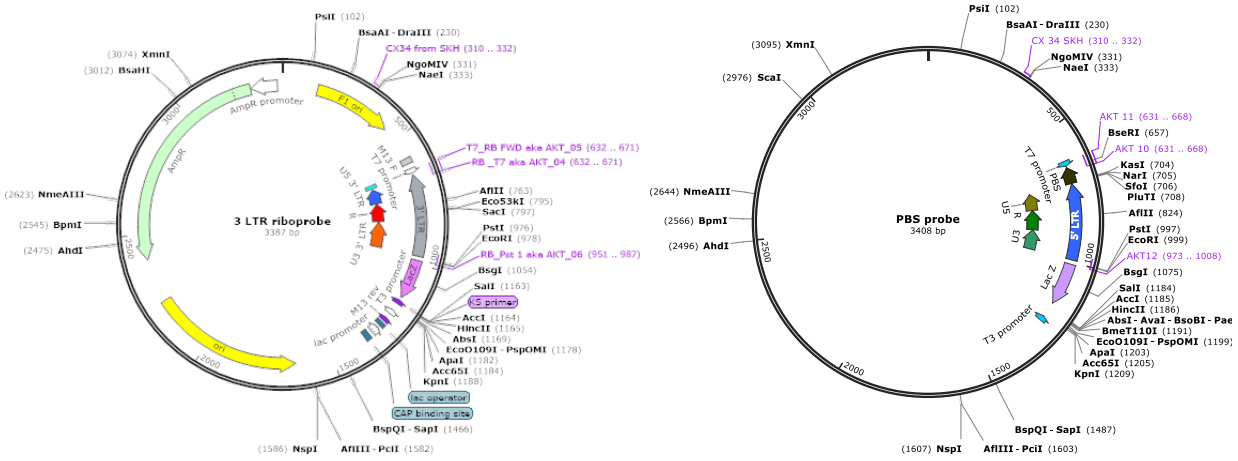

Figure 3.2.11 Plasmid maps showing the read-through probe and the read-in probe. Figure A shows the read-through probe (3' LTR riboprobe) and B the read-in probe (PBS probe). The pBluescript KS (-) plasmid was used as the backbone. HIV-1 LTR sequences were amplified from Telesnitsky bank plasmid 2232 which is a zipcode plasmid.

plasmid backbone. A second PCR was carried out with a forward primer from the T7 promoter region in the backbone and the sequence of interest. An overlap PCR was carried out to join the two fragments together. The PCR ligated products were gel purified and digested with NgoMIV and EcoRI. The plasmid backbone (pBluscript with Lac Z) was similarly digested. An overnight ligation reaction was setup in a 16°C heating block at 4°C overnight. The ligated plasmid was transformed into NEB5-α and plated onto LB plate with ampicillin. The selection plates with bacteria were incubated overnight at 37°C. Four colonies each were picked off each plate and inoculated into a 2 ml LB broth culture. Plasmid minipreps were carried out to isolate the plasmid. Confirmation of the read-through probe was by sequencing (AKT pg 146) and the PBS probe by a double digest with EcoRI and SfoI and sequencing (AKT pg 173B). After digesting both plasmids with the enzyme Sal I to leave 5′ overhangs, T7 polymerase was used in transcribing the antisense DNA into sense RNA incorporating α-^{32}P CTP. The α-^{32}P labelled RNA probes were extracted and resuspended in a 1:5 ratio of hybridization buffer and deionized formamide. The probes were stored at -20 °C for up to 4 weeks.

3.2.12 RNase Protection Assay for chimeric RNA

To determine the populations of RNA present in the extracted cell and viral RNA, a mixture of molar excess of probe and RNA was denatured at 95 °C for 5 minutes after which the sample was immediately transferred to a 65 °C water bath to hybridize overnight. RNA digest was carried out as described in section 3.2.7. The reaction was stopped similarly. The RNA in solution was precipitated in the -80 °C for 30 minutes with Ethanol and yeast tRNA used as a carrier. Centrifugation was carried out at 21,000 g for 30 minutes to pellet RNA. Dried RNA pellets were resuspended in loading dye, denatured at 95 °C for 5 minutes and immediately loaded onto an 8% acrylamide gel. The gel was run for an hour and 30 minutes at 500 V after which the gel was

transferred to a Whatman filter paper then covered with Saran wrap and dried at 78 °C for 50 minutes. The gel was allowed to cool for 20 minutes to prevent the gel from cracking. The dried gel was exposed to a Kodak phosphor screen in a cassette overnight. The exposed screen was scanned for radioactive phosphor signal using the Typhoon FLA 9500. The phosphor fluorescence setting at 200 µM and 750 pixels. The resulting data was analysed by comparing band sizes to the RNA ladder loaded on the gel.

CHAPTER FOUR

RESULTS

4.1 GFP co-expresses with HIV-1 Gag

Cells were selected in media with puromycin for 21 days until uninfected control cells died out of culture. GFP expression was used as a proxy for HIV-1 expression in TZ-5 library. To confirm the expectation of GFP co expressing only with HIV-1 since the LTR promoter drives HIV-1 expression as well as GFP in infected cells, intracellular Gag staining was carried out. Cells were gated into GFP+ and GFP- cells. Intracellular Gag staining showed 46.9% of cells were GFP positive and 53.1% of the cells gated were GFP-. Of the GFP+ gated cells, 44.5% were Gag positive and 1.6% Gag negative. This implied 96.5% of GFP+ cells expressed intracellular Gag thus HIV-1 while 3.5% of GFP+ cells did not. About 1.1% of the GFP- cells were Gag positive representing 2.1% of Gag positive GFP- cell population while 52.7% were Gag negative representing 97.9% of GFP- Gag negative cells. The results are summarised in figure 4.1 and table 4.1.

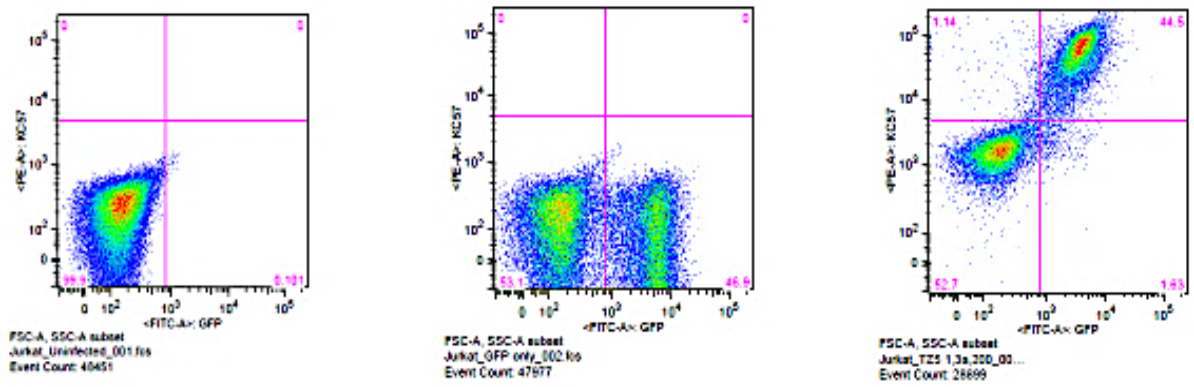

Table 4.1 Summary of intracellular Gag staining

Infected cells	**GFP- (53.1%)**	**GFP+ (46.9 %)**
Gag-	**97.9**	**3.5**
Gag +	**2.1**	**96.5**

Figure 4.1 Intracellular gag staining on zipcoded library TZ-5. The first panel shows gating for live cells in the population. The middle panel shows gating for GFP. GFP- cells are on the left bottom of the quadrat while GFP+ cells are on the right bottom quadrat. The panel to the right shows gating on cells for GFP on the bottom axis and intracellular gag on the upward axis. There was an observed shift on the upward axis for GFP+ cells indicating co-expression of gag and GFP in the pool of infected cells. Table 4.1 summarizes the data.

4.2 Cell sorting

GFP co-expressed with HIV-1 only thus cell sorting based on GFP expression was expected to differentiate cells into distinct GFP+ and GFP- populations. Confluent cell culture of the TZ-5 library was sorted into GFP+ and GFP- cells after staining with propidium iodide to differentiate dead cells from live cells. Previously reported (Read, Atindaana et al, 2019) population distribution of the zipcoded library 70% GFP+ to 30% GFP- was observed from the cell sorting. The GFP+ population was 63% of cell pool (70% of live cells) while the GFP- population was 26% (30% of live cells). Sorted cells were split two ways and put into 10ml culture.

4.3 Relative p24 concentration (ng/ml)

In figure 4.3, experiment 1 is represented by the blue bars while experiment 2 is in orange. Uninfected Jurkat cells were used in the mock qPCR to control for inherent interactions between the molecules released from Jurkat cells into cultured media and the assay. In both experiment 1 and 2 of figure 4.3, the mock culture was two orders of magnitude less in p24 (in ng/ml) than was detected in the GFP- pool which consistently recorded the least p24 values amongst the tested samples. In experiment 1 of figure 4.3, the GFP- cell sort culture recorded a p24 value of $0.28 \pm 4x10^{-2}$ ng/ml while the GFP+ cell sort culture had $20 \pm 4x10^{-1}$ ng/ml of p24. The total pool had a p24 value of $15.36 \pm 1x10^{0}$ ng/ml. In experiment 2 of figure 4.3, GFP- cell sort recorded a p24 value of $0.67 \pm 3x10^{-3}$ while GFP+ cell sort recorded a value of $32 \pm 3x10^{-1}$ with the total library recording $15.46 \pm 5x10^{-1}$ ng/ml of p24.

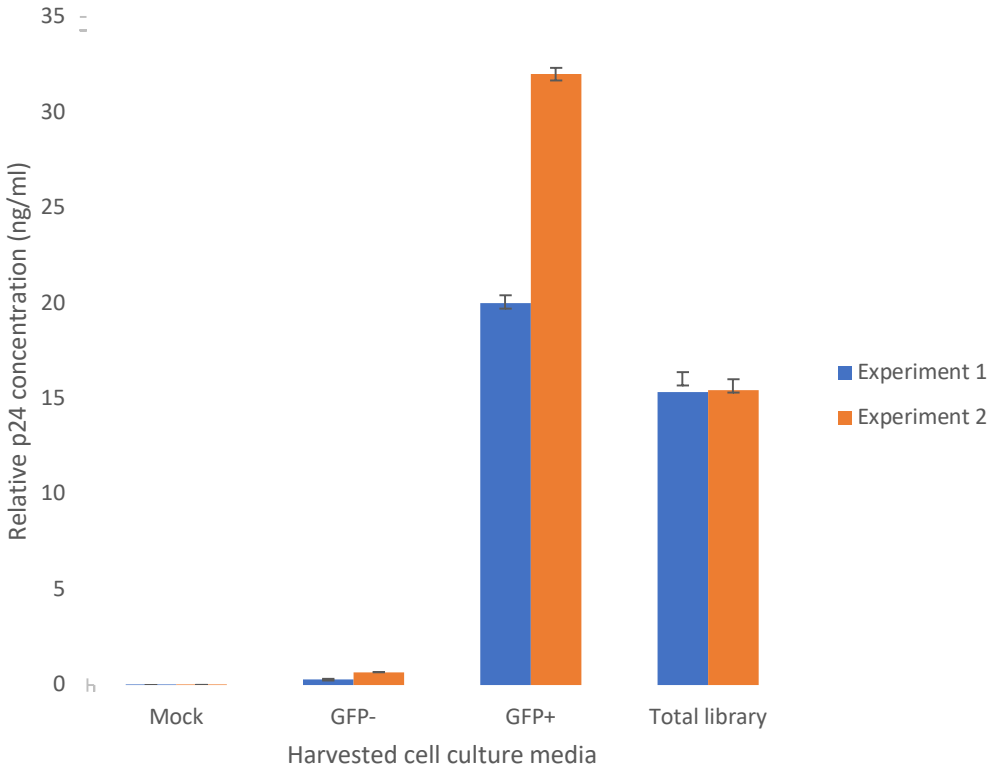

Figure 4.3 Results of RT-PCR to quantify the relative p24 (ng/ml) concentrations. The relative p24 (ng/ml) concentrations released into culture media in a 24-hour time period from 2 million cells in a 10 ml culture was determined by RT-PCR. Blue bars represent experiment 1 while orange bars represent experiment 2. GFP+ cell sort culture media recorded the highest concentration of p24 which was about five orders of magnitude greater than the mock experiment and two orders of magnitude greater than the GFP- cell sort culture. The total pool recorded about a quarter less p24 compared to GFP+ culture in experiment 1 and half less p24 in experiment 2.

4.4 Capture assay and RPA

The high affinity interactions between streptavidin and biotin were exploited to selectively enrich for subpopulations of RNA using a biotinylated probe and streptavidin coated beads. About 2 ug of RNA was captured from 10 ug of input total cell RNA using streptavidin coated beads and a biotinylated probe which binds RNA within the RT region of HIV-1 Pol gene. To determine the success of the capture assays carried out, 600 ng of RNA was hybridized to the probe pSKh 65 which was complementary to a region of HIV-1 spanning SD1 and flanking regions and labelled with α-^{32}P CTP.

The probe was expected to protect a band size of 128 bp for unspliced RNA and 72bp for spliced RNA. In the gel lanes for total cell RNA in figure 4.4, a band was visualized for both spliced and unspliced RNA as expected. The flow through sample showed an enrichment for spliced RNA since part of the complementary sequence for the biotinylated probe is absent in spliced RNA. Several washes were carried out to ensure almost no spliced RNA was detected in the final wash sample. The eluted RNA gel lane showed a selective enrichment for unspliced RNA. The probe did not interact with tRNA which was used as a carrier in sample preparations. Captured RNA using the Pol probe was used in subsequent work since it produced a much cleaner sample overall. The result is summarised in figure 4.4.

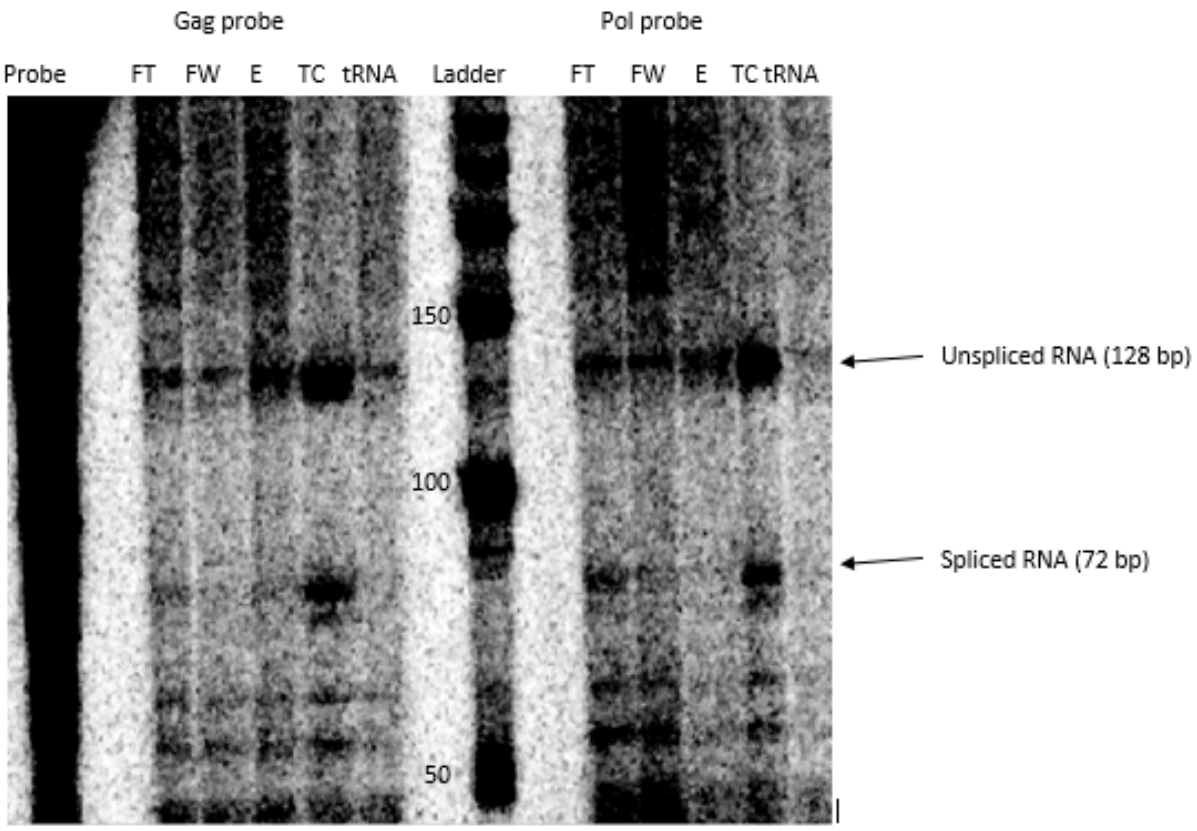

Figure 4.4 Capture assay RPA results. 8% acrylamide gel was used for RPA. Probe was pSKh 65 which binds to 128 bp of unspliced RNA and 72 bases of spliced RNA. FT is the flow-through sample, FW is the final wash sample and E is the elution sample. Two sets of captured samples were run on the gel. The first lane shows the probe in molar excess of the samples run. Overall, the elution sample from the Pol capture showed selective enrichment for unspliced RNA as compared to spliced and was used in cDNA synthesis and eventually HTS. The gel was exposed for longer (hence grainy picture) due to low quantity of HIV RNA present in samples.

4.5 High throughput sequencing (HTS)

To determine the abundance of various zipcodes in each cDNA and gDNA library generated, high-throughput sequencing was carried out. After cDNA synthesis from RNA, PCR was carried out to amplify zipcodes from genomic DNA, viral cDNA and cDNA from unspliced viral RNA and total cellular RNA with NEB Phusion polymerase. Phusion polymerase has an error rate of $4.4x10^{-7}$ (according to the manufacturer) making it a suitable choice for the zipcode PCR as few to no PCR errors were detected in the zipcode PCR product as evidenced with distinct families of zipcodes with exactly 20 bp zipcodes. About a 100 bp product was visualized on a 2% agarose gel.

Eight samples were submitted for sequencing in each replicate. These were genomic DNA from the total pool, GFP+ and GFP- cells. cDNA was generated from total unsorted viral media, GFP+ and GFP- pool as well as total cellular RNA and captured unspliced RNA. High through-put sequencing generated 132 zipcode families which were clustered in decreasing order of fraction of abundance according to the unsorted pool genomic DNA. The fraction of abundance for an individual pool (say GFP+ genomic DNA) was equal to 1. The top 74 families (table 4.5) which represents ≥ 0.974 of total reads (fraction of abundance) for each sample and were found in both replicate 1 and 2 were further characterized in this study.

Table 4.5 Distribution of the raw sample abundances from cells after high throughput sequencing

originalRank	zipCode	Unsorted poo	GFP+ gDNA 1	GFP- gDNA 1	Unsorted poo	GFP+ gDNA 2	GFP- gDNA 2
1	AACCATTGTGTTGCCTATCT	0.1607	0.1831	0.0142	0.1553	0.1778	0.0106
2	GACGAGGATTGTTTCGCAAT	0.1112	0.1404	0.1072	0.1205	0.1494	0.0801
3	TGGCTGAGTCGCTGGGTAAA	0.0931	0.1847	0.0315	0.0903	0.1771	0.0223
4	TACGGAGATTTTACACAACT	0.0819	0.1679	0.0618	0.0850	0.1673	0.0599
5	ATTGGATCTGAGTTACCTAG	0.0701	0.0702	0.0184	0.0637	0.0695	0.0136
6	ACATACGACGTGATTGGATG	0.0643	0.0009	0.1707	0.0679	0.0011	0.1984
7	GTTTGTCGATCGAAACCGAC	0.0450	0.0008	0.1823	0.0438	0.0009	0.1912
8	GCACGTCTAACTTGTGGACC	0.0376	0.0098	0.0600	0.0387	0.0106	0.0754
9	AGCGGTGCTGGGACTTTAAA	0.0343	0.0420	0.0187	0.0342	0.0405	0.0132
10	GACCGACCACGTTATTATCC	0.0308	0.0110	0.0449	0.0249	0.0097	0.0578
11	GCTGCCTAAGATAAAGACCG	0.0273	0.0321	0.0047	0.0252	0.0340	0.0022
12	CGAGCCAGACGAGTGACCAG	0.0268	0.0003	0.0520	0.0283	0.0001	0.0618
13	TTTCTCCTTGAGTAAGGTAC	0.0253	0.0006	0.0554	0.0240	0.0009	0.0565
14	GTAAACCTTAATTTATTAGT	0.0241	0.0254	0.0050	0.0246	0.0243	0.0013
15	TATACAGTAATAGGGTTCCA	0.0220	0.0299	0.0030	0.0224	0.0344	0.0013
16	GGCCGGATAAATCATTCTGG	0.0195	0.0084	0.0130	0.0200	0.0093	0.0127
17	CAAAAATGAGGAACACTGAA	0.0185	0.0003	0.0247	0.0164	0.0004	0.0268
18	AGCACCATTTTATGGACCTT	0.0147	0.0138	0.0077	0.0132	0.0119	0.0089
19	GTCAGCTCGTGATATGGCTG	0.0101	0.0064	0.0056	0.0097	0.0052	0.0066
20	GACGTAAGTAACGATTTGAC	0.0090	0.0006	0.0082	0.0078	0.0001	0.0059
21	TCTTATCATCCCATGGCCCG	0.0078	0.0102	0.0049	0.0083	0.0102	0.0031
22	CAATCATTAAGGGCTCAACT	0.0058	0.0002	0.0218	0.0061	0.0001	0.0223
23	TGGGTGGTGTGTGTTCGTTG	0.0058	0.0088	0.0005	0.0057	0.0092	0.0004
24	CTTGACCGACTAAGCCGGAT	0.0045	0.0024	0.0058	0.0046	0.0021	0.0055
25	TCATATTGGCTGAATGGACC	0.0037	0.0060	0.0013	0.0042	0.0053	0.0009
26	GCGGAGCCGGAATTTTCTAT	0.0030	0.0010	0.0079	0.0034	0.0010	0.0102
27	ATTACGTGACATCAACTAAC	0.0018	0.0028	0.0002	0.0017	0.0030	0.0002
28	CTGATAGTGGGTTGGTTGAT	0.0014	0.0016	0.0004	0.0011	0.0016	0.0002
29	GTTTGCAAAAACTGGTTACA	0.0014	0.0012	0.0028	0.0019	0.0013	0.0021
30	GATGGTACCCCACACGTTCT	0.0012	0.0025	0.0001	0.0010	0.0022	0.0002
31	CGTTAGGCAAAAACGGCAGA	0.0012	0.0001	0.0023	0.0014	0.0001	0.0029
32	GCCGCTGCTTCCAGTCGCGT	0.0011	0.0012	0.0018	0.0011	0.0015	0.0013
33	GTTGCCCCTTGCCATTAGGT	0.0010	0.0000	0.0020	0.0009	0.0000	0.0026
34	TGCAACAAGTCTTGACCTTA	0.0009	0.0000	0.0019	0.0008	0.0000	0.0018
35	AAAGTTTTTAACCACGCGGC	0.0009	0.0006	0.0023	0.0014	0.0011	0.0012
36	CCGCTACGGGCATGCATCCG	0.0008	0.0007	0.0001	0.0007	0.0008	0.0001
37	ATTAGGATTACGTATTGCGA	0.0008	0.0000	0.0010	0.0005	0.0000	0.0011
38	ACGTAAGTCGTCACCGGGAG	0.0008	0.0002	0.0002	0.0008	0.0002	0.0003
39	CAGTGTTAACTGCACTTCTG	0.0008	0.0018	0.0001	0.0009	0.0019	0.0002
40	TTAGCAGATGGCCGCAGGAA	0.0008	0.0000	0.0019	0.0009	0.0000	0.0021
41	TGAATACCAGCCGTCGTCTT	0.0007	0.0007	0.0009	0.0007	0.0008	0.0006
42	ACTCATCATGGTTTCAGATG	0.0006	0.0014	0.0002	0.0006	0.0013	0.0001
43	TCTGTTTTAATGGTCGAATA	0.0005	0.0006	0.0003	0.0004	0.0005	0.0002
44	TAGTTAGTGGGTCGTTAAAC	0.0005	0.0012	0.0002	0.0007	0.0011	0.0001
45	CGCTAGGGGTATTGCTACAC	0.0005	0.0002	0.0001	0.0003	0.0002	0.0001
46	GGACCTTCCTGGCCCGATCC	0.0005	0.0010	0.0002	0.0006	0.0012	0.0001
47	TATAGCGCAGCGGAAAAAAG	0.0004	0.0004	0.0003	0.0004	0.0004	0.0002
48	AGCCGTGGATCGAGTGGGAG	0.0004	0.0003	0.0008	0.0005	0.0006	0.0009
49	CTGACTGCTGGCTAGGAGGT	0.0003	0.0000	0.0010	0.0003	0.0000	0.0009
50	GGAATGTCAAGTGCGACCGC	0.0003	0.0005	0.0000	0.0003	0.0005	0.0001
51	AAACGGGGAGCGAGGACCCC	0.0003	0.0003	0.0001	0.0002	0.0002	0.0001
52	ATCCATATGGCGCATAACAT	0.0003	0.0002	0.0008	0.0004	0.0004	0.0005
53	GTCGCATGGTTGTGCACCGC	0.0003	0.0000	0.0008	0.0003	0.0000	0.0008
54	AAGCCTCCACTACGTGCGCT	0.0003	0.0006	0.0001	0.0006	0.0006	0.0000
55	TTTCCTAATGCGTGTATGGC	0.0003	0.0000	0.0010	0.0001	0.0000	0.0011
56	TTCACGCGGGATTAACCTCT	0.0003	0.0003	0.0001	0.0002	0.0002	0.0001
57	GGGCTAAGTTTGGTCAACAA	0.0003	0.0007	0.0002	0.0003	0.0006	0.0000
58	CGTTGTAGTAGGCAGCACTG	0.0003	0.0004	0.0001	0.0002	0.0003	0.0001
59	CAACTTCCCCGCTTCTTTAG	0.0003	0.0003	0.0000	0.0003	0.0003	0.0000
60	TGCATTGTCTCCTGACGACC	0.0002	0.0002	0.0005	0.0004	0.0003	0.0004
61	TCACGCGATACCGTCGCTCG	0.0002	0.0002	0.0000	0.0002	0.0002	0.0000
62	TGGATCTTTGATTCTCTTAA	0.0002	0.0003	0.0000	0.0001	0.0005	0.0001
63	ATCCGGCCGCGAACGGGGGG	0.0002	0.0003	0.0000	0.0003	0.0003	0.0000
64	TGAGCGGCGTAAGCGCTACC	0.0002	0.0000	0.0005	0.0003	0.0001	0.0004
65	AACTCCTGCTCTCCGGTCGG	0.0002	0.0004	0.0002	0.0004	0.0003	0.0001
66	CGACGACCCAGGTTTAAGTT	0.0002	0.0000	0.0002	0.0003	0.0000	0.0005
67	TCCGACAGTCAGTCCAACGG	0.0002	0.0004	0.0000	0.0002	0.0005	0.0000
68	AGACCTGTGCTCTATAAGCT	0.0002	0.0001	0.0001	0.0001	0.0002	0.0001
69	TTCGCTGCAATTCAATCGAC	0.0002	0.0003	0.0001	0.0003	0.0002	0.0000
70	TTGGCTATAATCCCGGAGCC	0.0001	0.0002	0.0000	0.0001	0.0002	0.0000
71	TCCCGAGGCCCACTTCACAC	0.0001	0.0000	0.0006	0.0002	0.0000	0.0006
72	TCATTTTAATGGGTGACCCG	0.0002	0.0001	0.0004	0.0002	0.0002	0.0002
73	CTAGTACCGGAGACGCATAT	0.0001	0.0000	0.0003	0.0001	0.0001	0.0002
74	AGATGCGTCACCTTTTAGTC	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001
	Total number of reads (out of 1)	**0.9823**	**0.9820**	**0.9588**	**0.9740**	**0.9782**	**0.9734**

The topmost 74 clones were used in the analysis for this study. Zipcodes were ranked according to read abundance in the unsorted total cell pool. The 74 clones represented ≥ 0.97 (97%) of reads recorded in each library. Individual fraction of abundances ranged from 0.184 to 0.000.

4.6 Clones have distinct GFP expression patterns

GFP expression was driven by the HIV-1 LTR thus the vector was designed so expression of GFP correlated with HIV-1 gene expression. The question however remained; did all GFP+ cells express HIV-1 to the same extent? To determine the extent of GFP expression by a clone, that is if a clone is predominantly GFP+ or GFP- within the library, the green proportion was calculated. Green proportion (GP) is a measure of the weighted reads which were captured for a particular clone within the GFP+ pool as a ratio of the weighted number of reads for the same clone within the sum of the GFP+ and GFP- pool as a fraction of 1. The ratio of GFP+ to GFP- determined during cell sorting was used in weighting reads.

That is $GP=\left[\frac{0.7*\text{number of reads for clone A in GFP+ pool}}{(0.3*\text{number of reads for clone A in GFP}-\text{ pool})+(\ 0.7*\text{number of reads for clone A in GFP+ pool})}\right]*1$

The GP for the clones was categorized into 10 groups (<0.10, 0.10-0.19, 0.20-0.29, 0.30-0.39, 0.40-0.49, 0.50-59, 0.60-0.69, 0.70-0.79, 0.80-0.89, >0.89). About 50% of the clones (37 clones in experiment 1 and 36 clones in experiment 2 = the sum of clones from the 2 topmost groups) were predominantly GFP+ indicating a transcriptionally active state while about 25% of the clones (18 clones = the sum of clones from the 2 bottommost groups) were predominantly GFP- indicating transcriptional silent state (figure 4.6 A). The remaining 25% of the clones were 70% to 30% transcriptionally active or silent. From figure 4.6 B, it was inferred there were more clones with lower green fractions in replicate 1 than in replicate 2. There was a correlation of 0.76 between the two replicates indicating that the green fraction of each clone is mostly unchanged between replicates except for some outliers.

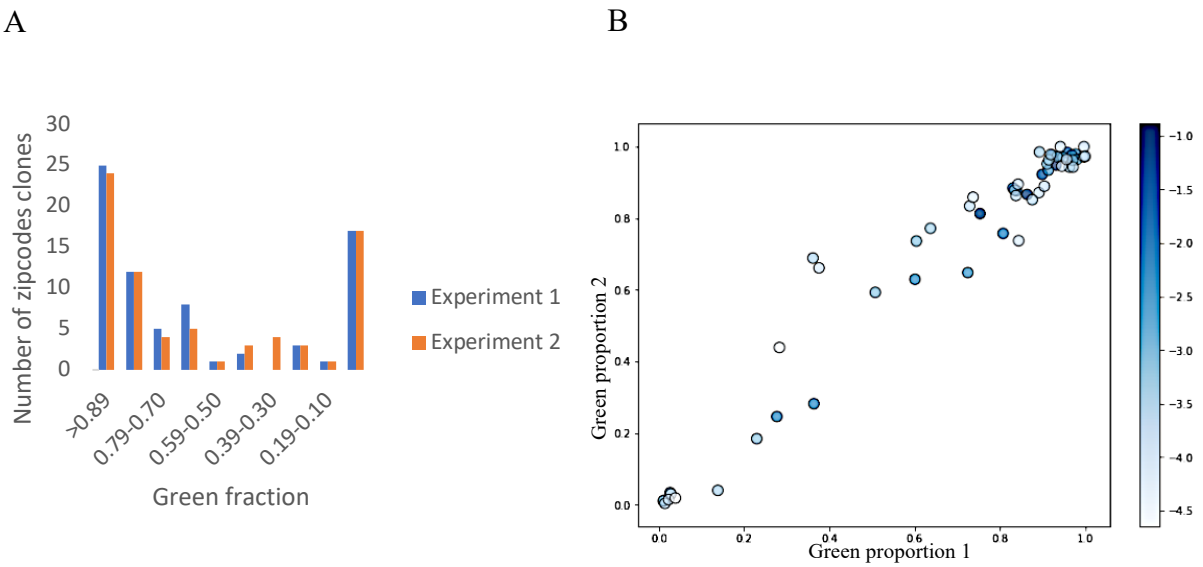

Figure 4.6 Clonal differences in GFP expression. Clones expressed GFP to different degrees and this trait was a stable clone phenotype in both replicates. The bar chart (A) shows the distribution of the 74 clones. About 50% of the clones expressed high amounts of GFP indicating high HIV-1 transcription while about 25% were transcriptionally silent clones in both replicates. The dot plot (B) shows the correlation of the green proportion ($\log_{10}$) for each clone in replicate 1 and 2 ($R^2 = 0.76$). The intensity of the colour bar shows increasing abundance of clones (all blue bars represent abundance) within the total GFP+ cell pool.

4.7 Different clones release different amounts of viruses per cell

Based on the green proportions calculated, different clones express GFP to varying amounts. The question to be addressed was if clones with similar abundance in the pool contributed equally to the virus pool since these clones varied in the extent to which GFP was expressed and thus HIV-1. To determine the contribution of each clone to the p24 value, the fraction of abundance of each clone in the viral pool determined by HTS was used. A plot of virus released per cell against clone abundance in cells (genomic DNA, figure 4.7 A and B) showed there was no correlation between clonal abundance and virus released. The most abundant clones did not release the most virus per cell. It was observed from figure 4.7 A and B that clones varied widely in the amount of virus released per cell. Plot C of figure 4.7 shows the correlation of virus release per cell in replicate 1 and 2. There was a correlation of 0.79 between the two replicates indicating the amount of virus released per cell may be around a fixed ratio for some clones.

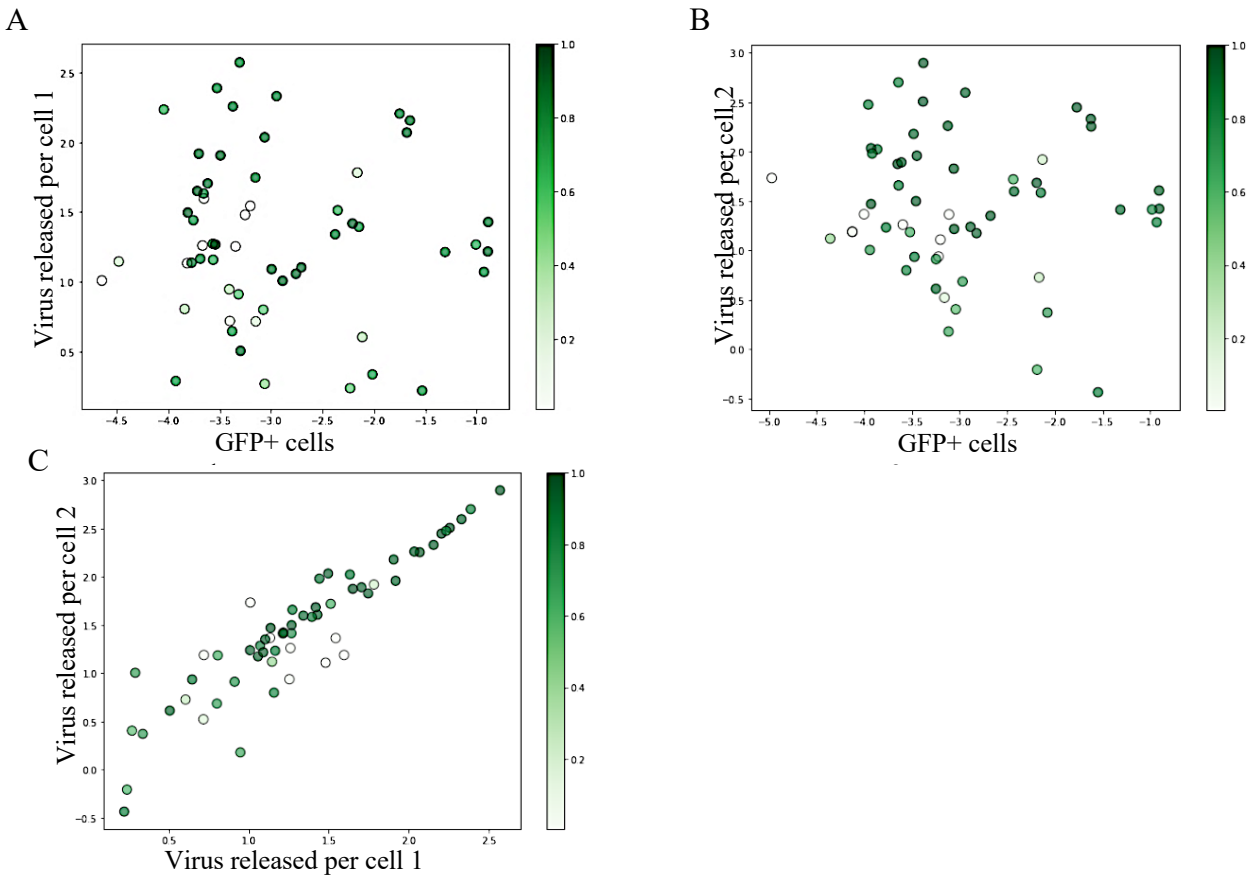

Figure 4.7 Variation in virus release per cell. A and B show a plot of virus released per cell (log_{10} scale) against clone abundance (log_{10} scale) in genomic DNA for experiment 1 and 2 repectively. The intensity of the colour bar shows increasing green proportions (all green bars represent green proportions) within the total GFP+ cell pool. Virus release spaned five orders of magnitude. No distinct patterns were observed for abundant clones or rare clones. There was a correlation of 0.79 for virus release per cell between the two replicates (C).

4.8 Clones with higher green proportions likely release higher virus per cell

The most abundant clones in the GFP+ cell gDNA pool did not release the most virus per cell indicating that a faster proliferation rate may not correlate with virus release. Since different clones expressed GFP to different extents (indicated by the differences in the green proportions calculated), a plot of virus released per GFP+ cell against the green proportions was generated to determine if clones which expressed GFP most of the time (higher green proportion) were most likely to release virus. The dot plot showed that among high green proportion clones (GP $\geq$ 0.89), there was a higher likelihood to release more virus on a per cell basis as clones shifted to the right (higher green proportions) tended to release at least one order of magnitude higher virus ($\geq$ 10 ng/ml of p24 per cell) in figure 4.8. Clones with medium to low GP (0.7 and below) varied in the amounts of virus released. Medium to low GP clones mostly released virus below 10ng/ml. The observed pattern was independent of clonal abundance (intensity of blue dot).

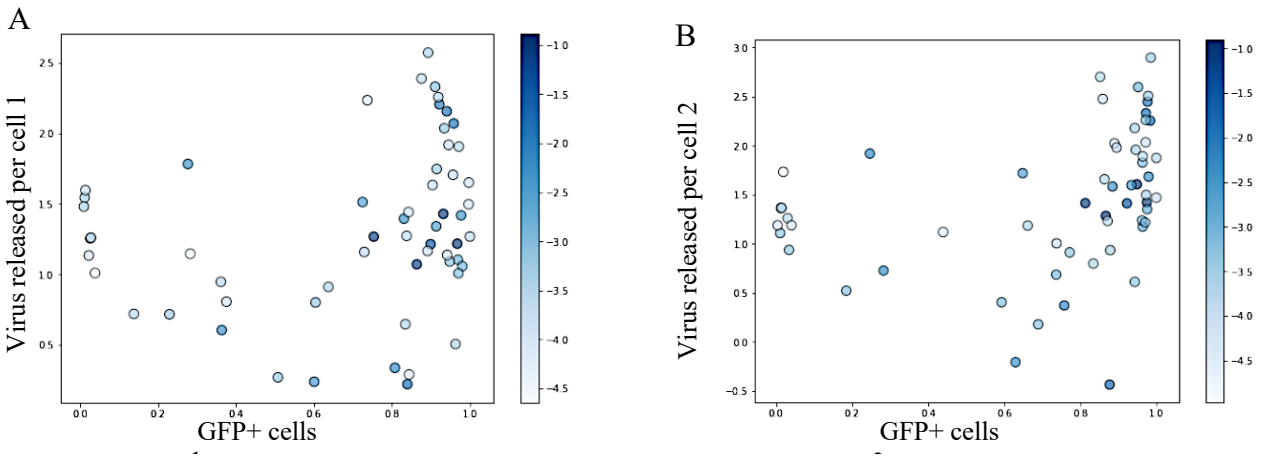

Figure 4.8 High green proportions may predict high virus release. A and B show plots of virus released per cell against green proportions (GP) all on $\log_{10}$ scale. The intensity of the colour bar indicates increasing abundance of clones (all blue bars represent abundance) within the total GFP+ cell pool. Clones with high green proportions were also observed to have high virus release although not for all clones. Among medium to low GP clones' virus release varied but was generally below 10 ng/ml of p24 per cell.

4.9 Different clones may have similar ratios of unspliced RNA to total viral cell RNA

Virus release appears to be independent of clonal abundance but high GFP expression may predict high virus release by the cell. It was necessary then to determinine if differences in virus released was as a result of differences in the amounts of unspliced RNA to total cell RNA generated in cells. This was fundamental as unspliced RNA is packaged into virions and also serves as the mRNA transcript used in generating Gag and Gag-Pol both of which are needed for virion assembly. If clones which released high amounts of virus per cell produced high amounts of unspliced RNA relative to the total cell RNA while clones with low virus released per cell generated low amounts of unspliced RNA relative to the total cellular viral transcripts, then it was expected that a plot of the ratio of unspliced RNA to total cell RNA against total cell RNA would generate a perfect linear correlation. However if cells generated about the same amounts of unspliced RNA relative to the total viral cell RNA, then we would predict a horizontal line showing no correlation. It was observed that a plot of the ratio of unspliced RNA to total cell RNA against total cell RNA showed clones may generate similar amounts of unspliced RNA relative to total cell RNA except for outliers as shown in figure 4.9 A and B. Clones which generated low amounts of unspliced RNA relative to total cell RNA showed much variation moreso in figure 4.9 B. No patterns were observed for high virus relase (intensity of red dot) in Figure 4.9. Between the two replicates there was no correlation.

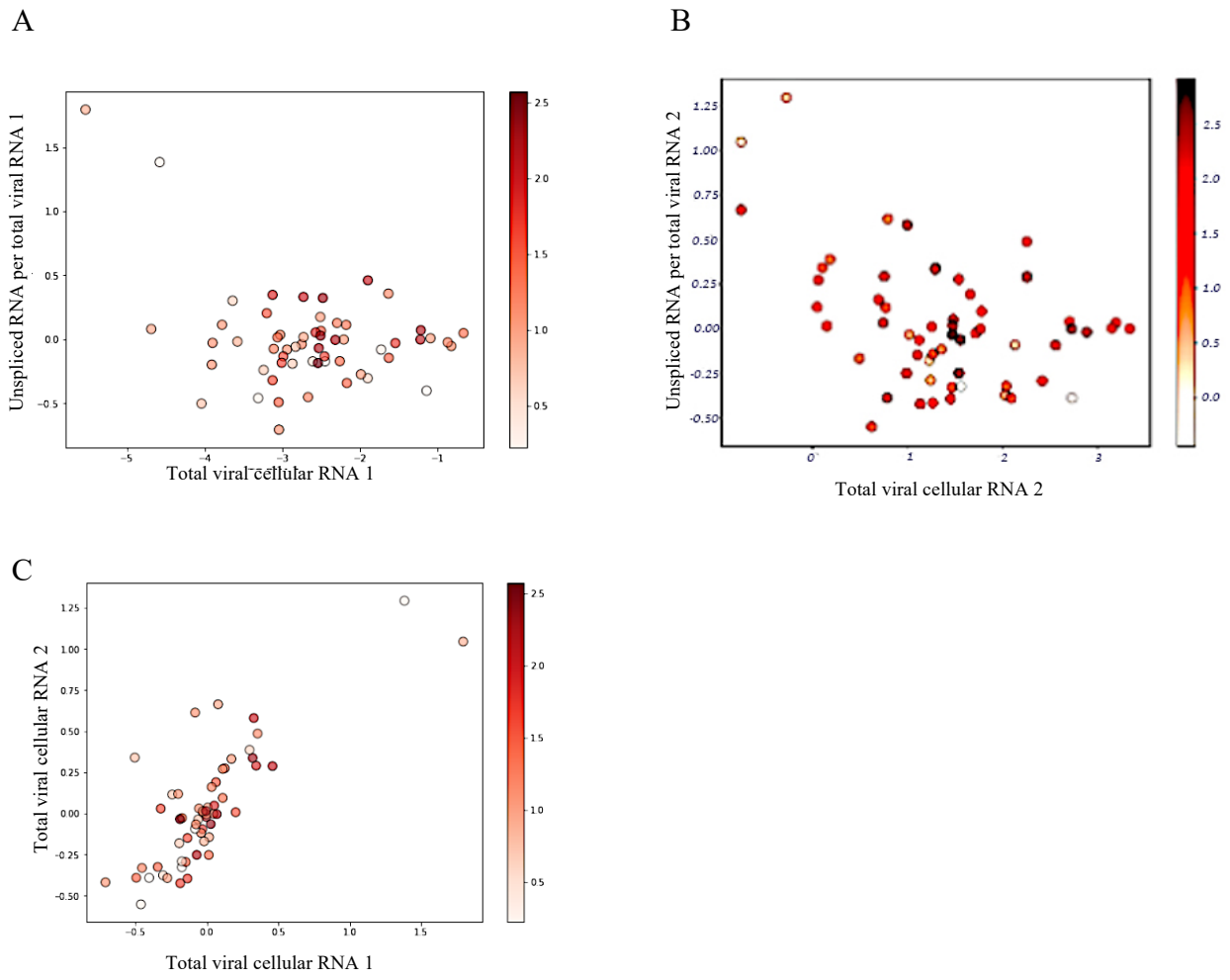

Figure 4.9 Clonal differences in unspliced RNA relative to total viral cell RNA. A plot of ratio of unspliced RNA to total viral cellular RNA against total viral cellular RNA all on $\log_{10}$ scale. The intensity of the colour bar showed increasing virus release per cell (all red bars represent virus release per cell) within the total GFP+ cell pool.

4.10 Cells with higher unspliced RNA are more likely to release virus

If different clones generate similar amounts of unspliced RNA compared to the total viral cellular RNA, then on a per cell basis does amounts of unspliced RNA differ among clones. The prediction was that cells with a higher amount of unspliced RNA per cell may release more virus. This also suggested cells with an oversplicing phenotype would release less virus as the primary transcript would be used in generating more spliced mRNA variants. In figure 4.10 A and B plots of unspliced RNA per GFP+ cell against virus release per GFP+ cell showed clones which released more virus per cell most likely had a higher unspliced RNA per cell than clones which had less virus released. This observation was independent of clonal abundance as the most abundant clones (more intense blue colour) were not the topmost clones.

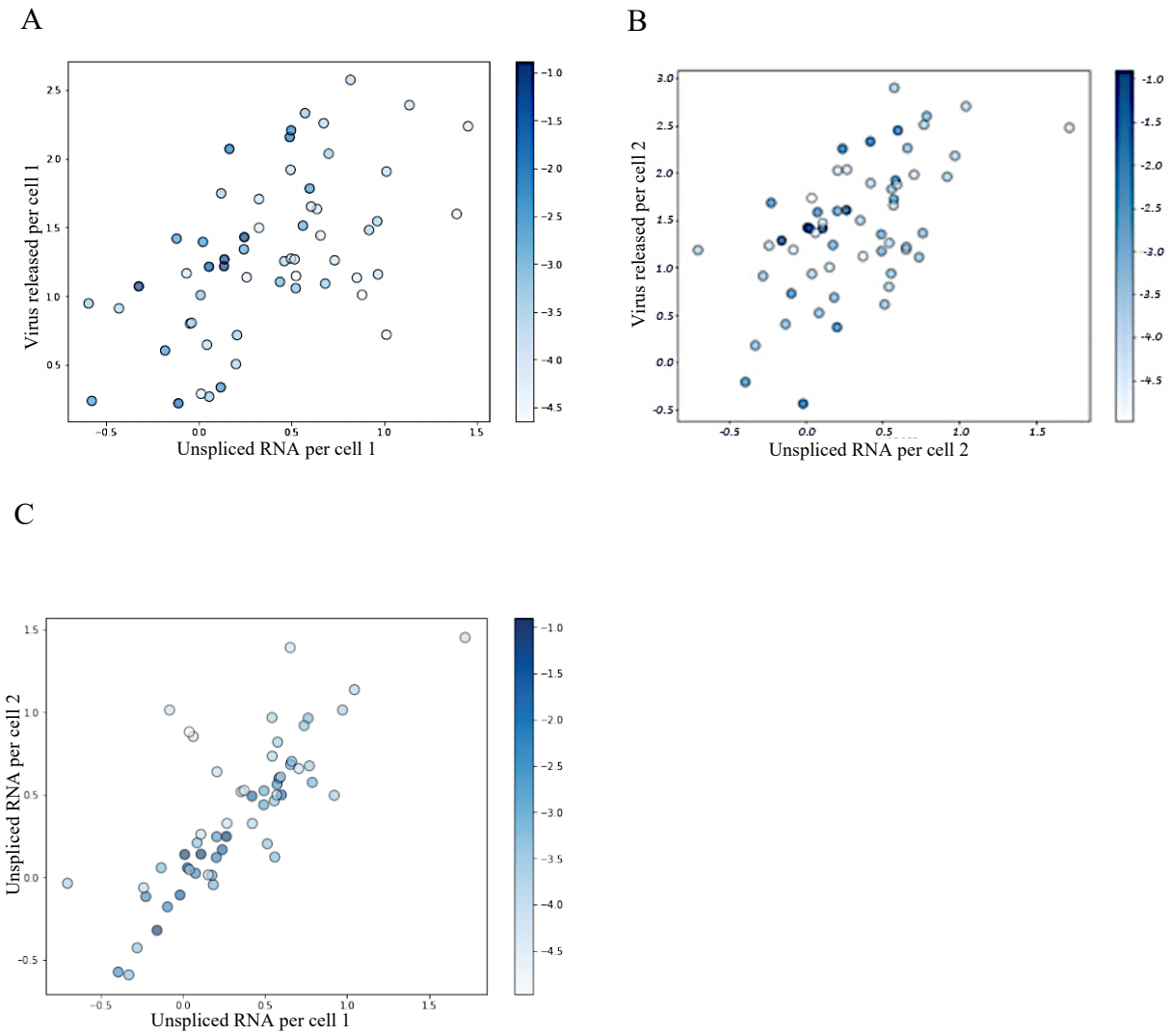

Figure 4.10 Cell differences in amount of unspliced RNA generated against virus release

Plot A and B show virus released per cell against unspliced RNA per cell on $\log_{10}$ scale. Cells which released more virus per cell likely had a greater amount of unspliced RNA per cell in both experiments (plot A-replicate 1, plot B- replicate 2) although not perfectly. Plot C has a R^2 of 0.59. The intensity of the colour bar shows increasing cell abundance within the total GFP+ cell pool.

4.11 Higher amounts of Unspliced RNA may indicate higher green proportion and virus release per cell

Using magnetic beads coated with Thymidine tails (oligo dT) allowed polyadenylated mRNA transcripts to be isolated out of a mixture of RNA. Esquiaqui et al., (2020) have shown that polyA purification of HIV-1 total viral cellular mRNA preferentially enriched for spliced RNA as these have longer polyA tails than unspliced HIV-1 mRNA. Using poly A purification of the total mRNA, spliced RNA was isolated and confirmed by RPA. High throughput sequencing of zipcodes generated from polyA purified cDNA was carried out. Comparing spliced mRNA to unspliced mRNA showed more abundant clones generated about equal amounts of spliced and unspliced RNA as these clones' clustered mostly to the top right of the plot (in blue) in figure 4.11 A. Clones with a higher green proportion indicative of higher GFP+ expression were shifted more towards the unspliced mRNA axis in figure 4.11 B. This observation was interesting as GFP+ is expressed as a Nef spliced product and so the prediction was that clones with a higher green proportion make more spliced mRNA (in green) compared to unspliced if GFP expression was independent of HIV-1 expression. However, the opposite was observed. Clones which released the most virus per GFP+ cell were also shifted more towards the unspliced mRNA axis indicating generating more unspliced mRNA per cell may contribute to a higher virus release per cell (in red) in figure 4.11 plot C.

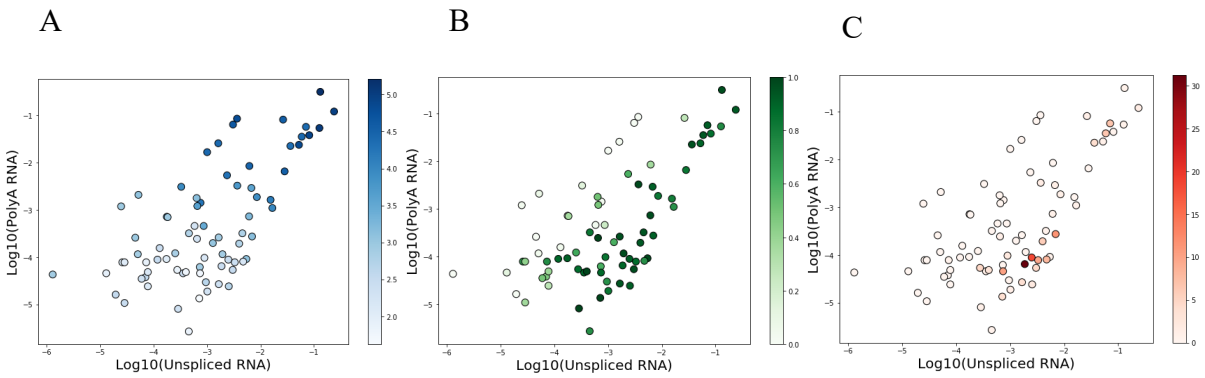

Figure 4.11 Comparison of spliced RNA (polyA purified mRNA) and unspliced mRNA. Most abundant clones generated the most amounts of spliced and unspliced mRNA. Clones which had a higher green proportion were more shifted to the unspliced mRNA axis indicating higher amounts of unspliced RNA. Clones which released more virus per cell generally generated more unspliced RNA. Plot A shows clonal abundance, B shows Green proportions, C virus release

4.12 Chimeric host-provirus RNA generated in HIV-1 gets packaged into virions

Generation of chimeric RNA occurs about 10% of the time during transcription. Determining the fate of these RNAs is fundamental if the virus released per cell for clones is to be determined/calibrated correctly. Regular HIV-1 transcription generates RNA with a conformation of R-U5----//----U3-R (with poly A tail) from provirus with a conformation of U3-R-U5----//----U3-R-U5 (figure 4.121 A and B). This conformation allowed us to distinctively generate probes which could detect chimeric RNA since these RNA subsets deviate from the conformation but will have specific ends irrespective of integration sites of the provirus from which they were generated. To determine if chimeric RNA were generated in the TZ-5 library, two different probes were designed to protect specific chimeric species as RNases A and T1 digested single stranded RNA but RNA which was protected by the antisense probe was not digested. The read-through probe was designed to protect 181 bp of the HIV-1 LTR R-U5 region which could be detected at the 5' LTR region in the case of regular HIV-1 transcription. The U3-R species (240 bp) was detected at the 3' LTR for regular HIV-1 transcription. Detection of the U3-R-U5 species indicated the generation of a read-through or read-in chimeric RNA species (figure 4.121C). A second probe was designed to protect chimeric RNA and most importantly distinguish between read-through and read-in chimeric RNA which was detected. The read-in probe had the added sequences of HIV-1 tRNA lysine primer binding site (PBS) which could only be generated from the 5' end during transcription. For the read-in probe 283 bp of the U3-R-U5 was protected while 342 bp of the U3-R-U5- PBS was protected as shown in figure 4.122. While most of the RNA generated was from regular HIV-1 transcription (indicated by darker bands for U3-R, R-U5 and R-U5-PBS in figure 4.122) in about 1:1 ratio in cells, chimeric read-through and read-in RNA was detected in total viral cellular RNA and viral RNA using both probes. However, more read-in RNA (indicated

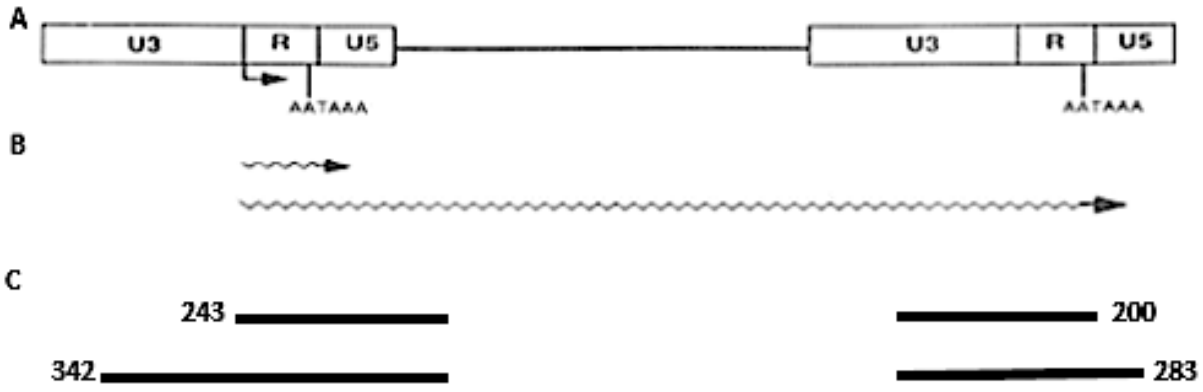

Figure 4.121 Schematic of HIV-1 with the RNA species protected by the read-in probe. A) Schematic showing the organization of the 5' and 3' LTRs with transcription start and the poly adenylation signals. Figure B shows transcription initiation and poly adenylation from the 5' LTR and C shows the chimeric RNA species protected by the read-in probe. Figure was modified from Cherrington and Ganem 1992.

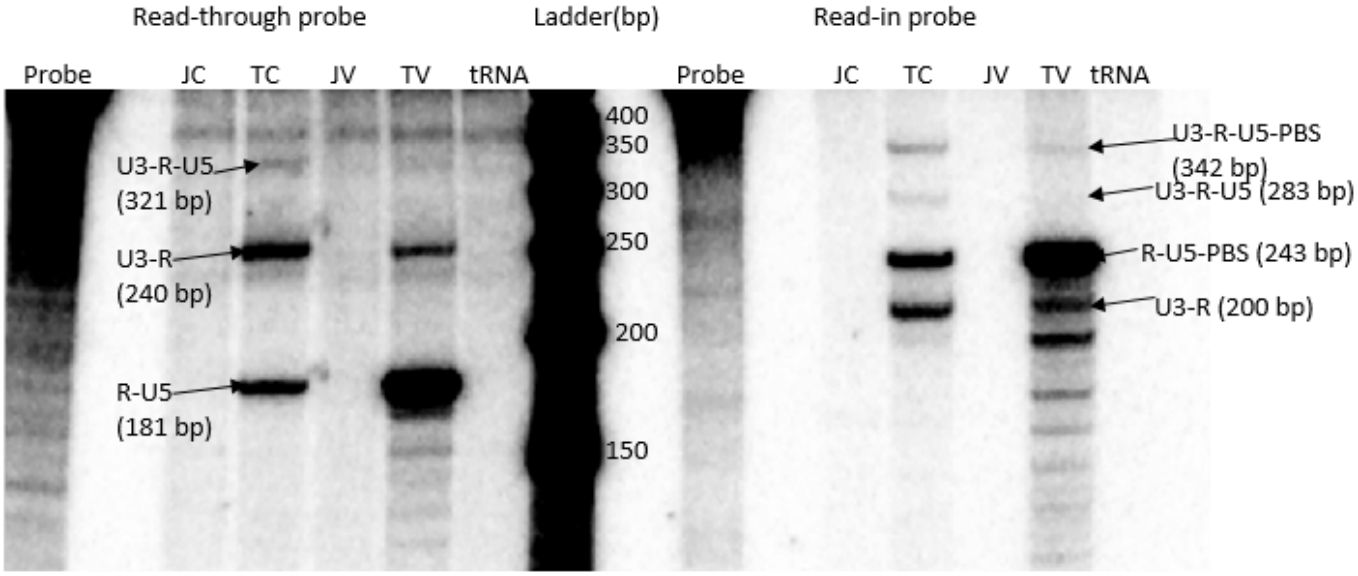

Figure 4.122 Detection of chimeric RNA in cells and virus. Using both the read-through and read-in probe, chimeric RNA could be detected in TZ-5 cells (TC) and virus (TV). Read-in RNA may be generated more frequently than read-through RNA. It was important that the probes did not have complementarity to RNA generated by uninfected Jurkat cells (JC) or RNA extracted from filtered centrifuged media collected from cultured Jurkat cells (JV). Yeast tRNA was used as a carrier in RNA extraction and the gel showed no interaction between tRNA and probes. The unbound probe was about 500 bp distinguishing it from full length bound probe (U3-R-U5 or U3-R-U5-PBS).

by the presence of a darker band) seemed to be generated and packaged into virions than read-through RNA. An interesting phenotype was discovered from the RPA. While the U3-R and R-U5 bands protected in cells were about equal sizes, in viruses a much bigger band was detected for R-U5. This observation was made using both probes. Experiments in section 4.13 were prompted by this observation.

4.13 Chimeric RNA is detected in cells and viral RNA of HIV-1 transfection and infection using different cell lines and constructs

HIV-1 RNA generated from different constructs and cell lines as well as wildtype infection was probed for the presence of chimeric RNA and also to address if equal amounts of R-U5 was detected in virions. This was done to assess the effects of integration within the host genome in driving the generation of host-provirus chimeras and whether it was comparable to a wildtype HIV-1 infection. The same amount of RNA as determined by Nanodrop quantification was used in probe hybridization. All virus samples end with a V while cell samples end with a C. The JME-1 construct was a HIV-1 KFS (construct with a KPN-1 frameshift mutation). The construct was transfected into HEK 293 T-cell to generate RNA. As this is a transfection, HIV-1 was expressed from the unintegrated plasmid in the cell cytoplasm with no modulation from integration sites. GPP was an HIV-1 construct which expressed Gag, Pol with Puromycin driven by the SV-40 promoter. This construct was used in infecting HEK 293 T-cells and puromycin selection carried out for 3 weeks. The TZ-5 construct was same as previously described. It was an infection of Jurkat cells with the TZ-5 library which had a puromycin cassette driven by an EF1-α promoter and expressed GFP. The wildtype infection was carried out by infecting CEM-SS cells with pNL4-3. Comparing both cell and viral RNA from all samples showed chimeric RNA could be generated and detected either in a transient transfection or infection. In figure 4.13 B, it may be confirmed

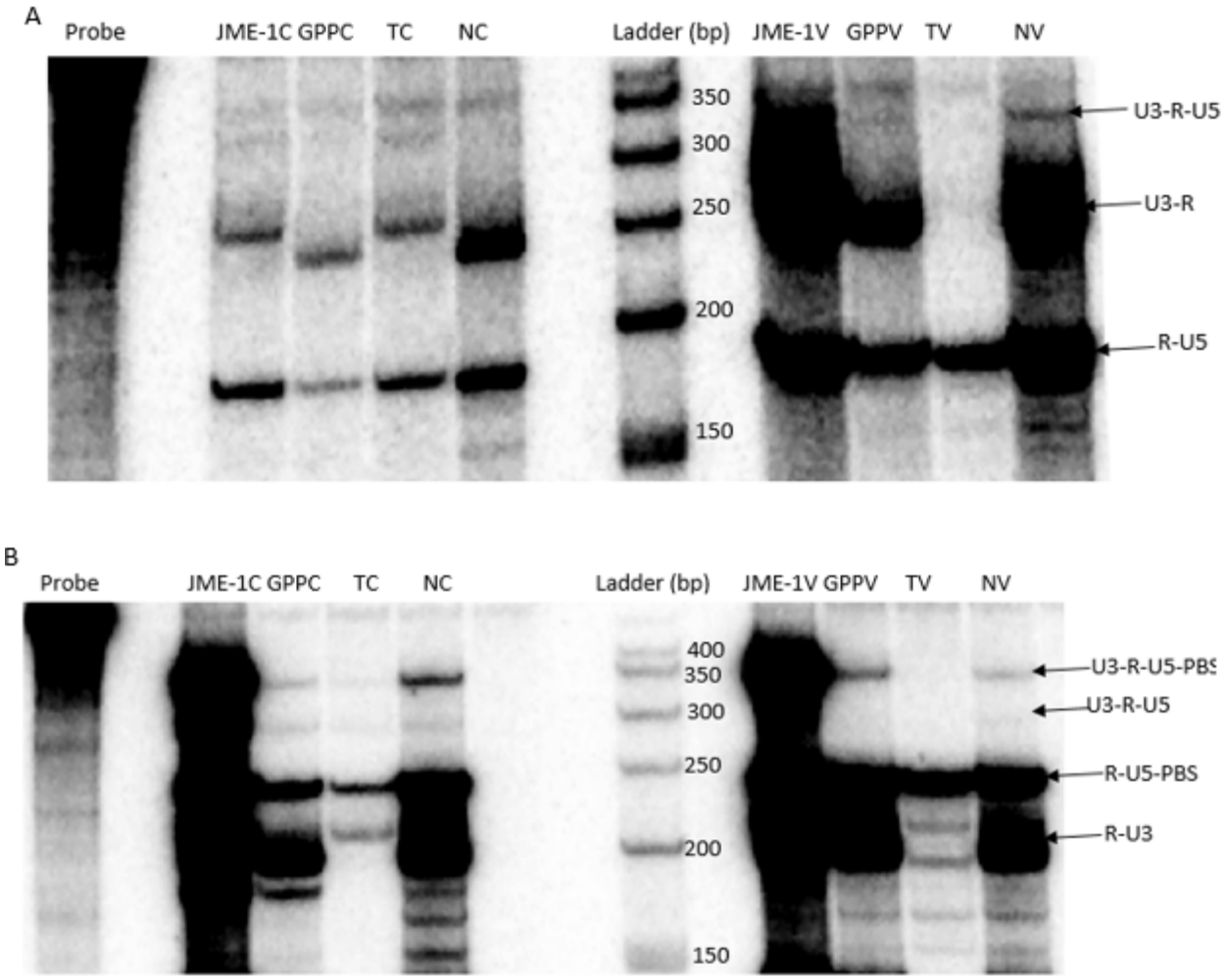

Figure 4.13 Visualization of detected chimeric RNA from different constructs in different cells using A) Read-through probe and B) Read-in probe for transiently transfected HEK 293 T-cells (JME-1), infected and puromycin selected HEK-293 T-cells (GPP), TZ-5 and wildtype infection using pNL4-3. Again read-in chimeric RNA is better detected than read-through suggesting more read-in RNA may be generated than read-through in cells and thus packaged into virus as well. Viral cellular RNA and viral RNA from the GPP construct recapitulated the pattern of detected chimeric RNA in wildtype infection and may be a better suited construct to study the generation of chimeric host-provirus RNA generation and packaging.

that read-in chimeric RNA may be generated in higher quantities than read-through RNA (4.13 A). Comparing across the different constructs used, the GPP infection looked the most similar to wildtype infection in terms of the ratios/ quantities of chimeric RNA generated and packaged into virus. The GPP construct maybe a more realistic construct to use in studying host-provirus chimeric RNA generation and packaging phenomenon in cell lines.

CHAPTER FIVE

DISCUSSION

5.1 Discussion

The current gold standard for estimating the latent viral reservoir, the quantitative viral outgrowth assay (QVOA) relies on the ability to induce cells to release virus to infect lymphoblast from normal healthy donors. This assay has the disadvantage of taking a long time, being labour intensive, time consuming and unfortunately does not detect all replication competent proviruses (Barton and Palmer, 2016). Combining elements of the zipcode library with other variations of this assay may broaden our current understanding of the latent reservoir.

Various assays and techniques with different through-puts and efficiencies have been employed to determine the approximate size of the latent viral reservoir. While a combination approach of sorts is needed in understanding the intricacies of HIV-1 infection, over the last couple years, research efforts have been geared towards understanding the outcome of infection on a cellular level. A review (Kok et al., 2017) of single cell omics reveal different methods designed to study the outcomes of infection in individual cells with the central theme of how a single HIV-1 integration modulates cellular functions.

The zipcoded library could be applied to answer varying research questions such as proviral gene expression levels and abundance, virus release, the extent of latency reversal by different agents on different proviruses among others. The low MOI used in establishing infections allowed for a single integrant per cell and most importantly the presence of the unique 20bp zipcode genetic marker for each integrant makes the zipcoded library a well-suited system to study infection outcomes in cells/clonal populations. PCR and Sanger sequencing techniques have been used to confirm the integration site of the single integrant in the cell with a unique zipcode per cell.

GFP is used as a marker for gene expression in the zipcoded library. GFP is in the *nef* open reading frame which is dependent on the HIV-1 LTR promoter. While GFP expression allows the differentiation of cells into active HIV-1 gene expressing and transcriptionally silent clones, GFP has a half-life of approximately 26 hours in mammalian cells (Corish and Tyler-Smith, 1999). The long half-life of GFP poses a challenge for gene expression studies as cells may have long stopped expressing HIV-1 but GFP is still detected. Persistent GFP may be a particular problem in HIV-1 gene expression studies as Tritel and Resh (2000) observed using pulse chase and gradient centrifugation techniques that majority of the nascent HIV-1 Gag is degraded within 2 hours of synthesis. A subpopulation of newly synthesized Gag was observed to bind membranes in a 5-to-10-minute time frame, aggregate as Gag multimers with time and ultimately end up in packaged virions in a 2-to-6-hour time frame. Results of the intracellular Gag staining performed shows about 97% of cells which expressed GFP stained with intracellular Gag while 98% of Gag negative cells were also GFP negative. Although it cannot be outrightly dismissed, there may be a minor contribution of persistent GFP protein to reporting for HIV-1 gene expression in this system.

FACS analysis of the population of TZ-5 library to sort into GFP positive and negative pools revealed population distribution of 70% GFP+ cells to 30% GFP- cell population. A similar pattern of population gene expression was observed by Read, Atindaana et al, (2019) using the B2 zipcoded library. This distribution may be an indication of minimal gene expression silencing amongst infected cells in an *in vitro* system as compared to patients.

Pollack et al., (2017) in examining the contribution of defective proviruses to detected RNA in supernatant observed supernatant HIV-1 RNA reflected virus production among intact proviruses to a high accuracy as defective proviruses maybe lacking in some gene products necessary for overall virion assembly. GFP+ pool recorded the highest relative p24 value in both experimental

sets as expected while the GFP- pool recorded the least since few to no viruses were expected in the culture supernatant. Spontaneous reactivation and cell sorting and collection errors may account for minimal p24 detected in GFP- supernatant. In experiment 1 of figure 4.3, unsorted total population p24 value reflected about 76.8% of the GFP+ pool estimated indicating the significant p24 contribution is due to the 70% GFP+ population.

After confirmation of virus released among GFP+ cells and the total unsorted pool, RNA was extracted using the Trizol method. All RNA samples were confirmed to be DNA free before proceeding to cDNA synthesis and zipcode PCR. Proliferation of cells was said to drive HIV-1 persistence (Wagner et al., 2014). Integration within oncogenes and genes associated with cellular proliferation favours the generation of clonally expanded cells through proliferation. Cells were observed to be clonally expanded across all the different samples sequenced. Although clonal expansion was observed, no clone accounted for ≥ 18.5% of the raw reads for fraction of abundance as seen in table 4.5. This observation may be due to the relative short period (8 days) culture was allowed to expand after recovery from cryopreservation. Cultures allowed to expand for longer periods may show larger fractions of relative abundance due to much more clonally expanded cells in the population.

Although HIV-1 is known to preferentially integrate into active transcription units (Schroder et al., 2002) clones in the zipcode library were observed to either predominantly express HIV-1 or not based on the green proportions calculated in figure 4.6 A. Ho et al., (2013) observe in non-inducible proviruses isolated from patient samples there was no difference observed between integration into active transcription units and thus integration site differences do not account for differences observed in HIV-1 expression. However, some clones in the zipcode library were observed to be split about halfway between the GFP+ and GFP- pools. Epigenetic signatures at

sites of integration may be acting as an on/off switch in the provirus thus permitting gene expression in some cells in the clone and transcriptionally silencing the provirus some of the time. Repressive chromatin markers have been identified to be recruited to the LTRs of HIV-1 silencing the LTR through deacetylation and methylation of histones (Friedman et al., 2011, Kauder et al., 2009). Although methylation patterns are reported to be faithfully inherited for about 80 cell divisions, studies have also shown the DNA methyltransferase-1 (DNMT1) has a preference for methylating hemi-methylated DNA rather than unmethylated DNA (Ooi and Bestor, 2008). This implies among cells of the same clone, there may be different HIV-1 expression levels based on the different methylation profiles inherited through mitosis. Stable Tat expression during acute infection in Jurkat cells was recorded to prevent the establishment of latent infections (Donahue et al., 2012). When Tat was occluded from cultured cells, cells went into a state of permanent latency (Kamori and Ueno, 2017). Differences in Tat expression among clones may also act as a switch regulating HIV-1 expression.

Among infected cells in patients, it would be predicted that cells which made more virus may not be very abundant since the cells would be more exposed to immune pressure, lysis by virion budding and cytopathic effects of HIV-1 but accurate quantification of virus released is needed in order to assess the effects of latency reversing agents on reactivated latent clones. Clones were tested to determine if all GFP+ cells released virus to the same extent. That is, would any GFP+ clone chosen from the pool release virus in similar amounts to the next or virus release per cell was a distinct phenotype of each clone? This was important because while clonal expansion and cell proliferation are known drivers for HIV-1 persistence, the contribution of these to virus released is not very well defined. It was observed that there were about 5 orders of magnitude difference amongst clones in amounts of virus released per cell in the zipcode library (figure 4.7).

Bui et al., (2016) observed a similar range of virus release per cell from their study on reactivation of inducible proviruses from the latent pool. In the TZ-5 library, the most abundant clones were not responsible for the highest amounts of virus released (figure 4.8) indicating proliferation and clonal expansion advantages may not be the determining factors in amounts of virus released. Rather some low abundance clones were observed on average to release high amounts of virus per cell. Putting both observations together, experimental cultures with high numbers of clonally expanded cells may report low levels of virus released over time. For clones with high virus released per cell, there was less variation between the two replicates implying virus release per cell may be a distinctive phenotype of some clones.

Since proliferation/clonal expansion did not correlate with virus, the fundamental question of how varied GFP expression patterns contributes to the extent of virus release had to be addressed. Different GFP expression patterns were observed for different clones and this was used in generating the green proportions for each clone. Most clones with green proportions above 0.9 were observed to release at least ≥ 10ng/ml p24 per cell in both replicates. Exploiting differences in epigenetic signatures in cells within the same clone, the percentage of cells isolated from a patient which may be reactivated or remain non-inducible in the latent pool can be estimated. To predict the likelihood of a cell to release virus in the latent reservoir, another layer of data analysis for QVOA could be performed where 'green proportions' would be calculated and plotted against the virus released. This may be used as a measure of the decay rate of the latent pool since different clones would be induced to different extents.

For any cell to release virus it first has to be assembled. Virion assembly requires the incompletely spliced Env product along with unspliced Gag and Gag-Pol mRNA. The first intron of HIV-1 is completely spliced out in all spliced transcripts allowing unspliced RNA to be captured by probes

complementary to sequences within the intron. A probe complementary to sequences in HIV-1 RT within the *pol* gene captured RNA with no spliced RNA detected (figure 4.4). Pol capture was also preferred since all RNA captured were expected to at least have an intact RT function and hence be reverse transcription competent. Pollack et al., (2017) showed HIV-1 RNA is detected in proviruses with mutations in Psi, D1, hypermutations and internal deletions. Gag was detected in the supernatant of cells with mutations in D1. The ratio of unspliced RNA generated in the cell to total cell RNA plotted against the total viral cellular RNA in figure 4.9 A and B showed almost no variation amongst abundant clones. Variation was observed amongst least abundant clones. However, clones with higher amounts of unspliced RNA per GFP+ cells were more likely to have more virus released per cell (figure 4.10). Gag is transcribed from unspliced mRNA and Gag is known as the driving machinery behind virion assembly thus it is expected that clones with higher amounts of unspliced RNA detected in the cells have a higher likelihood of higher virus release. Shehu-Xhilaga et al., (2001) have also shown changing the ratios of unspliced RNA Gag to Gag-Pol does not affect HIV-1 RNA packaging and by extension virus release. However, mutations which favour Gag-Pol protein led to decreased virion assembly without packaged genomic RNA. Poly A purification of viral cellular RNA has been shown to selectively enrich for spliced RNA due to tail length differences between spliced and unspliced RNA (Esquiaqui et al., 2020). Comparing spliced RNA against unspliced RNA in the library in figure 4.11, more abundant clones generated equal amounts of unspliced and spliced RNAs confirming clonal abundance does not determine amounts of RNAs and by extension virus release in clones. Clones with higher green proportions were observed to generate higher amounts of unspliced RNA than spliced RNA. This observation conclusively rules out aberrant Nef splicing affecting GFP expression in this library and thus reporting for HIV-1 expression incorrectly. Clones which released the most virus per cell

were observed to also produce more unspliced RNA to spliced RNA and overlapped with clones which had a high green proportion. It was previously observed in this study, higher green proportions predict a higher likelihood of virus release. The higher amounts of virus released may be indicative of the higher amounts of unspliced RNA generated in clones with high green proportions indicating HIV-1 expression. Bringing it all together, clones with higher green proportions have higher amounts of unspliced RNA and may predict higher amounts of virus released per cell.

Recent studies have identified splicing of HIV-1 into host genes triggering the expression of those genes. Specifically, over-expression of BACH 2 and STAT5B in regulatory T-cells in an active infection confer proliferation and survival properties on the cell (Cesana et al., 2017). Although these chimeric RNAs are in the minority, identification of such host provirus chimeras would introduce many more layers in determining correlates of virus release.

HIV-1 has been shown to not distinguish between ends/tails of RNA which gets encapsidated in a dimer, packaging RNA almost twice as long as its genome (Nikolaitchik et al., 2013). Dimerization was observed to be the determining factor for encapsidation as RNA with two copies of Psi or DIS which self-dimerized was packaged. An and Telesnitsky (2004) determined about 7% of read-through HIV-1 RNA was packaged into genomes. Encapsidation of RNA with host tails presents the opportunity to determine the integration sites of provirus and also to determine if certain genomic features were associated with the generation of chimeric RNA. Using zipcoded cells also allowed us to quantify the number of chimeric RNA generated in the clone and packaged as compared to regular RNA.

Using probes complementary to sequences in the LTR of HIV-1 both read-through and read-in chimeric RNA was detected in cells and virus (figure 4.12). In virus RNA we observed a

disproportionate band for R-U5 and a much smaller U3-R although proportionate amounts are observed in cells. This may imply truncated transcription before termination at the 3' LTR. Indeed, Sunshine et al., (2016) using HIV-1 latency models learned most of the detected chimeric RNA were located at the ends of the genomic RNA. Of chimeric RNA sequenced, 94% aligned at the 5' LTR of HIV-1 however, truncated transcripts were identified. Particularly transcripts were identified with host sequences linked to truncated Nef and U3 sequences. Identification of Nef sequences at the 5' host provirus junction indicates RNA which terminated at Nef was used in reverse transcription and interestingly was successful. Truncation of transcription before termination at the 3' LTR may not impair virus replication since HIV is terminally redundant and thus reverse transcription uses the R-U5 portion of the 5' LTR and only the U3 portion of the 3' LTR.

It is noteworthy that the experiments by Sunshine et al (2016) which observed truncated transcripts and this study both used cells with a single provirus integrant. Although the disproportionate amounts of R-U5 RNA detected in virus in the TZ-5 library could be due to the discussed factors, different constructs were tested to determine if the same phenotype was observed across board (figure 4.13). The TZ-5 library was the only one to exhibit this phenotype. Further studies may be needed to validate this result and also tease out the exact mechanism. The GPP construct showed the most similar pattern of detected RNA to the wildtype virus NL4-3. This construct may be better suited and thus can be adapted to study chimeric RNA in cells and virus. Different populations of chimeric RNA in cells with more than one integrated provirus may not show disproportionate amounts of R-U5 phenotype due to recombination.

Active transcription from an upstream transcription machinery in *cis* to an HIV-1 provirus has been shown to switch transcription initiation from 5' LTR to 3' LTR in reactivated CD4+ T-cells

(Lenasi et al., 2008). Promoter occlusion, which is when one elongation complex prevents the assembly of another has also been shown to occur when HIV-1 integrates downstream of an actively transcribed gene. Lenasi and colleagues (2008) report active transcription from an upstream gene may block transcription from the 5' LTR. Blockage of the 5' LTR consequently allows transcription to be activated from the 3' LTR. This phenomenon may explain why we see a much-pronounced read-in band in viruses implying that the more dominant read-in process may be shifting transcription to the 3' LTR. Transcription from the 3' LTR will generate transcripts which lack the Psi and DIS of HIV-1 and thus will not be packaged.

Read-in RNA was observed to be generated to a higher degree than read-through RNA. This phenomenon may be due to orientation of integration of the proviruses as HIV-1 integration in the same orientation as an upstream host gene, has been shown to enhance transcription by approximately 4-fold (Han et al., 2008). Integration in an opposite orientation to transcription of host genes was shown to decrease HIV-1 gene expression by inhibiting the assembly of elongation factors.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

Clonal expansion and cell proliferation have been cited as correlates of HIV-1 persistence. Various methods have been used to estimate the latent viral reservoir with virus release being central to these. HIV-1 can be transcriptionally regulated on several levels. Assembly of HIV-1 virions requires both spliced and unspliced products thus a shift in equilibrium either way may affect virus release. This study evaluated the correlates of virus release among clones expressing GFP.

Gag was co-expressed with GFP in the TZ-5 library with approximately 97% of Gag+ cells being GFP+. About 98% of GFP- cells were also detected to be Gag-. GFP expression was therefore a good correlate for HIV-1 gene expression in the library. Previously determined population GFP expression of the library of 70% GFP+ cells and 30% GFP- in the zipcoded B2 library was also observed in the TZ-5 library. GFP expression may be fixed among clones in the library. Significant amounts of virus were released into culture supernatant for GFP+ culture as compared to GFP- culture indicative of active reverse transcriptase released from lysed cells in supernatants.

High throughput sequencing revealed minimum amounts of clonal expansion/proliferation in the library. Based on clone abundance in GFP+ and GFP- sorts, green proportions were calculated. Clones either had a high green proportion or a low green proportion. A few clones expressed GFP in between the two extremes. Differences in epigenetic signatures due to different inherited patterns may account for variability in GFP expression in some clones. Tat expression levels could also contribute to variability in GFP expression within clone since binding of Tat to the TAR element is necessary for recruiting transcription factors to RNAP II for transcriptional elongation. Next, clonal abundance, green proportions and the amounts of unspliced RNA generated in cells were examined to determine correlates of virus release per cell. There were about five orders of

magnitude difference in amounts of virus released per cell against GFP+ cell abundance. Similar variation has been observed by Bui et al., (2016) from reactivated single integrant infected cells. Amounts of unspliced RNA generated showed a moderate correlation with virus release indicating higher amounts of unspliced RNA per cell in some clones leads to high virus release. Clones with green proportion of ≥ 0.9 were observed to at least have 10ng/ml p24 per cell. Compared to poly A purified spliced RNA, clones were distributed according to abundance. Higher green proportion clones generated higher amounts of unspliced RNA, and this overlapped with high virus release.

Chimeric DNA is generated during transcription about 10% of the time. Read-through and Read-in was detected in TZ-5 viral cellular RNA and packaged RNA in virus. More read-in was detected than Read-through RNA. This observation may be due to chimeric RNA with Psi and DIS being packaged irrespective of the length. Using the zipcode library to detect chimeric RNA in virus detected disproportionate amounts of R-U5 to U3-R. Control experiments with other constructs revealed this phenotype to be distinct to TZ-5 library. Premature termination of transcription may be occurring in this library thus not making it a suitable system to study chimeric RNA generation. The pattern of detected RNA for GPP was most similar to wildtype virus infection and may be more suited to this study.

In conclusion, high green proportions may predict higher virus release per cell amongst clones. While these are the trends, there are significant numbers of clones that do not follow this general trend indicating other parameters must also be contributing to higher virus release. Read-through and read-in RNA are detected in the library but read-in is detected in higher amounts in virus.

In addition to recapitulating the variation in virus release in reactivated latent cells, the zipcode library has the added advantage of being easily adapted and relevant for such studies in different lab settings. Currently, the gold standard for estimating the latent reservoir QVOA is a binary sort

of system where cells are either inducible or non-inducible. As a calibration measure for reactivated and non-induced cells of the latent reservoir, measures of virus release per cell plotted against 'green proportions' which is calculated from HIV-1 expression could be used to estimate the decay rate for the clones.

6.2 Recommendations

The study could be repeated with a library with higher clonal diversity to confirm observations. A shorter half-life GFP could be used in a new library to validate if using the long half-life GFP may contribute to cell sorting errors. Differential levels of Tat expression could be used to study if the HIV-1 expression switch was regulated exclusively by Tat or epigenetic modifications also play a role. Reproducibility of data was found to follow a pattern for most abundant clones. Huge discrepancies were observed among low abundant clones. A limiting dilution could be carried out to isolate rare clones which can be expanded and further characterised.

Premature termination and or RNA degradation may be occurring in some transcripts of the TZ-5 library. Control experiments could be performed substituting SV-40 as the promoter for puromycin in place of EF-1α promoter. A qPCR assay could be designed to estimate amounts of chimeric RNA generated within the clone compared to full length transcripts.

REFERENCES

1. Abram, M. E., Ferris, A. L., Das, K., Quinoñes, O., Shao, W., Tuske, S., ... & Hughes, S. H. (2014). Mutations in HIV-1 reverse transcriptase affect the errors made in a single cycle of viral replication. *Journal of virology*, *88*(13), 7589-7601.
2. An, W., & Telesnitsky, A. (2004). Human immunodeficiency virus type 1 transductive recombination can occur frequently and in proportion to polyadenylation signal readthrough. *Journal of virology*, *78*(7), 3419-3428.
3. Archin, N. M., Keedy, K. S., Espeseth, A., Dang, H., Hazuda, D. J., & Margolis, D. M. (2009). Expression of latent human immunodeficiency type 1 is induced by novel and selective histone deacetylase inhibitors. *AIDS (London, England)*, *23*(14).
4. Archin, N. M., Liberty, A. L., Kashuba, A. D., Choudhary, S. K., Kuruc, J. D., Crooks, A. M., ... & Richman, D. D. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, *487*(7408), 482.
5. Arts, E. J., & Hazuda, D. J. (2012). HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine*, a007161.
6. Barton, K. M., & Palmer, S. E. (2016). How to define the latent reservoir: tools of the trade. *Current HIV/AIDS Reports*, *13*(2), 77-84.
7. Barton, K., Winckelmann, A., & Palmer, S. (2016). HIV-1 reservoirs during suppressive therapy. *Trends in microbiology*, *24*(5), 345-355.
8. Besnard, E., Hakre, S., Kampmann, M., Lim, H. W., Hosmane, N. N., Martin, A., ... & Svensson, J. P. (2016). The mTOR complex controls HIV latency. *Cell host & microbe*, *20*(6), 785-797.
9. Blood, G. A. C. (2016). Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy*, *43*(3), 203.

10. Briggs, J. A., Wilk, T., Welker, R., Kräusslich, H. G., & Fuller, S. D. (2003). Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO journal*, *22*(7), 1707-1715.

11. Bronke, C., Almeida, C. A. M., McKinnon, E., Roberts, S. G., Keane, N. M., Chopra, A., ... & John, M. (2013). HIV escape mutations occur preferentially at HLA-binding sites of CD8 T-cell epitopes. *AIDS (London, England)*, *27*(6).

12. Brown, T. R. (2015). I am the Berlin patient: a personal reflection. *AIDS research and human retroviruses*, *31*(1), 2-3.

13. Bruner, K. M., Hosmane, N. N., & Siliciano, R. F. (2015). Towards an HIV-1 cure: measuring the latent reservoir. *Trends in microbiology*, *23*(4), 192-203.

14. Bui, J. K., Mellors, J. W., & Cillo, A. R. (2016). HIV-1 virion production from single inducible proviruses following T-cell activation ex vivo. *Journal of virology*, *90*(3), 1673-1676.

15. Carter, J., & Saunders, V. A. (2007). *Virology: principles and applications*. John Wiley & Sons.

16. Cesana, D., de Sio, F. R. S., Rudilosso, L., Gallina, P., Calabria, A., Beretta, S., ... & Vicenzi, E. (2017). HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. *Nature communications*, *8*(1), 498.

17. Chamond, N., Locker, N., & Sargueil, B. (2010). The different pathways of HIV genomic RNA translation.

18. Chen, H. C., Martinez, J. P., Zorita, E., Meyerhans, A., & Filion, G. J. (2017). Position effects influence HIV latency reversal. *Nature structural & molecular biology*, *24*(1), 47.

19. Chen, Z., Telfier, P., Gettie, A., Reed, P., Zhang, L., Ho, D. D., & Marx, P. A. (1996). Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *Journal of virology*, *70*(6), 3617-3627.

20. Cherrington, J., & Ganem, D. (1992). Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *The EMBO journal*, *11*(4), 1513-1524.

21. Corish, P., & Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein engineering*, *12*(12), 1035-1040.

22. Crabtree, G. R. (1989). Contingent genetic regulatory events in T lymphocyte activation. *Science*, *243*(4889), 355-361.

23. Crooks, A. M., Bateson, R., Cope, A. B., Dahl, N. P., Griggs, M. K., Kuruc, J. D., ... & Archin, N. M. (2015). Precise quantitation of the latent HIV-1 reservoir: implications for eradication strategies. *The Journal of infectious diseases*, *212*(9), 1361-1365.

24. Dilley, K. A., Nikolaitchik, O. A., Galli, A., Burdick, R. C., Levine, L., Li, K., ... & Hu, W. S. (2017). Interactions between HIV-1 Gag and viral RNA genome enhance virion assembly. *Journal of virology*, *91*(16), e02319-16.

25. Donahue, D. A., Kuhl, B. D., Sloan, R. D., & Wainberg, M. A. (2012). The viral protein Tat can inhibit the establishment of HIV-1 latency. *Journal of virology*, *86*(6), 3253-3263.

26. Emery, A., Zhou, S., Pollom, E., & Swanstrom, R. (2017). Characterizing HIV-1 splicing by using next-generation sequencing. *Journal of Virology*, *91*(6).

27. Eriksson, S., Graf, E. H., Dahl, V., Strain, M. C., Yukl, S. A., Lysenko, E. S., ... & Abdel-Mohsen, M. (2013). Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS pathogens*, *9*(2), e1003174.

28. Esquiaqui, J. M., Kharytonchyk, S., Drucker, D., & Telesnitsky, A. (2020). HIV-1 spliced RNAs display transcription start site bias. *RNA*, *26*(6), 708-714.

29. Février, M., Dorgham, K., & Rebollo, A. (2011). CD4+ T cell depletion in human immunodeficiency virus (HIV) infection: role of apoptosis. *Viruses*, *3*(5), 586-612.

30. Fletcher, C. V., Staskus, K., Wietgrefe, S. W., Rothenberger, M., Reilly, C., Chipman, J. G., ... & Anderson, J. (2014). Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proceedings of the National Academy of Sciences*, *111*(6), 2307-2312.

31. Friedman, J., Cho, W. K., Chu, C. K., Keedy, K. S., Archin, N. M., Margolis, D. M., & Karn, J. (2011). Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. *Journal of virology*, *85*(17), 9078-9089.

32. Ganser-Pornillos, B. K., Yeager, M., & Sundquist, W. I. (2008). The structural biology of HIV assembly. *Current opinion in structural biology*, *18*(2), 203-217.

33. Gee, A. H., Kasprzak, W., & Shapiro, B. A. (2006). Structural differentiation of the HIV-1 Poly (A) Signals. *Journal of Biomolecular Structure and Dynamics*, *23*(4), 417-428.

34. Geels, M. J., Cornelissen, M., Schuitemaker, H., Anderson, K., Kwa, D., Maas, J., ... & van Beelen, M. (2003). Identification of sequential viral escape mutants associated with altered T-cell responses in a human immunodeficiency virus type 1-infected individual. *Journal of virology*, *77*(23), 12430-12440.

35. Global HIV and AIDS statistics - 2018 factsheet UNAIDS http://www.unaids.org/en/resources/fact-sheet accessed February 2019.

36. Greger, I. H., Proudfoot, N. J., Demarchi, F., & Giacca, M. (1998). Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic acids research*, *26*(5), 1294-1300.

37. Gupta, R. K., Abdul-jawad, S., McCoy, L. E., Mok, H. P., Peppa, D., Salgado, M., ... & Grant, P. (2019). HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. *Nature*, 1.

38. Han, Y., Lin, Y. B., An, W., Xu, J., Yang, H. C., O'Connell, K., ... & Siliciano, R. F. (2008). Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell host & microbe*, *4*(2), 134-146.

39. Henderson, A., Holloway, A., Reeves, R., & Tremethick, D. J. (2004). Recruitment of SWI/SNF to the human immunodeficiency virus type 1 promoter. *Molecular and cellular biology*, *24*(1), 389-397.

40. Henrich, T. J., Hanhauser, E., Marty, F. M., Sirignano, M. N., Keating, S., Lee, T. H., ... & Hill, A. L. (2014). Antiretroviral-free HIV-1 remission and viral rebound after allogeneic stem cell transplantation: report of 2 cases. *Annals of internal medicine*, *161*(5), 319-327.

41. Hiener, B., Horsburgh, B. A., Eden, J. S., Barton, K., Schlub, T. E., Lee, E., ... & Sinclair, E. (2017). Identification of genetically intact HIV-1 proviruses in specific CD4+ T cells from effectively treated participants. *Cell reports*, *21*(3), 813-822.

42. Ho, Y. C., Shan, L., Hosmane, N. N., Wang, J., Laskey, S. B., Rosenbloom, D. I., ... & Siliciano, R. F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*, *155*(3), 540-551.

43. Jain, V., Hartogensis, W., Bacchetti, P., Hunt, P. W., Hatano, H., Sinclair, E., ... & Pilcher, C. D. (2013). Antiretroviral therapy initiated within 6 months of HIV infection is associated with lower T-cell activation and smaller HIV reservoir size. *The Journal of infectious diseases*, *208*(8), 1202-1211.

44. Jordan, A., Bisgrove, D., & Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *The EMBO journal*, *22*(8), 1868-1877.

45. Josefsson, L., von Stockenstrom, S., Faria, N. R., Sinclair, E., Bacchetti, P., Killian, M., ... & Shao, W. (2013). The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proceedings of the National Academy of Sciences*, *110*(51), E4987-E4996.

46. Kamori, D., & Ueno, T. (2017). HIV-1 Tat and viral latency: What we can learn from naturally occurring sequence variations. *Frontiers in microbiology*, *8*, 80.

47. Kaneko, S., & Manley, J. L. (2005). The mammalian RNA polymerase II C-terminal domain interacts with RNA to suppress transcription-coupled 3′ end formation. *Molecular cell*, *20*(1), 91-103.

48. Karn, J., & Stoltzfus, C. M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harbor perspectives in medicine*, *2*(2), a006916.

49. Kauder, S. E., Bosque, A., Lindqvist, A., Planelles, V., & Verdin, E. (2009). Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS pathogens*, *5*(6), e1000495.

50. Kharytonchyk, S., Monti, S., Smaldino, P. J., Van, V., Bolden, N. C., Brown, J. D., ... & Summers, M. F. (2016). Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proceedings of the National Academy of Sciences*, *113*(47), 13378-13383.

51. Kok, Y. L., Ciuffi, A., & Metzner, K. J. (2017). Unravelling HIV-1 latency, one cell at a time. *Trends in microbiology*, *25*(11), 932-941.

52. Lenasi, T., Contreras, X., & Peterlin, B. M. (2008). Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell host & microbe, 4*(2), 123-133.

53. Luzuriaga, K., Gay, H., Ziemniak, C., Sanborn, K. B., Somasundaran, M., Rainwater-Lovett, K., ... & Persaud, D. (2015). Viremic relapse after HIV-1 remission in a perinatally infected child. *New England Journal of Medicine*, *372*(8), 786-788.

54. Maldarelli, F., Wu, X., Su, L., Simonetti, F. R., Shao, W., Hill, S., ... & Coffin, J. M. (2014). Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*, *345*(6193), 179-183.

55. McNamara, L. A., & Collins, K. L. (2011). Hematopoietic stem/precursor cells as HIV reservoirs. *Current opinion in HIV and AIDS*, *6*(1), 43.

56. Merson, M. H., O'Malley, J., Serwadda, D., & Apisuk, C. (2008). The history and challenge of HIV prevention. *The lancet*, *372*(9637), 475-488.

57. Moudgil, T., & Daar, E. S. (1993). Infectious decay of human immunodeficiency virus type 1 in plasma. *Journal of Infectious Diseases*, *167*(1), 210-212.

58. Muriaux, D., Fossé, P., & Paoletti, J. (1996). A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry*, *35*(15), 5075-5082.

59. Nikolaitchik, O. A., Dilley, K. A., Fu, W., Gorelick, R. J., Tai, S. H. S., Soheilian, F., ... & Hu, W. S. (2013). Dimeric RNA recognition regulates HIV-1 genome packaging. *PLoS pathogens*, *9*(3), e1003249.

60. Nyamweya, S., Hegedus, A., Jaye, A., Rowland-Jones, S., Flanagan, K. L., & Macallan, D. C. (2013). Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Reviews in medical virology*, *23*(4), 221-240.

61. Ohlmann, T., Mengardi, C., & López-Lastra, M. (2014). Translation initiation of the HIV-1 mRNA. *Translation*, *2*(2), e960242.

62. Onafuwa-Nuga, A., & Telesnitsky, A. (2009). The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiology and Molecular Biology Reviews*, *73*(3), 451-480.

63. Ooi, S. K., & Bestor, T. H. (2008). Cytosine methylation: remaining faithful. *Current Biology*, *18*(4), R174-R176.

64. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., & Ho, D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, *271*(5255), 1582-1586.

65. Pinzone, M. R., VanBelzen, D. J., Weissman, S., Bertuccio, M. P., Cannon, L., Venanzi-Rullo, E., ... & Groen, K. (2019). Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nature communications*, *10*(1), 728.

66. Pollack, R. A., Jones, R. B., Pertea, M., Bruner, K. M., Martin, A. R., Thomas, A. S., ... & Hao, H. (2017). Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. *Cell host & microbe*, *21*(4), 494-506.

67. Procopio, F. A., Fromentin, R., Kulpa, D. A., Brehm, J. H., Bebin, A. G., Strain, M. C., ... & Hoh, R. (2015). A novel assay to measure the magnitude of the inducible viral reservoir in HIV-infected individuals. *EBioMedicine*, *2*(8), 874-883.

68. Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, *108*(4), 501-512.

69. Read, D. F., Atindaana, E., Pyaram, K., Yang, F., Emery, S., Cheong, A., ... & Kidd, J. M. (2019). Stable integrant-specific differences in bimodal HIV-1 expression patterns revealed by high-throughput analysis. *PLoS pathogens*, *15*(10), e1007903.

70. Reeves, D. B., Duke, E. R., Wagner, T. A., Palmer, S. E., Spivak, A. M., & Schiffer, J. T. (2018). A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nature communications*, *9*(1), 4811.

71. Roberts, J. D., Bebenek, K., & Kunkel, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science*, *242*(4882), 1171-1173.

72. Rowland-Jones, S. L., & Whittle, H. C. (2007). Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1?. *Nature immunology*, *8*(4), 329.

73. Saliou, J. M., Bourgeois, C. F., Ayadi-Ben Mena, L., Ropers, D., Jacquenet, S., Marchand, V., ... & Branlant, C. (2009). Role of RNA structure and protein factors in the control of HIV-1 splicing. *Front Biosci (Landmark Ed)*, *14*, 2714-2729.

74. Sallusto, F., Geginat, J., & Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.*, *22*, 745-763.

75. Schiralli Lester, G. M., & Henderson, A. J. (2012). Mechanisms of HIV transcriptional regulation and their contribution to latency. *Molecular biology international*, *2012*.

76. Schröder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, *110*(4), 521-529.

77. Sebastian, N. T., & Collins, K. L. (2014). Targeting HIV latency: resting memory T cells, hematopoietic progenitor cells and future directions. *Expert review of anti-infective therapy*, *12*(10), 1187-1201.

78. Sharp, P. M., & Hahn, B. H. (2010). The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2487-2494.

79. Shaw, G. M., & Hunter, E. (2012). HIV transmission. *Cold Spring Harbor perspectives in medicine*, a006965.

80. Shehu-Xhilaga, M., Crowe, S. M., & Mak, J. (2001). Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *Journal of virology*, *75*(4), 1834-1841.

81. Sunshine, S., Kirchner, R., Amr, S. S., Mansur, L., Shakhbatyan, R., Kim, M., ... & Sui, S. H. (2016). HIV integration site analysis of cellular models of HIV latency with a probe-enriched next-generation sequencing assay. *Journal of virology*, *90*(9), 4511-4519.

82. Tritel, M., & Resh, M. D. (2000). Kinetic analysis of human immunodeficiency virus type 1 assembly reveals the presence of sequential intermediates. *Journal of virology*, *74*(13), 5845-5855.

83. Verdin, E., Paras Jr, P., & Van Lint, C. (1993). Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *The EMBO journal*, *12*(8), 3249-3259.

84. Vermeire, J., Naessens, E., Vanderstraeten, H., Landi, A., Iannucci, V., Van Nuffel, A., ... & Verhasselt, B. (2012). Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti-and retroviral vectors. *PloS one*, *7*(12), e50859.

85. Wagner, T. A., McLaughlin, S., Garg, K., Cheung, C. Y., Larsen, B. B., Styrchak, S., ... & Frenkel, L. M. (2014). Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*, *345*(6196), 570-573.

86. Wang, Z., Gurule, E. E., Brennan, T. P., Gerold, J. M., Kwon, K. J., Hosmane, N. N., ... & Ho, Y. C. (2018). Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. *Proceedings of the National Academy of Sciences*, *115*(11), E2575-E2584.

87. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., & Jones, K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, *92*(4), 451-462.

88. West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5′→ 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, *432*(7016), 522.

89. Whitney, J. B., Hill, A. L., Sanisetty, S., Penaloza-MacMaster, P., Liu, J., Shetty, M., ... & Smith, J. Y. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature*, *512*(7512), 74.

90. Woodham, A. W., Skeate, J. G., Sanna, A. M., Taylor, J. R., Da Silva, D. M., Cannon, P. M., & Kast, W. M. (2016). Human immunodeficiency virus immune cell receptors, coreceptors, and cofactors: Implications for prevention and treatment. *AIDS patient care and STDs*, *30*(7), 291-306.

91. World Health Organization. (2018). Global health observatory (GHO) data.

92. Wu, G., & Zaman, M. H. (2012). Low-cost tools for diagnosing and monitoring HIV infection in low-resource settings. *Bulletin of the World Health Organization*, *90*, 914-920.

93. Wu, Y. (2004). HIV-1 gene expression: lessons from provirus and non-integrated DNA. *Retrovirology*, *1*(1), 13.

APPENDIX

Cell sorting

Entire procedure was carried out on ice or at 4°C

1. Centrifuge cells to pellet at 2000 g for 5 minutes at 4 °C
2. Aliquot cell pellets into 96 well plates
3. Wash cell pellet 2x with 100 ul PBS with 10% FBS and centrifuge at 1500 rpm for 2 minutes
4. Discard supernatant
5. Add 1ul of propidium iodide to 1 ml of PBS with 10% FBS and dispense 100ul to wells with cells
6. Incubate at on ice in the dark for 20 mins
7. Wash cells 2x with 100 ul PBS with 10% FBS and discard supernatant
8. Transfer cells into polystyrene tubes in a total of 200 ul of PBS with 10% FBS
9. Proceed to cell sorting

Harvesting of viral media and cell pellets

Viral media was harvested by centrifuging culture at 2000 g for 5 min. The supernatant was collected and gently filtered through a 0.2-micron filter to eliminate all cells present. Amount of virus released was quantified by qRT-PCR. Viral RNA was extracted from the media.

Plasmids and primers

The plasmids generated for the read-through and read-in probes have been sequenced to confirm the plasmid and have been banked in the Telesnitsky plasmid bank TL# 2336, 2337 respectively. The data sheets of all primers received with their respective names have been filed in my lab notebook.

RT assay

The relative p24 concentration was used to estimate reverse transcriptase concentration being in the virus.

1. An aliquot of 5 ul of viral media was put into Eppendorf tubes
2. Add 5 ul of 2x lysis buffer and incubate at room temperature for 15 minutes
3. Add 90 ul of 1x qPCR buffer
4. Carry out 3 5-fold serial dilutions of viral media and 5 serial dilutions of p24 standard
5. Prepare PCR master mix by adding components described below

Reagent	**Volume (µL) per 1 reaction**
DEPC-H_20	12.2
10x qPCR buffer	2
10mM dNTPs	0.4
25mM MgCl2	1.6
40µM Forward MS2 primer	0.15
40µM Reverse MS2 primer	0.15
0.8 µg/µL MS2 RNA	0.1
100 mM DTT	0.2
20x EvaGreen	1
RNase inhibitor (rRNasin)	0.2
Promega GoTaq	0.2

6. Aliquot 18 ul of master mix into qRT PCR plates specific for light cycler
7. Add 2 ul of sample to respective wells *

*Each sample is run in duplicate

8. Centrifuge plate at 1500 rpm for 2 min and put into light cycler

9. Check for successful self-test
10. Run the program as described below

Program	**Cycle(s)**	**Temperature (˚C)**	**Time (min)**
Pre-Incubation	1	37	30
3 step amplification	45	95	10
Melting	1	-	-

Viral RNA extraction

Viral RNA (with 1ml 20% sucrose and PBS) was centrifuged at 25000 rpm for 2 hours at 4 degrees Celsius and the Trizol extraction carried out as follows

1. An aliquot of 500 ul Trizol was used to resuspend viral pellet and allowed to sit for 5 minutes. Transfer to a 1.5 ml Eppendorf tube
2. Add 100 ul of Chloroform Isoamyl Alcohol (24:1) per 500 ul Trizol used.
3. Shake tube by hand vigorously for 15 secs and incubate at room temperature for 2- 3 minutes
4. Centrifuge the mixture at top speed for 15 mins at 4 degrees Celsius. The upper layer contains your RNA
5. Remove aqueous layer into another Eppendorf tube and add 5-10 ug of RNase free glycogen
6. Add 250 ul of Isopropanol per 500 ul of Trizol used and incubate at room temperature for 10 minutes
7. Centrifuge at top speed for 10 min at 4 °C.
8. Wash RNA pellet with 500 ul 75% EtOH per 500 ul Trizol used.

*Pellet can be stored at -20 °C

9. Dry pellet for 5- 10 minutes at room temperature

10. DNase treatment

Add 50 ul DNase mix and incubate for 15 minutes at 37°C.

5ul 1M $MgCl_2$

1ul 0.5M DTT

1ul RNase-free DNase RQ1 DNase (Promega)

0.25ul rRnasin (40U/ul Promega)

42 ul TE

11. Add 25ul DNase stop mix to each sample

2.5ul 0.5M EDTA

12.5ul 3M NaOAc

2.5ul 10% SDS

7.5ul DEPC-H_2O

12. Phenol Chloroform Isoamyl alcohol extraction

Add equal volume of Phenol: Chloroform: Isoamyl alcohol 125ul total, mix and centrifuge at top speed for 3 min. Pipette the top RNA layer.

Add 2.5 volumes 100% EtOH (312ul virus), mix, freeze -20°C at least 30min., centrifuge top speed for 15 min. at 4c, dry pellet 5-10 min. RT

Resuspend RNA pellet in 20ul TE

PCR amplification of zipcode regions

Zipcode regions were amplified from genomic DNA and cDNA using Phusion PCR as described below. Carryout entire procedure on ice. Pre-set thermocycler to heat up.

Master Mix

Reagent	Volume (ul) per 1 sample
5x Phusion buffer	5
10 mM dNTP	0.4
10 uM forward primer	1
10 uM reverse primer	1
Phusion DNA polymerase	0.2
PCR water	10.4
Template DNA (100 ng)	2
Total volume	20

PCR Cycle conditions

Step	Temperature (°C)	Time (secs)	Number of cycles
Initial denaturation	98	30	1
Denaturation	98	10	25
Annealing	59	10	25
Elongation	72	10	25
Final extension	72	240	1
Hold	4		