Host Transcriptome Dynamics During HIV-1 Latency

Thesis submitted in partial fulfilment for
the degree of

Master of Science

as a part of the

Integrated Ph. D. program
(Biological Science)

by

Chhavi Saini

HIV-AIDS laboratory
Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
(*A Deemed University*)

Jakkur, Bengaluru-560064, India.

March 2019

I dedicate this thesis to my Family.

Declaration

I hereby declare that the work enveloped in this thesis entitled, "**Host Transcriptome Dynamics During HIV-1 Latency**" has been the result of constant investigations carried out by me under the supervision of **Prof. Udaykumar Ranga** at the HIV-AIDS Laboratory, Molecular and Genetics Laboratory, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India and the work has not been submitted elsewhere for the award of any degree or diploma.

In keeping with the general practice in reporting the scientific observations, due acknowledgement has been made whenever the work described is based on the findings of other investigators. Any omission that might have occurred due to oversight or error in judgement is regretted.

Chhavi Saini

Certificate

I hereby certify that the work described in this thesis entitled, "**Host Transcriptome Dynamics During HIV-1 Latency**" is result of the investigation carried out by **Miss. Chhavi Saini** towards her Master of Science as part of the Integrated PhD program at the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India under my guidance and that the results presented in the thesis have not previously formed the basis for the award of any other diploma, degree or fellowship.

Dr. Udaykumar Ranga
Professor,
HIV-AIDS Laboratory,
Molecular Biology & Genetics Unit,
Jawaharlal Nehru Centre for Advanced Scientific Research.
Ph: +91-80-22082830
E-mail: udaykumar@jncasr.ac.in

Acknowledgment

I would like to express my sincere gratitude to my guide, Prof. Udaykumar Ranga who gave me a wonderful opportunity to work in one of the emerging and crucial field of HIV-1 biology. His constant guidance and insightful discussions inspired me to pursue this field of research and have helped to sharpen my scientific ideas.

I thank Prof. Ravi Manjithaya, Prof. James Chelliah, Prof. Hemalatha Balaram, Prof. Manisha Inamdar, Prof. A. Anand, prof. Tapas Kumar Kundu, Prof. Sheeba Vasu, Prof. N.K Subhedar, Prof. G.R Ramesh, Prof. Namita Surolia, Prof. Kaustuv Sanyal and Prof. V Nagaraja for their course work. I am very much thankful to Dr. Kushagra Bansal for his invaluable guidance in NGS data analysis which is an essential part of my work.

I would like to thank seniors in the laboratory for their constant support and guidance – Disha Bhange, Arun Panchapakesan, Malini Menon, Sutanuka Das, Shrestha Pal, Deepak Selvam. I would like to thank other lab members for making environment relaxed and for their help whenever needed – A. Brahma, Haider Ali, Kavita, Yuvraj, Neelam, Swati, Swathi, Jyotsna, Ishfaq, Vijeta, Anish. I thank my lab assistant Sridhar and Lakappa for their help whenever needed. I am grateful to Harshit Prajapati and Archisman Banerjee for help in data analysis. I am very much thankful to Dr. Narendra Nala for his great help in flow cytometry.

I also thank my friends and seniors in JNCASR Rashi, Shubham, Ankit, Kuladeep, Saheli, Irine, Harshit, Geetika, Vijay, Preeti, Anjali, Divyesh, Laxmi Mam, Utsa, Arnab and Pallabi for their encouragement and for all the wonderful time we had together.

As always, I am deeply obliged to my family – Mom, Dad, and Roshy for their constant support, encouragement and well wishes.

Contents

Abbreviation

AIDS	Acquired immunodeficiency syndrome
ANOVA	Analysis of Variance
AP-1	Activator protein-1
BHQ	Black hole quencher
CD	Cluster of Differentiation
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CDK	Cyclin dependent kinase
CRF	Circulating recombinant forms
CTD	C-terminal domain
DEAE	Diethylaminomethyl
D2EGFP	Destabilised enhanced green fluorescent protein
DMEM	Dulbecco's Modified Eagle's Medium
DSIF	DRB sensitivity inducing factor
EF1α	Elongation factor-1 alpha
EGFP	Enhanced green fluorescent protein
FACS	Flow activated cell sorting
FAM	Fluorescein amidite
FBS	Fetal bovine serum
HDAC	Histone deacetylase
HDACi	Histone deacetylase inhibitor
HEXIM	Hexamethylene bisacetamide-inducible protein
HIV-1	Human immunodeficiency virus
HMBA	Hexamethylene bisacetamide
IRES	Internal ribosome entry site
LEF-1	Lymphoid enhancer-binding factor-1
NELF	Negative elongation factor
NFAT	Nuclear factor of activated T-cell
NF-κB	Nuclear factor-KappaB
NGS	Next Generation Sequencing

PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffer saline
qPCR	Quantitative polymerase chain reaction
PIC	Protease inhibitor cocktail
PMA	Phorbol-12-myristate-13-acetate
P-TEFb	Positive transcription Elongation factor b
PTM	Post translational modifications
RNA-seq	RNA sequencing
RPMI	Roswell Park Memorial Institute
snRNP	Small nuclear ribonucleoprotein
Sp1	Specificity protein 1
TAR	Transactivation response element
Tat	Transactivator of transcription
TNFα	Tumor necrosis factor alpha
TSA	Trichostatin
U3	Unique to 3’
U5	Unique to 5’
VSV-G	Vesicular stomatitis virus glycoprotein

Chapter 1

Introduction

Chapter 1: Introduction

1.1 General

Human Immunodeficiency Virus (HIV) infects host immune cells harboring CD4 receptors, like $CD4^+$ T-cells, macrophages and dendritic cells, this weakens the immune system and with time leads to a condition known as Acquired Immunodeficiency Syndrome (AIDS). Based on genetic characteristics and viral antigens HIV is classified into – HIV-1 and HIV-2. HIV-1 has four groups – M (major), N (nonmajor), O (outlier) and P. HIV-1M is further divided into various subgroups or clades, which are nearly at same genetic distance from each other – A-D, F-H, J and K; subtype C is predominant and responsible for 50% of global infections [1]. There are also circulating recombinant forms of the virus within the M group, formed in dually infected individuals. (Figure 1).

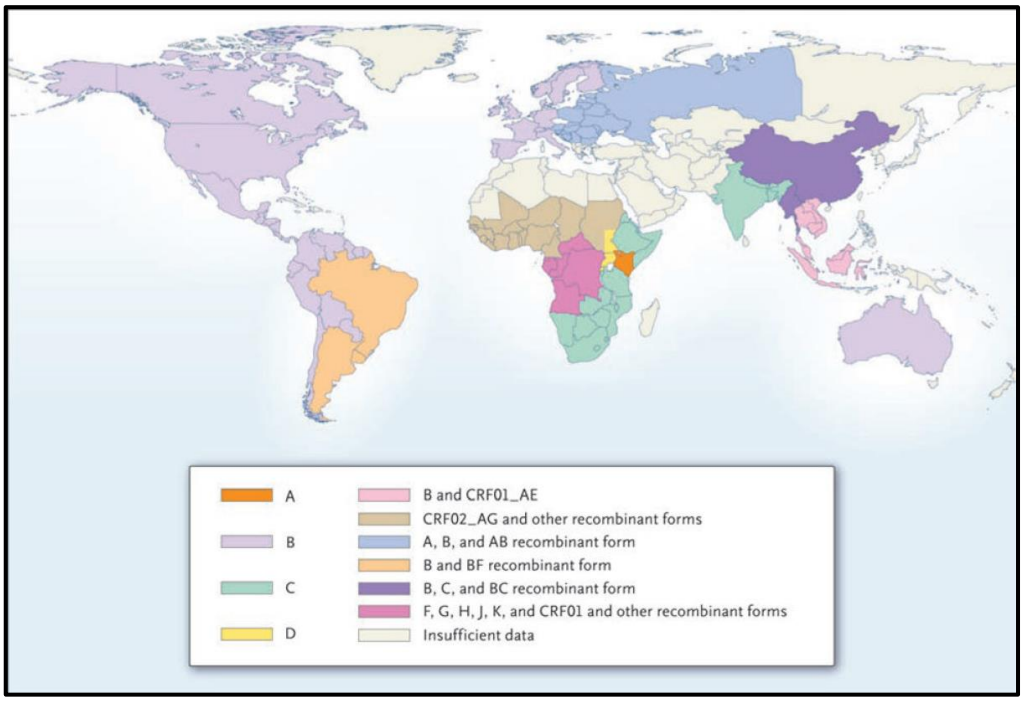

Figure 1. Global Distribution of HIV-1 Subtypes and Recombinant Forms
(Reproduced with permission from Taylor, Barbara S., et al 2008, copyright © from Massachusetts Medical Society)

HIV-1 encodes fifteen proteins expressed as nine ORFs. Structural protein Gag which gets processed into – MA (matrix), CA (capsid) and NC (nucleocapsid) and p6; and Env – SU

(surface, gp120) and TM (transmembrane, gp41) forms the virion core and outer envelope. HIV-1 replication and integration into the host genome are mediated by viral reverse transcriptase and integrase. Other essential proteins Tat and Rev are required for viral mRNA export and gene regulation respectively. Accessory proteins like Vif, Nef, Vpr, Vpu are essential for virion assembly (figure 2) [2]. To enter in the host cell HIV-1 binds to CD4 and its co-receptor CCR5, in the cytoplasm virus reverse transcribes its ssRNA to dsDNA by viral encoded reverse transcriptase, and then pre-integration complex mediates the entry of dsDNA into the nucleus followed by its integration in the host genome. This provirus can stay in transcriptionally active form, producing more viral particles or can enter in a transcriptionally inactive form, known as the latency.

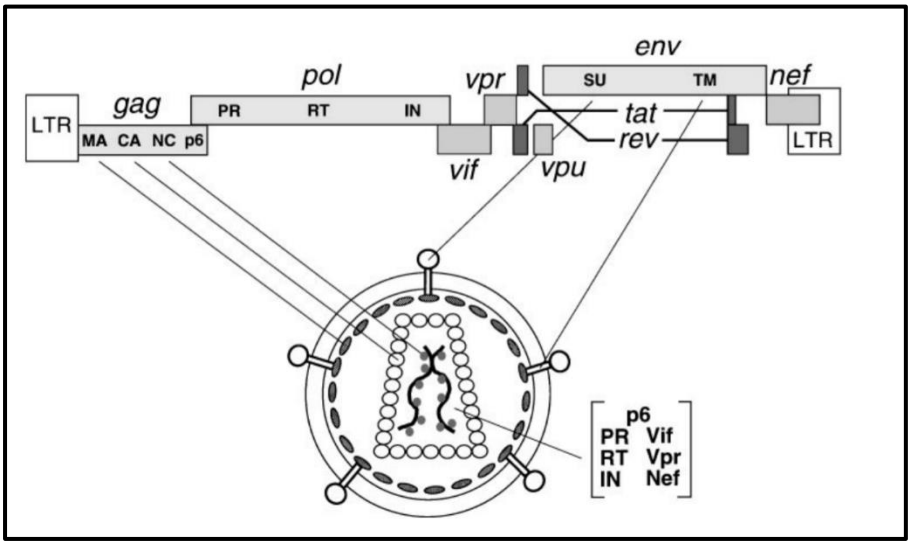

Figure 2 Organisation of HIV-1 Genome and Virion
(Reproduced with permission copyright © from Frankel, Alan D., and John AT Young 1998)

1.2 HIV-1 Transcriptional Unit

HIV-1 promoter long terminal repeat (LTR) is divided into U3, R and U5 regions. U3 is the viral enhancer region having binding sites for many cellular transcriptional factors like – AP-1, USF, NF-κB, NFAT, Sp1, etc. [3]. HIV-1 transcription initiation begins when these transcription factors bind together at the LTR along with RNAPII and transcribes early HIV-1 transcript, forming early viral proteins – Tat, Nef, and Rev. Like eukaryotic transcription, HIV-1 transcription is also abortive, due to binding of negative factors of transcription (DSIF, NELF) to RNAPII causing proximal promoter pausing [4]. Viral protein Tat (transactivator of

transcription) removes this pause by binding to eukaryotic transcription elongation factor P-TEFb and recruiting it to the LTR; this enhances the rate of viral transcription by several folds. Tat is a unique transactivator protein because it binds to early HIV-1 RNA element called TAR (Trans-activation response element) thus recruiting many cellular factors at viral promoter for elongation of viral transcripts. P-TEFb is a complex of cyclin-dependent kinase 9 (CDK-9) and cyclin-T1, this is the only known cyclin to form interaction with N-terminal cysteine-rich domain in Tat. During transcription initiation, P-TEFb first inactivates negative factors of transcription by CDK-9 mediated phosphorylation of NELF by dissociating it from RNA polymerase II, and phosphorylation of DSIF modifies it to a positive factor of transcription. Also, CDK-9 phosphorylates Ser-2 of each heptad repeat in RNAPII α-CTD tail, thus leading to productive transcription [5]. P-TEFb is present in an inactive state on 7SK snRNP complex, bound to HEXIM1/2 by CDK-9. Tat competes with HEXIM1/2 and sometimes with 7SK snRNP to release P-TEFb since Tat has a ten times higher affinity for CDK-9 than HEXIM1 [6]. Tat has to contend with other cellular proteins like Brd4 which binds to the majority of active P-TEFb. P-TEFb also recruits super elongation machinery to HIV-1 promoter, resulting in robust elongation of viral transcripts (Figure 3) [6].

Tat post-translational modifications also have functional significance in HIV-1 transcription. Acetylation in its cysteine and arginine-rich regions is essential for the recruitment of host chromatin modeling complexes to the promoter [7]. In the first step of transcription, PCAF acetylates Tat K28 residue in the cysteine-rich domain, enabling Tat to bind to P-TEFb. During elongation step p300 further acetylates K50 residue in arginine-rich domain [8], this modification allows Tat to dissociate from TAR and P-TEFb and binds to RNAPII. K50 acetylated Tat recruits chromatin modeling complexes like – PCAF, SWI/SNF at RNAPII to promote transcription elongation by nucleosomal assembly and disassembly [9]. After transcription termination, Tat K50 residue gets deacetylated by Sirtuin-1 (SIRT1) so that Tat is available for the next round of transcription (Figure 4) [10].

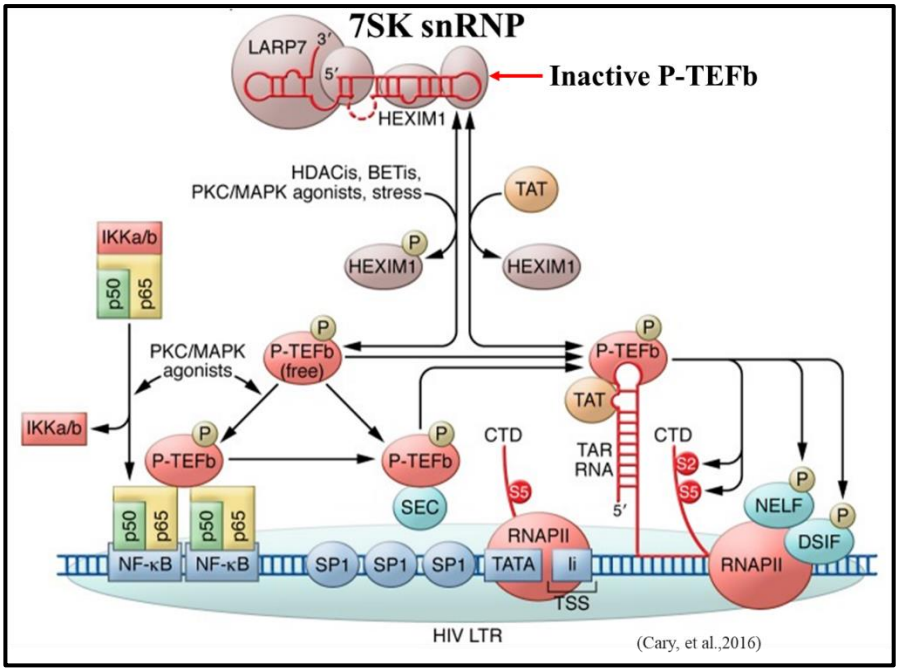

Figure 3 HIV-1 Trans-activation (Reproduced with permission copyright © from Cary, Daniele C., Koh Fujinaga, and B. Matija Peterlin, 2016)

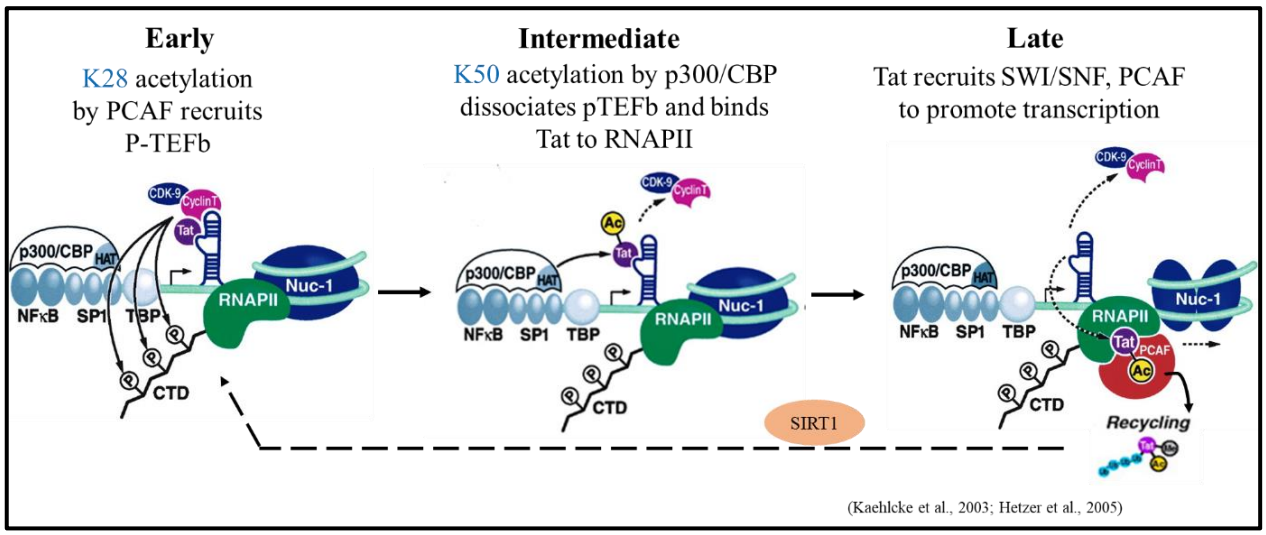

Figure 4 HIV-1 Transactivation in Context of Tat Posttranslational Modifications (Reproduced with permission copyright © from Hetzer, Claudia, et al. 2005; Kaehlcke et al., 2003)

1.3 HIV-1 Latency

Followed by active transcription, HIV-1 gradually enters a state of transcriptional silence, known as latency. It was assumed previously that the state of infected host cell - active or at rest decides the viral state. However, soon it was realized that virus could enter latency

irrespective of the cellular state; observing the latent form of the virus in an activated primary cell model of HIV-1 [11]. HIV-1 infects active CD4+ T-cells, many of which die immediately due to cytopathic effects, and few surviving cells enter to resting state harboring the latent virus. This latent reservoir is a stable niche for the virus, surviving for several months, and further expanding through homeostatic proliferation [12]. Antiretroviral drugs can only suppress new infections in active T-cells, persistence of multiple reservoirs of the replication-competent virus poses a challenge for viral eradication (Figure 5).

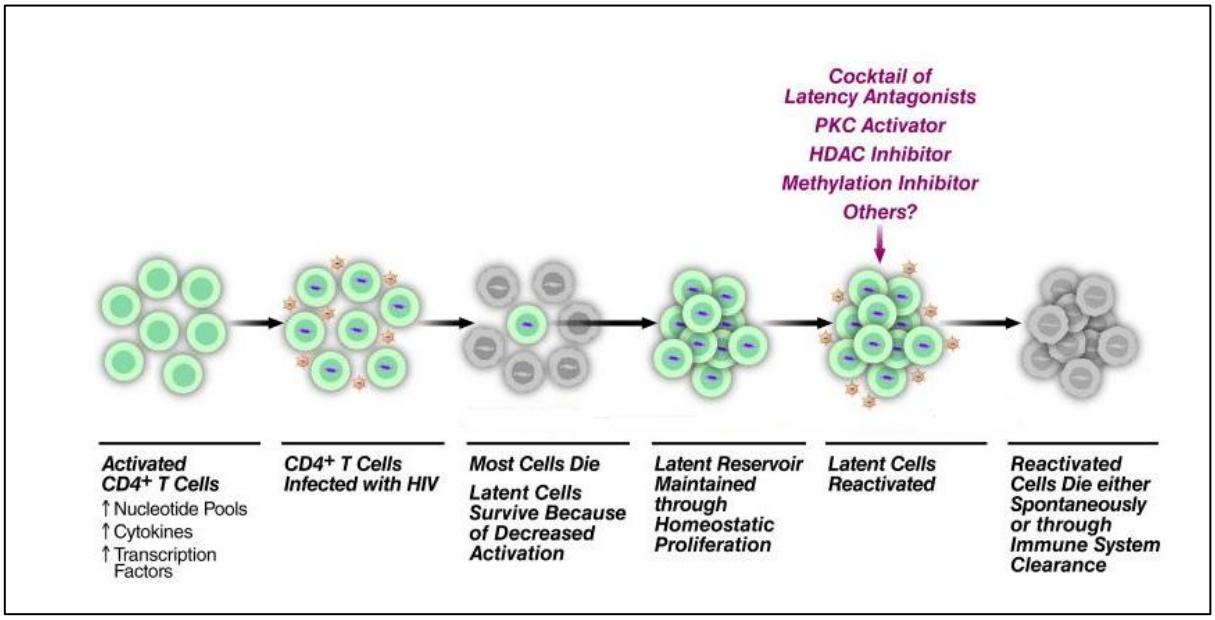

Figure 5 HIV-1 Latent Reservoir (Reproduced with permission copyright © from Ruelas, Debbie S., and Warner C. Greene, 2013)

The mechanism of latency is still not understood. However the involvement of complex host factors for maintenance of latency has been extensively studied, a few are described below.

1. Sequestration of host factors

Latent HIV-1 gets activated by various T-cell signaling pathway, since both require the involvement of common transcriptional factors like – NF-κB, NFAT, and AP-1, with some DNA binding proteins which recruits transcriptional factors to the promoter such as – Sp1, LEF-1, USF, etc. In resting cells all these factors are sequestered by their cognate cytoplasmic inhibitors, thus keeping HIV-1 into latency [13]. Also as previously mentioned, limited availability of P-TEFb either by its inactive form on 7SK snRNP complex or its binding to cellular Brd4 keeps HIV-1 into latency (Figure 6).

2. Epigenetic silencing

Per the histone code hypothesis, combined modifications at histone tails determine the state of gene expression [14] and are also known to be valid with HIV-1 provirus. HIV-1 proviral DNA is organized with histone octamer in the host genome; two important LTR bound histones - NUC-1 and NUC-0 have overlapping binding sites with transcription factors, and their remodeling is required for initiation of transcription [13]. Modifications of NUC-1 by SWI/SNF remodeling complex is crucial for the elongation of viral transcription. In some of the HIV-1 latency models NUC-1 and NUC-0 are constitutively deacetylated, some reports have shown HDACs are recruited at the LTR by distinct binding to NF-kB p-50 homodimer, LSF, etc. Latent virus can be activated by HDACi like trichostatin and valproic acid, owing to the role of HDACs in latency maintenance [13].

Epigenetic modifications of nucleotides have been known to play a role in latency, like chronically latent cell line ACH-2 maintains latency by CpG island methylation at HIV-1 LTR [15]. Facultative heterochromatin protein Polycomb Repressor 2 (PCR2) mediates gene silencing by di- and tri – methylation of K27 in H3 through its enzymatic subunits EZH1 and EZH2, significantly contribute to HIV-1 latency [16] (Figure 6).

3. Transcriptional interference (TI)

TI is the cis suppression of one transcriptional process by another. TI is an asymmetric process happens when promoters are (i) convergent, (ii) in tandem or (iii) divergent to each other [17]. HIV-1 DNA often gets integrated into transcriptionally active regions; its transcription can be inhibited by an ongoing transcription from host promoter, preventing pre-initiation complex from assembling onto the LTR. Transcriptional interference from 5'LTR can inhibit transcription from 3'LTR [18]. Eukaryotic transcriptional machinery removes histones ahead of RNAPII and assembles deacetylated form them behind it; this process ensures non-functional TSS of the cryptic promoter within the gene. The same proof-reading mechanism renders HIV-1 promoter cryptic, leading to TI mediated latency (Figure 6).

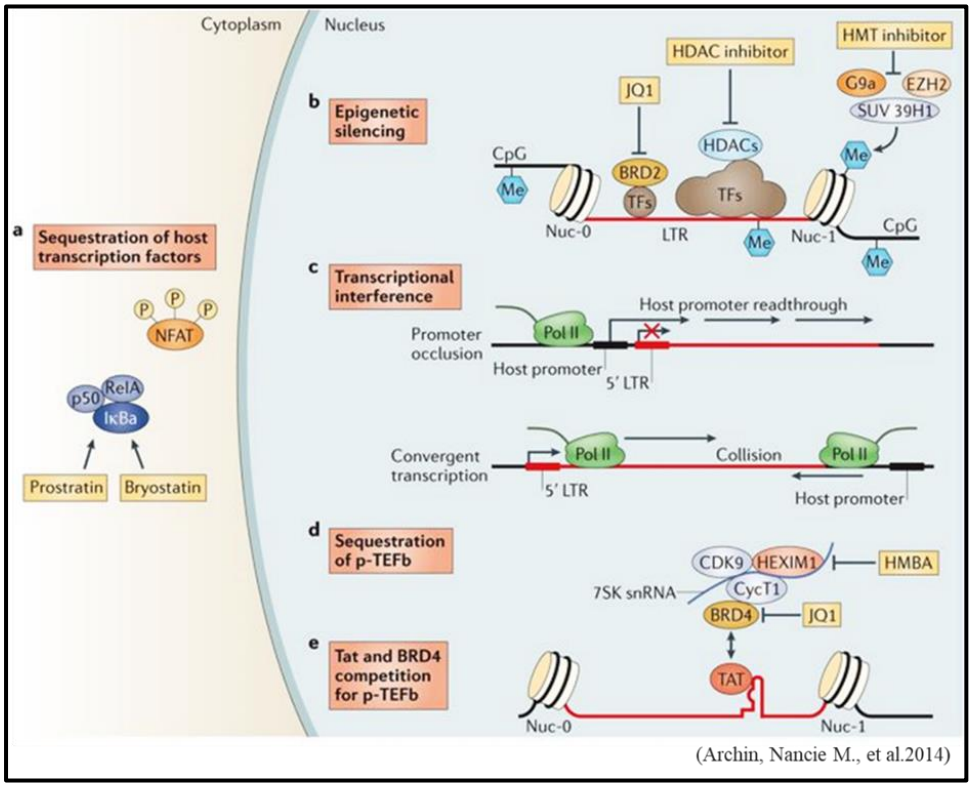

Figure 6 HIV-1 Latency Establishment (Reproduced with permission copyright © from Archin, Nancie., et al. 2014)

4. Role of the microprocessor machinery

Association of the microprocessor machinery with TAR element, therefore with HIV-1 LTR, has been shown to mediate RNAi-independent suppression of HIV-1 transcription. The microprocessor complexes Drosha/Dgcr8, Xrn2 and Rrp6 along with termination factors such as Setx and Xrn-2 acts on RNAPII and causes pausing and premature termination of HIV-1 transcription. Ribonuclease activity of these factors cleaves TAR RNA, and Rrp6 generates TAR derived small RNA which further inhibits viral transcription. The same study has shown knockdown of Drosha and Dgcr8 reduces repressive chromatin marks (H3K9me3 & H3K27me3) from HIV-1 LTR and increases initiation markers such as H3K36me3 for HIV-1 transcription, stating the secondary role of microprocessor machinery in latency establishment (Figure 7) [19].

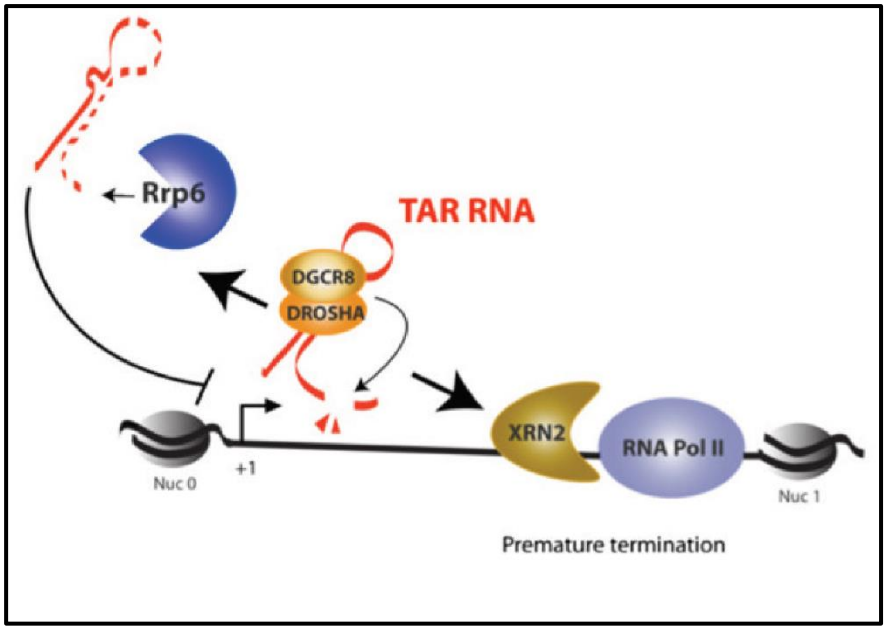

Figure 7 Microprocessor Mediated Premature Termination and Transcriptional Silencing of HIV-1 (Reproduced with permission copyright © from Wagschal, Alexandre, et al., 2012)

5. FACT protein complex

In eukaryotes, facilitates chromatin transcription (FACT) complex-mediated nucleosomal disassembly and reassembly during the process of transcription elongation. However, some of its components SUPT16H and SSRP1 a heterodimer inhibits HIV-1 transcription by binding to Tat and sterically hindering binding of P-TEFb at the TAR element. Tat has been shown to bind specifically SUPT16H, without disrupting the heterodimer. Depletion of FACT components in latently infected primary CD4+ T-cells reversed latency, small molecules disrupting the interaction of FACT proteins with Tat-LTR can be considered for antiretroviral therapy (Figure 8) [20].

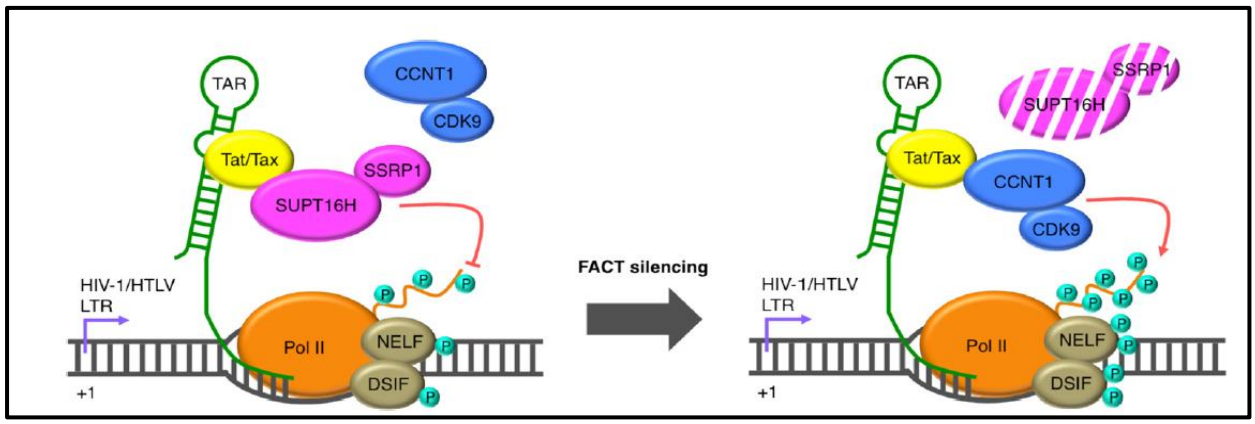

Figure 8 Working Model of Tat mediated Transcriptional Silencing (Reproduced with permission copyright © from Huang, Huachao, et al, 2015)

1.4 Tat - A Multifunctional HIV-1 Protein

Tat is an intrinsically disordered protein; this enables Tat to bind to the repertoire of host proteins and modulates expression of host genes. Tat is divided into six domains; first five are exon1 encoded and the second exon encodes C-terminal domain (Figure 9). Exon1 encodes N-terminus of Tat, containing domains required for Tat mediated trans-activation and for cellular interactions, they are - RNA binding domain, nuclear localization signal, and protein transduction domain [21]. C-terminal Tat domain is believed to have a role in viral pathogenesis, by reducing MHC I complex expression on cell surface, thus escaping immune system [22].

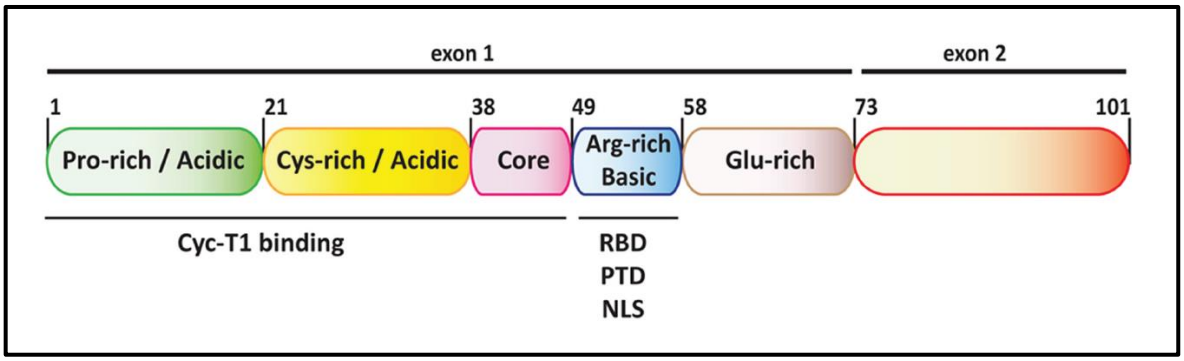

Figure 9 HIV-1 Tat Protein Domains (Reproduced with permission copyright © from Clark, E., Nava, B., & Caputi, M. (2017)

As mentioned above Tat can be modified at various residues as showed in figure 10 by host proteins [21].

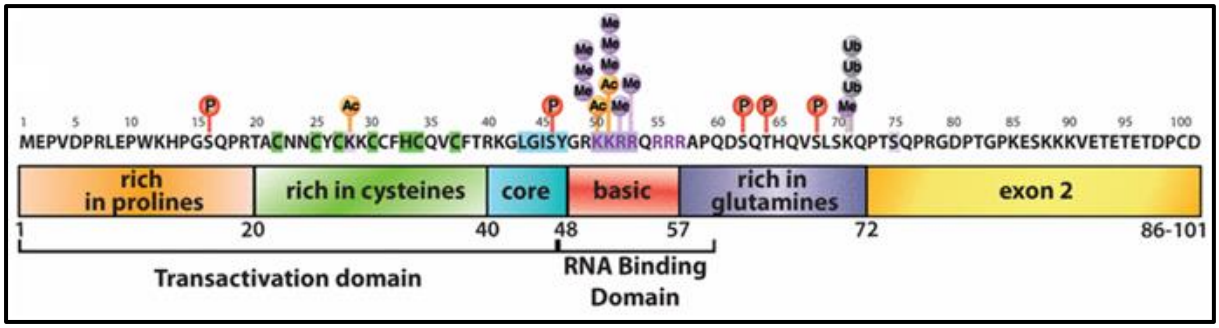

Figure 10 Post-translational Modifications of Tat (Reproduced with permission copyright © from Hetzer, Claudia, et al. 2005)

Tat modulates many cellular processes; some are mentioned below.

1. Tat-induced signaling pathways

   (i) Apoptosis: Upon HIV-1 infection Tat mediates up or down-regulation of many apoptotic genes leading to rapid turnover of T-cells. Tat regulates expression of apoptosis-associated genes like - cytokine, cell survival genes – BCL-2, SOD (superoxide dismutase). It activates mitochondrial apoptosis by disrupting microtubule networks or by direct signal transduction causing TNF-α or Fas-Fas ligand induced apoptosis. Tat upregulates expression of PTEN and PP2A, which increases the unphosphorylated form of FOXO3a leading to T-cell apoptosis [21]. Tat can cause apoptosis in bystander T-cells by increased production of TNF-α related apoptotic inducing ligand (TRAIL) in macrophages [23].

   (ii) Cell proliferation: Kaposi sarcoma lesions are clinical manifestation during HIV-1 infection. Tat and FGF (fibroblast growth factor) interplay lead to the formation of lesions in KS. Further, Tat enhances KS cell survival by inducing PI-3K pathway dependent cell survival factors like- VEGF2, insulin growth factor receptor I (IGF-IR) [23].

2. Extracellular Tat

   Tat is secreted from infected cells by binding to PI (4,5-P_2) in plasma membrane followed by insertion in the membrane by its Trp11 residue. The extracellular Tat which is in nanomolar range can bind to various target cells through receptors like – CD26, LRP, CXCR4 and heparin sulfate proteoglycan present on lymphocytes, macrophages, neutrophils, etc. Once internalized by endosome mediated pathway, Tat can directly or indirectly modulate host gene expression (Figure 11) [24].

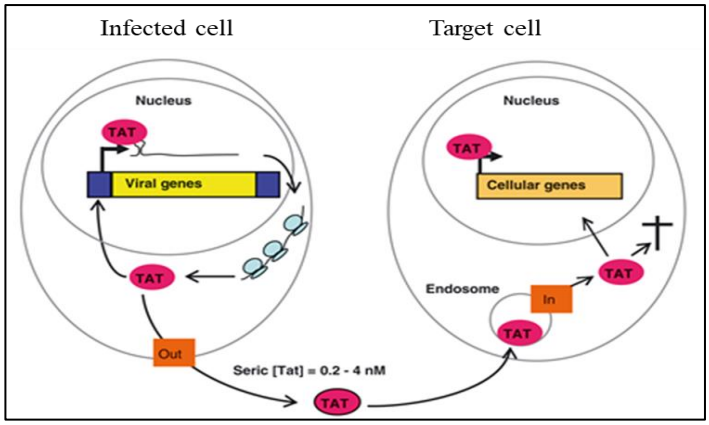

Figure 11 Tat Secretion and Internalization
(Reproduced with permission copyright © from Debaisieux, Solène, et al. 2012)

3. Regulation of cellular gene expression

Tat can modulate the expression of cellular genes like cytokines, cell cycle-related proteins, surface, and chemokines receptors, mRNA processing genes to enhance viral proliferation. Following are the potential mechanisms for modulation of cellular gene expression – (i) binding to TAR-like element in 5' untranslated region, (ii) direct binding to promoter sequences, (iii) associating with transcriptional factors. Tat binds to 5'UTR of IL-6 and TNF-β promoter, IL-6 is linked to pathologies associated with AIDS like B-cell lymphoma, Kaposi sarcoma, etc. Tat indirectly induces the expression of interferon-stimulated genes by JAK and MAPK pathway by upregulating expression of MAP2K6, MAP2K3, and IRF7. Also, recent studies have shown the interaction of Tat with T-cell master transcription factor ETS1, along with RUNX1 and GATA3 modulating expression of nearly 400 cellular genes. Thus, promoting T-cell activation as well as viral activation [21].

Recently, a genome-wide mapping study using ChIP was performed in Jurkat cells to identify Tat occupied regions in the genome. It was observed that Tat occupies a maximum of DNA repeat elements, of which 58% were Alu repeats (figure 12). Tat is known to upregulate Alu transcription in Jurkat cells by increasing the activity of TFIIIC. In turn Alu increases viral replication by antagonising interferon-induced PKR activation and reducing cellular protein synthesis under stress [25]. Since Tat bound Alu repeats are located in introns and intergenic regions much away from TSS, it was found that Tat can mediate long-range interaction with its target gene (Figure 13).

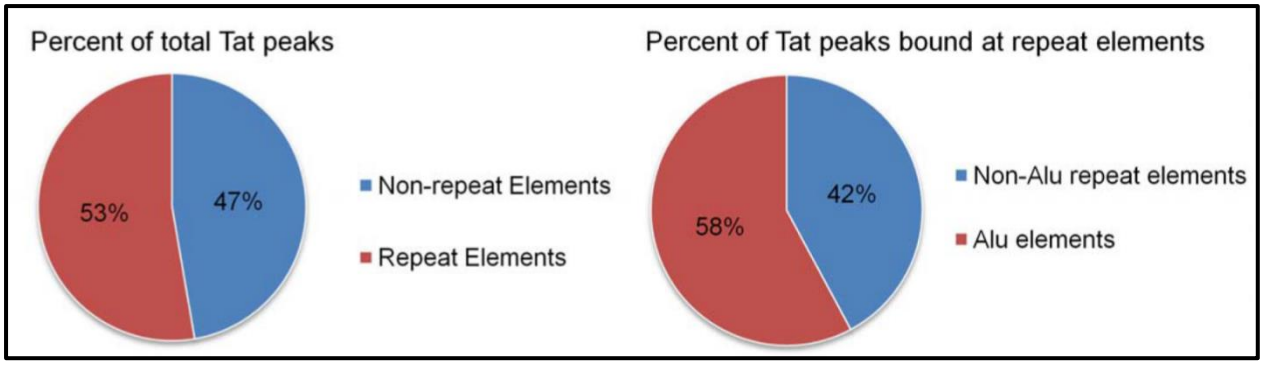

Figure 12 Distribution of Tat-binding Peaks (Reproduced with permission copyright © from Marban, Celine, et al. 2011)

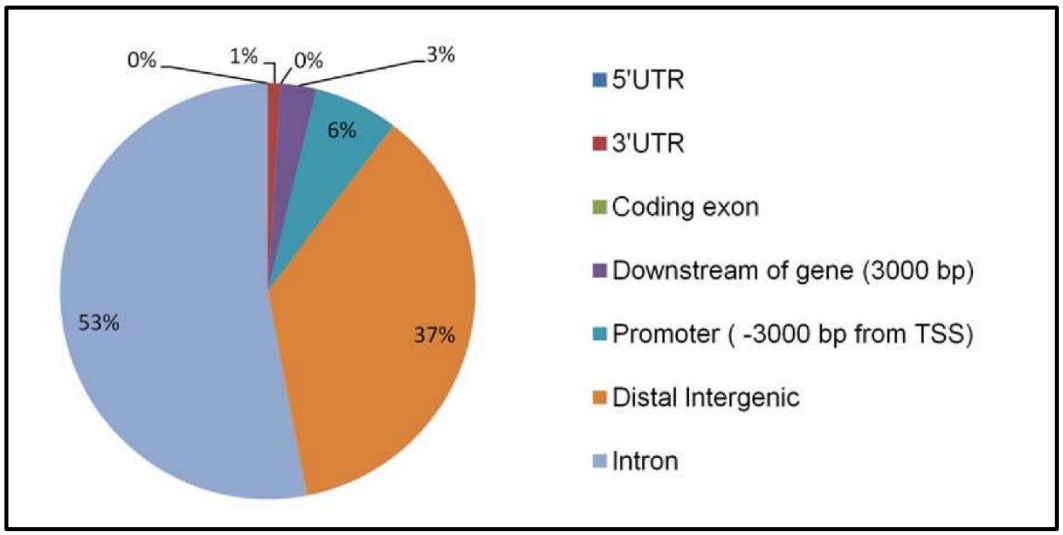

Figure 13 Distribution of Tat-bound Alu (Reproduced with permission copyright © Marban, Celine, et al. 2011)

1.5 Next Generation Sequencing

Before the era of transcriptomics, in late 1970s transcript analysis was done using low-throughput and time-consuming Sanger sequencing and ESTs (expressed sequence tags) libraries. Later tags based high-throughput techniques like SAGE (serial analysis of gene expression), CAGE (cap analysis of gene expression) and MPSS (massively parallel signature sequencing) were developed, these were costly, and only small proportions of transcripts were analysed. In 1995, Microarray was used for transcriptome analysis; this technique quantifies the abundance of transcripts by their hybridization to complementary probes, this approach was cost-effective but was inefficient for the discovery of novel transcript since it requires prior knowledge of transcripts. Introduction of Next Generation Sequencing in the 2000s has

revolutionized the field of transcriptomics. In 2004, the first RNA-seq report was published covering 10^5 transcripts and in 2008 Solexa/Illumina technologies covered 10^9 transcripts, which is sufficient for human transcript analysis [26]. In NGS workflow (Figure 14), isolated RNA is randomly fragmented and converted to cDNA libraries, either by rRNA depletion or by mRNA enrichment by polyA. These libraries are tagged with a unique adapter and sequenced by sequence-by-synthesis approach using NGS platform like – Illumina, Roche 454, PacBio, etc. All the NGS systems are equipped with pair-end reads sequencing, which provides more accurate read alignment to the reference genome as compared to single-end read (Figure 15).

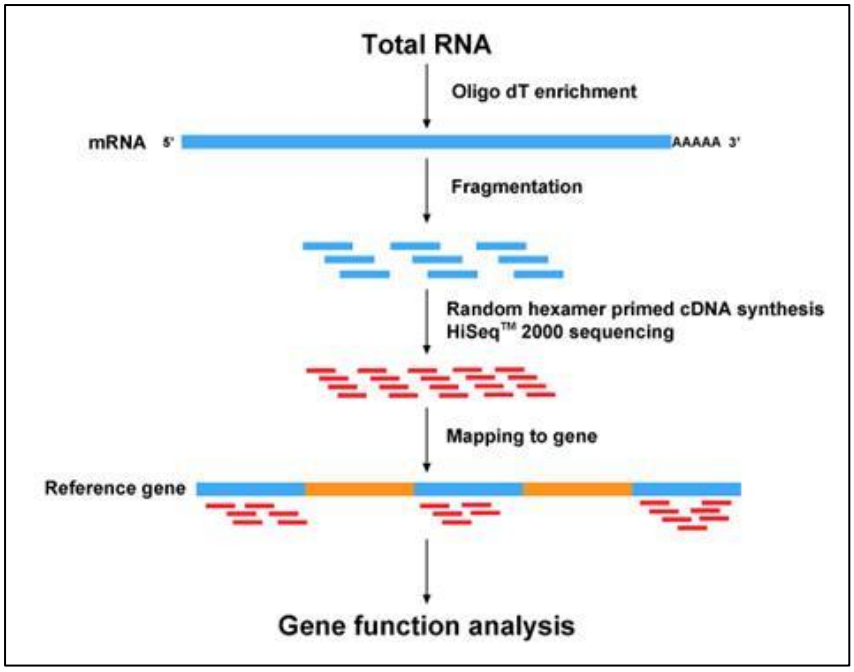

Figure 14 RNA Sequencing Workflow (Reproduced with permission copyright © from bgisequence.com)

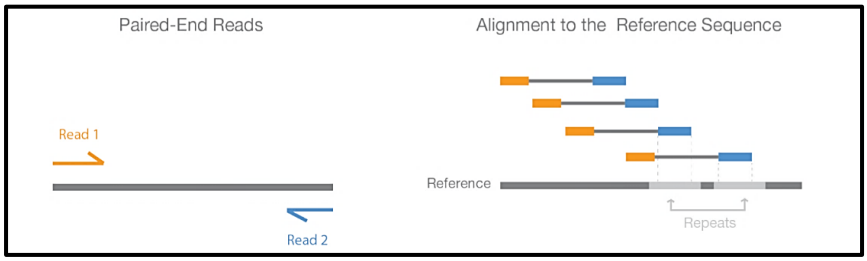

Figure 15 Paired-end read Sequencing and Alignment (Reproduced with permission copyright © from Illumina)

In post-infection scenario HIV-1 active state is achieved by its master regulatory circuit, where Tat mediates elongation of viral transcripts from its promoter and forms positive-feedback loop at the LTR. If Tat is continuously available to the virus, then mechanism through which latent

state is achieved is not understood. Based on previous studied from our laboratory (Chakraborty et al. unpublished) showing Tat occupancy at viral promoter in latent cell population and faster latency establishment by strong promoters (more Tat), we speculate the existence of Tat-negative feedback loop, which drives HIV-1 into latency. To understand effect of Tat mediated latency, we engineered a novel HIV-1 Jurkat reporter cell line encoding the expression of two different fluorescent proteins that permit the delineation of several distinct phases through the latency cycle of HIV-1. The cells representing the distinct latency phases have been sorted, cellular RNA extracted, and the RNA was subjected to whole transcriptome analysis to evaluate the influence of the viral transcription circuit on the expression of the cellular transcripts.

Our work is novel in terms, we have considered whole latency cycle (Figure16), to understand the Tat mediated cellular gene expression pattern which might mediate latency establishment by Tat post-translational modifications or by recruiting negative factors of transcription specifically on HIV-1 LTR. We hypothesise that Tat a has diametrically opposite effect on HIV-1 transcription – keeping HIV into active as well as driving it into latency (figure 17).

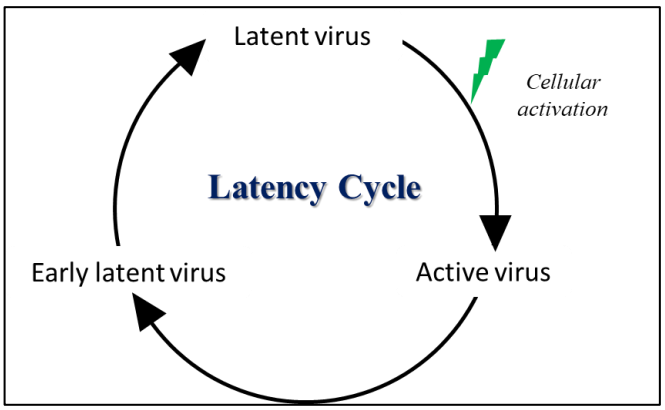

Figure 17 Schematic for HIV-1 Latency Cycle

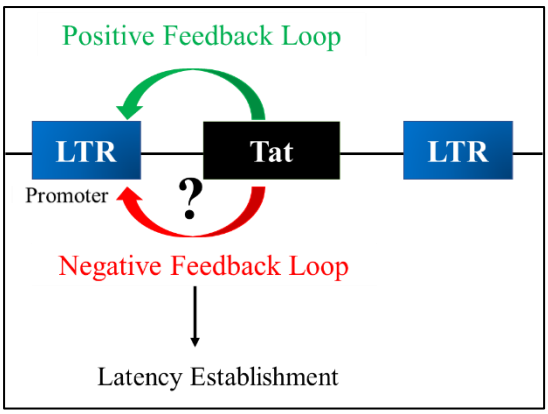

Figure 16 Hypothetical Model Explaining Diametrically Opposite Roles of Tat

Chapter 2

Material and Methods

Chapter 2: Material and Methods

2.1 Viral Construct

The viral construct used for the study was a dual reporter sub-genomic subtype C virus, having only HIV-1 master regulatory circuit, i.e. LTR and Tat protein. In the construct d2EGFP is directly under the LTR, followed by IRES separated Tat-mScarlet fused protein. D2EGFP protein has a degradation domain from mouse ornithine decarboxylase, reducing its half-life to 2hrs, this would give real transactivation state of the virus, unlike other conventional EGFP which remain in the system for a longer time even after promoter is switched off. Only RFP expression implies the presence of Tat in the system. Table 2 explains the working of the sub-genomic virus (Figure 18). To normalize for the environmental variations of the Jurkat cells, irrespective of viral infection we have time-point control Jurkat cells, which were infected with the lentiviral construct, EF-1α_EGFP expressing EGFP under EF-1α constitutive promoter (Figure 19). Table 1 explains working of HIV-1C construct

Figure 18 Sub-genomic Viral Construct

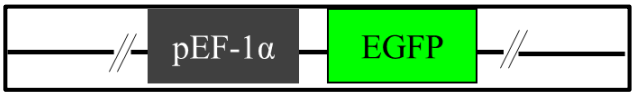

Figure 19 Time-point Control Lentiviral Construct

State of LTR (on/off)	Reporter protein expression
on	D2EGFP & RFP
off	RFP (only Tat is present)
off	No expression of either

Table 1 LTR Activity and Expression of Reporter Proteins

2.2 Immunoblotting to check the integrity of Tat-RFP fusion protein

Since Tat is fused to huge RFP (mScarlett), we wanted to check the integrity of Tat-FLAG-RFP fusion protein. Immunoblotting was performed in HEK 293T cells, 5million cells were transfected with plasmid encoding fusion protein (Tat-FLAG-RFP); only Tat, mScarlet and 3X FLAG (fused p50) as positive controls; plain cells were used as negative controls. The lysate was prepared 48 hrs post-transfection, cells were washed with ice-cold 1XPBS, centrifuged at 800 rpm for 5 min, at 4 ^{o}C. Cells were lysed in ice-cold RIPA buffer added with 1X PIC, kept on vortex for 30min at 4 ^{o}C, followed by centrifugation at 16,000 g for 30 min at 4 ^{o}C. For immunoblotting 12% SDS gel was casted, protein samples (100 µg) were prepared in 4X SDS loading dye, heated at 95 ^{o}C for 10 min. The transfer was done on PVDF membrane, for 1.5 hrs at 90 V. The membrane was blocked in 5% skim milk and probed with specific primary antibody (prepared in 1XPBS, 0.2% BSA) o/v at 4 ^{o}C. Blots were then probed with specific secondary antibody conjugated with HRP (prepared in blocking solution, 1:10,000 dilution) for 1 hr at RT and developed. Vinculin was used as loading control.

Primary antibody	Company	Dilution	Host
dsRed	Clontech (632496)	1:1000	Rabbit
Anti- Tat	NIH (5A5.3)	1:1000	Mouse
Anti -FLAG	Sigma (F3165)	1:2000	Mouse
Anti-vinculin	Sigma (V4505)	1:2000	Mouse

Table 2 List of antibodies for Western Blotting

2.3 Cell culture, virus preparation and infection

Both the virus – sub-genomic HIV-1C and EF-1α_GFP were prepared in HEK 293T cells, using packaging plasmids pVSV-G, pCMV-Rev, and psPAX. For transfection 0.3million HEK 293T cells were seeded in 6-well plate in DMEM media supplemented with 10% FCS, maintained at 37^{o}C, 5% CO_2; cells were allowed to grow until they were 40%-50% confluent (1.5 days). 1 hr before transfection cells were given media change, DNA mix (1/10th of total media) was prepared by adding packaging vectors along with either of lentiviral plasmids. To DNA mix, 2X HBS and 2.5M $CaCl_2$ (1/20th of DNA mix) were added and mixed vigorously,

followed by dropwise adding to the well. 6 hrs post-transfection media was changed to remove residual plasmids and after 24hrs imaging was done by epifluorescence microscopy to check the expression of GFP and RFP. Post 48hrs of transfection virus was harvested and stored in -80°C.

Jurkat cell line was used for infections and Next Generation sequencing analysis. 0.6million Jurkat cells were infected at 10% infectivity in 10µg/ml DEAE containing media (RPMI with 10% FCS at 37°C, 5% CO_2). 6hrs post infection media was changed to remove residual virus, 48hrs post infection cells were activated with cellular activators (TNF-α (10ng/ml), PMA (5ng/ml), HMBA (5nM) and TSA (0.1µM)) and 10% infected cell population was sorted; for HIV-1C infected cells double positive cells were sorted, and for time-point control cells GFP positive cells were sorted. Latency kinetics was determined from latency kinetics.

2.4 Fluorescently activated cell sorting for NGS

To understand the mechanism of latency by NGS, we selected five distinct cell populations resembling different phases of viral infection from latency kinetics plots. Populations sorted for HIV-1C infected Jurkat cells were S1, S2, S3, S4 and S5, and its corresponding time-point control cells were T1, T2, T3 and T5; it is important to note that S3 and S4 cell populations were sorted on the same day, hence have only one corresponding time-point control T3 (Figure 21).

2.5 RNA Isolation, Quantification and Quality Check

Monophasic lysis method was used for isolating total cellular RNA from all time point sorted Jurkat cells. Around 2 million Jurkat cells were pelleted down at 800rpm for 5 min, followed ice-cold PBS wash. Cells were then lysed in 300µl of TRI reagent (Sigma #T9424), the tube was gently inverted 4-5 times and incubated at RT for 5 min. To this 60µl of chloroform was added, mixed vigorously, incubate at room temperature for 5 min, allowed phases to separate, followed by centrifugation at 12000rpm for 15min at 4°C. Then aqueous layer (containing total RNA) was aspirated carefully, to this 75µl of isopropanol was added and incubated at RT for 10min to allow RNA to precipitate and centrifuged at 12000rpm for 10min at 4°C. The supernatant was thrown, and the pellet was washed with 75% alcohol at 7500rpm for 5min, the pellet was allowed to dry at RT and then re-suspended in 50ul of nuclease-free water (Himedia #ML064). RNA purity and quantity were determined by taking A_{260}/A_{280} and A_{260}/A_{230} ratio.

RNA quality and integrity were checked on 1.5% non-denaturing agarose gel by observing bands corresponding to – 28srRNA, 18srRNA, and 5srRNA. Also, QC of RNA samples was done by Agilent 2100 Bioanalyzer System using Agilent Bioanalyzer RNA 6000 nano kit, which provides RNA Integrity Number (RIN) based on its electropherogram profile; good quality RNA has RIN between 7-10.

2.6 Quantification of Tat transcripts and Tat-RFP protein levels

Tat transcripts were quantified by qPCR across all time points (S1 – S5) from HIV-1C infected samples in triplicates, with GAPDH as a control. All the RNA samples were given DNase treatment by AccuRT Genomic DNA Removal Kit (abm #G488) following the manufacture's protocol. cDNA was prepared from 500ng of total RNA using OneScript cDNA synthesis kit (abm #G234) following the manufacturer's protocol. qPCR for Tat transcripts was done with the c-Tat specific probe using MyTaq mix (Bioline # BIO-25041) and for GAPDH transcripts were quantified by SYBR based chemistry.

Table 3 List of Primers and Probe for qPCR

Tat	Probe	FAM-CATGGCAGGAAGAAGCGGAGACAGCG-BHQ
	Forward primer	ATCATCCAGGAAGTCAGCCCGAAAC
	Reverse primer	TTTGATGATCCTCACTGCTCGGAGG
GAPDH	Forward primer	GAGCTGAACGGGAAGCTCACTG
	Reverse primer	GCTTCACCACCTTCTTGATGTCA

2.7 Next Generation Sequencing

2µg of total RNA from all time-points - infected cells (S1, S2, S3, S4, S5), time-point control cells (T1, T2, T3, and T5) in duplicates and plain Jurkat cells were given for mRNA Sequencing (Illumina, Hiseq 2000). Sequencing was done for paired-end reads (2 X 150bp) and for 25-30M reads. Before proceeding for data analysis, we plotted correlation between the duplicates for each sample.

NGS data analysis workflow involves, quality check of reads by FastQC software, low quality reads from both the ends were trimmed by trimmomatic. Using Picard all the PCR duplicates were removed, and reads were aligned on the reference genome using HISAT2. Cufflinks software assemble transcripts based on reads aligned to reference files, its output file cuffdiff_output can be used for further differential expression of genes across different time-points.

For differential gene expression across different time points and between test (HIV-1C) and control condition (time-point control), we used cuffdiff_output file, which has FPKM values for all the genes. We first compared the differential expression of genes across different time points within the sample (HIV-1C and time-point controls), then compared differentially expressed genes between HIV-1C samples and time-point controls. Details mentioned in (Figure 31).

Chapter 3

Results

Chapter 3: Results

3.1 Western blot to check integrity of Tat-RFP protein in HEK 293T cells

We observed single band of 54kDa corresponding to Tat-RFP fusion protein, when probed with all three antibodies (except for anti-FLAG antibody, which might have non-specificity). Immunoblotting is difficult to perform in infected Jurkat cells due to overall low protein production. Our understanding is that Tat-RFP construct is intact in Jurkat cells also as observed in HEK293T cells (figure 20).

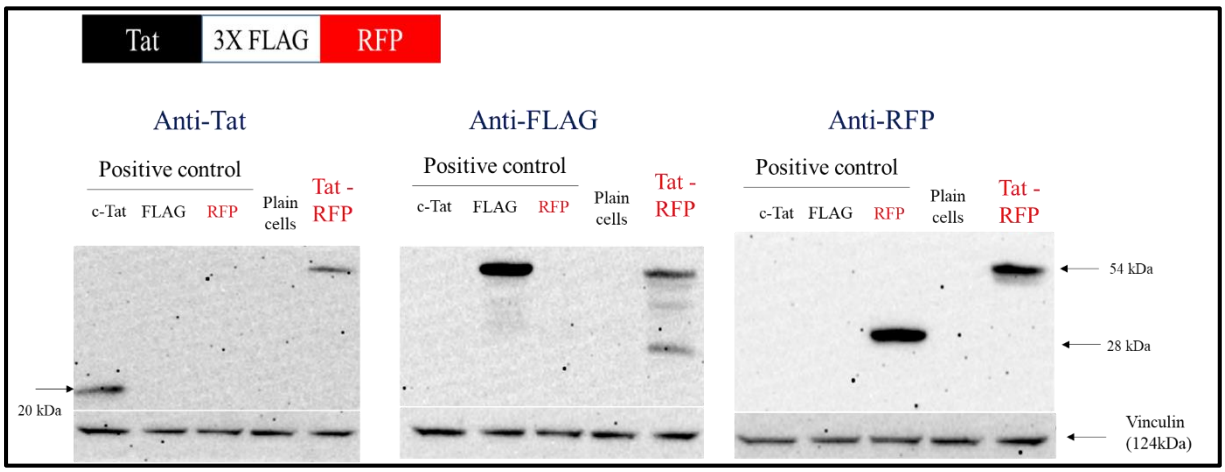

Figure 20 western blot for Tat-RFP fusion Protein (A) Construct of Fusion Protein (B) Blot Probed With all Three Antibody.

3.2 Latency kinetics and Cell Sorting

48hrs post-infection cells were activated with global activators, 10% of the total population were positive for both the reporter proteins – d2EGFP and RFP in HIV-1C infection; and only EGFP positive for time-point control infection (EF-1α_EGFP) (Figure 22). These infected cells were sorted, and latency kinetics was established. Then these sorted cells were allowed to enter into latency for 15 days, followed by activation with global activators, and sorting each of five distinct population (S1, S2, S3, S4 and S5) on their corresponding days. From time-point control infected cells, we sorted only EGFP positives population (T1, T2, T3 and T5) on their corresponding days (Figure 21and 23).

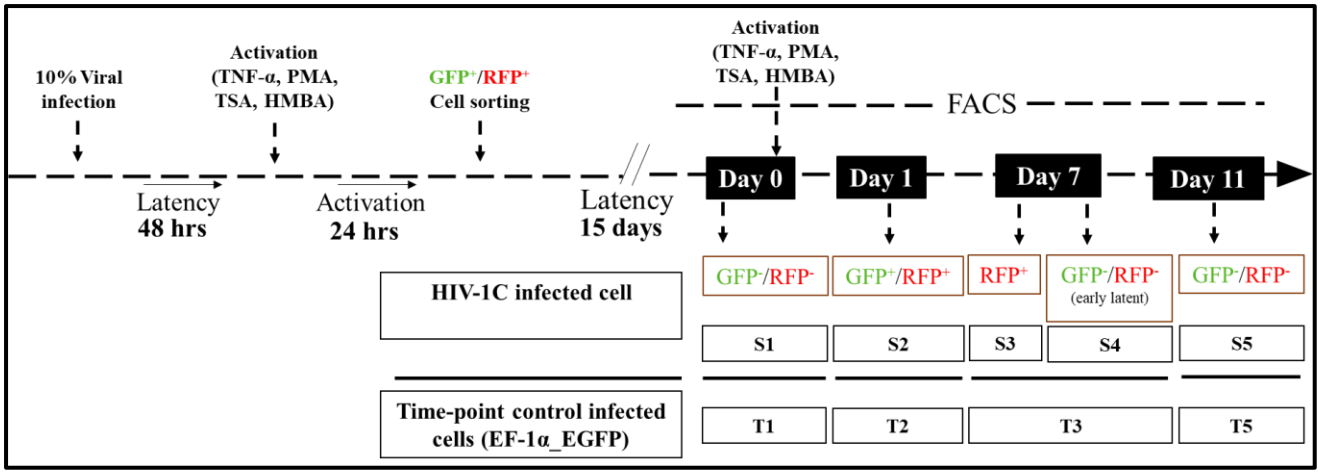

Figure 22 Experimental Paradigm for Sorting Distinct Cell Population on Different Days

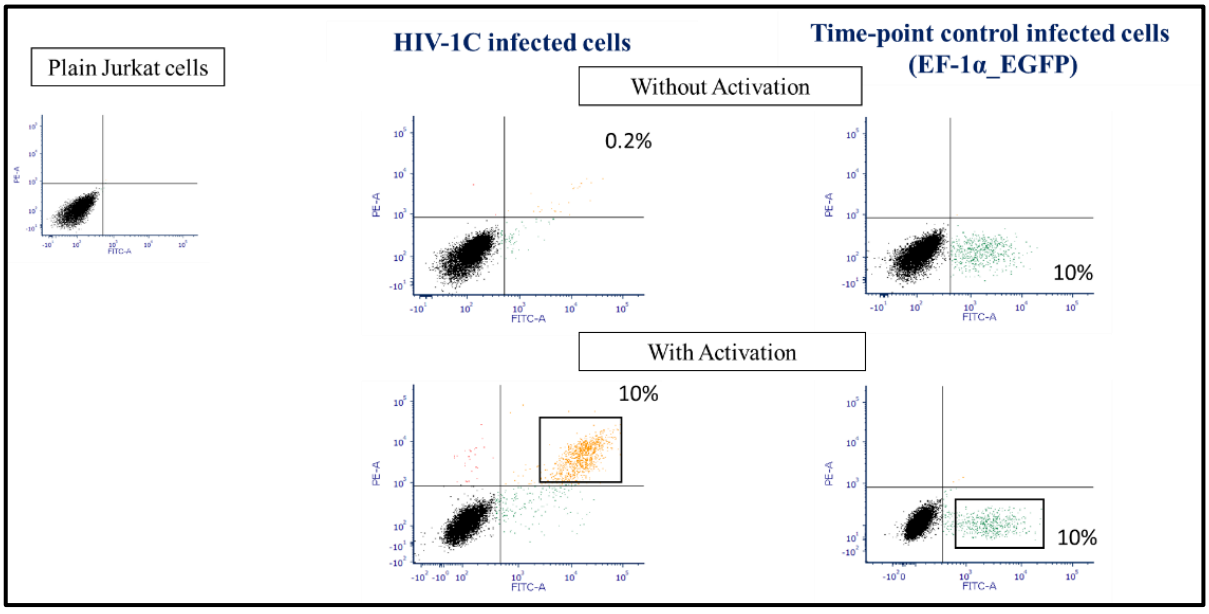

Figure 21. 10% Infectivity for Both HIV-1C and EF-1α_EGFP in Jurkat cells, observed after activation with cellular activators. (EF-1α_EGFP shows constitutive EGFP expression when not activated, right top panel)

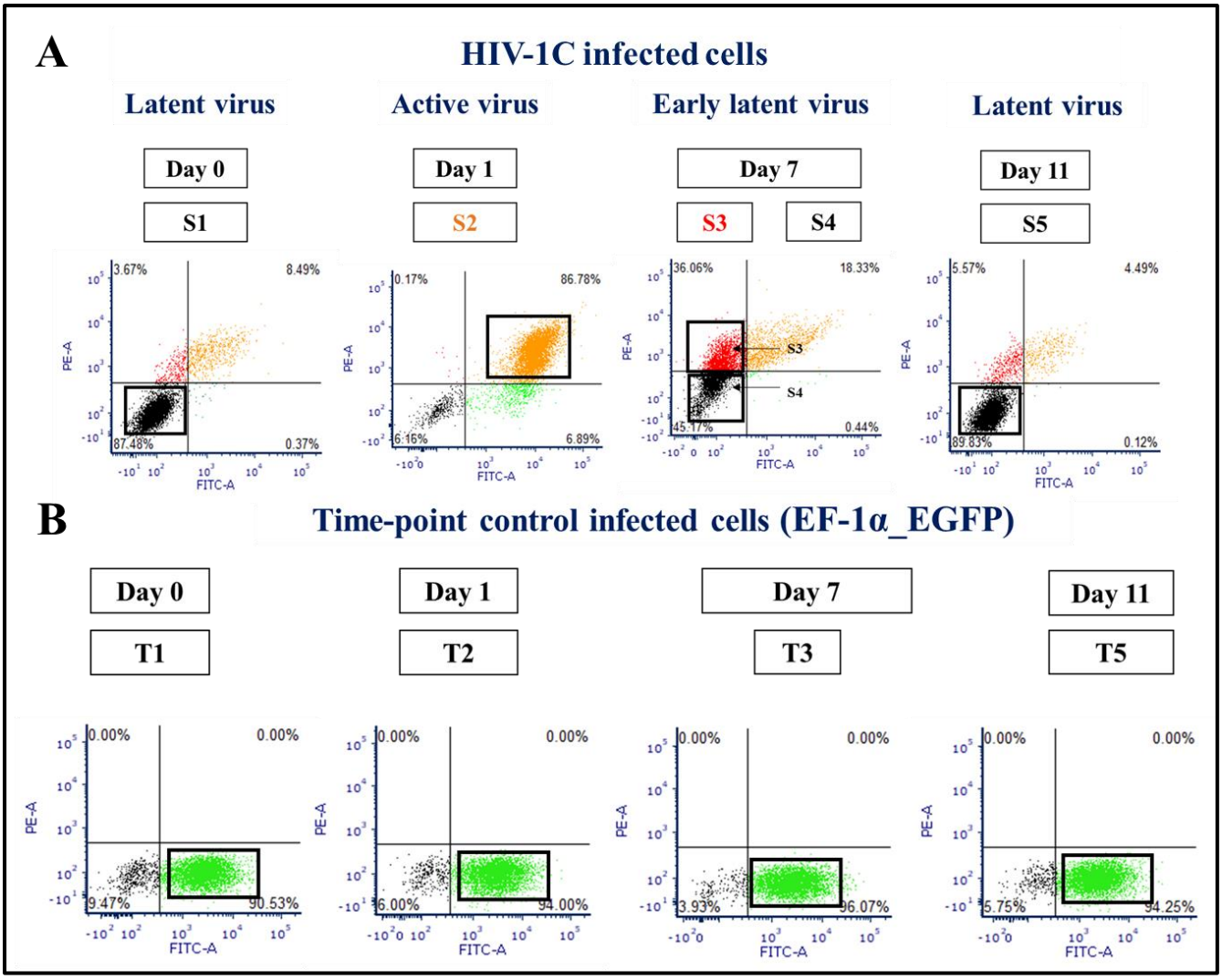

Figure 23 Flow Cytometry Profiles of All the Sorted Population (A) for HIV-1C Infected cells (B) for Time-point Control Cells. (Y-axis: RFP signal & X-axis: GFP signal; solid line box represents sorted population)

3.3 RNA isolation and quality control

Total cellular RNA was isolated by monophasic lysis method as mentioned in section (). RNA purity was checked by A_{260}/A_{280} and A_{260}/A_{230} ratio, all samples had ratio between 1.8-2.0. RNA samples were of good quality, observed as sharp bands of 28s, 18s, and 5s rRNA on non-denaturing agarose gel (1.5%) (Figure 24).

RNA integrity number (RIN) was between 9-9.9 (Table 4) calculated Agilent 2100 Bioanalyzer System, this software generates an electropherogram and gel-like image (figure 25); RIN value is calculated based on ribosomal RNA ratio. Good quality RNA has RIN value above 7.

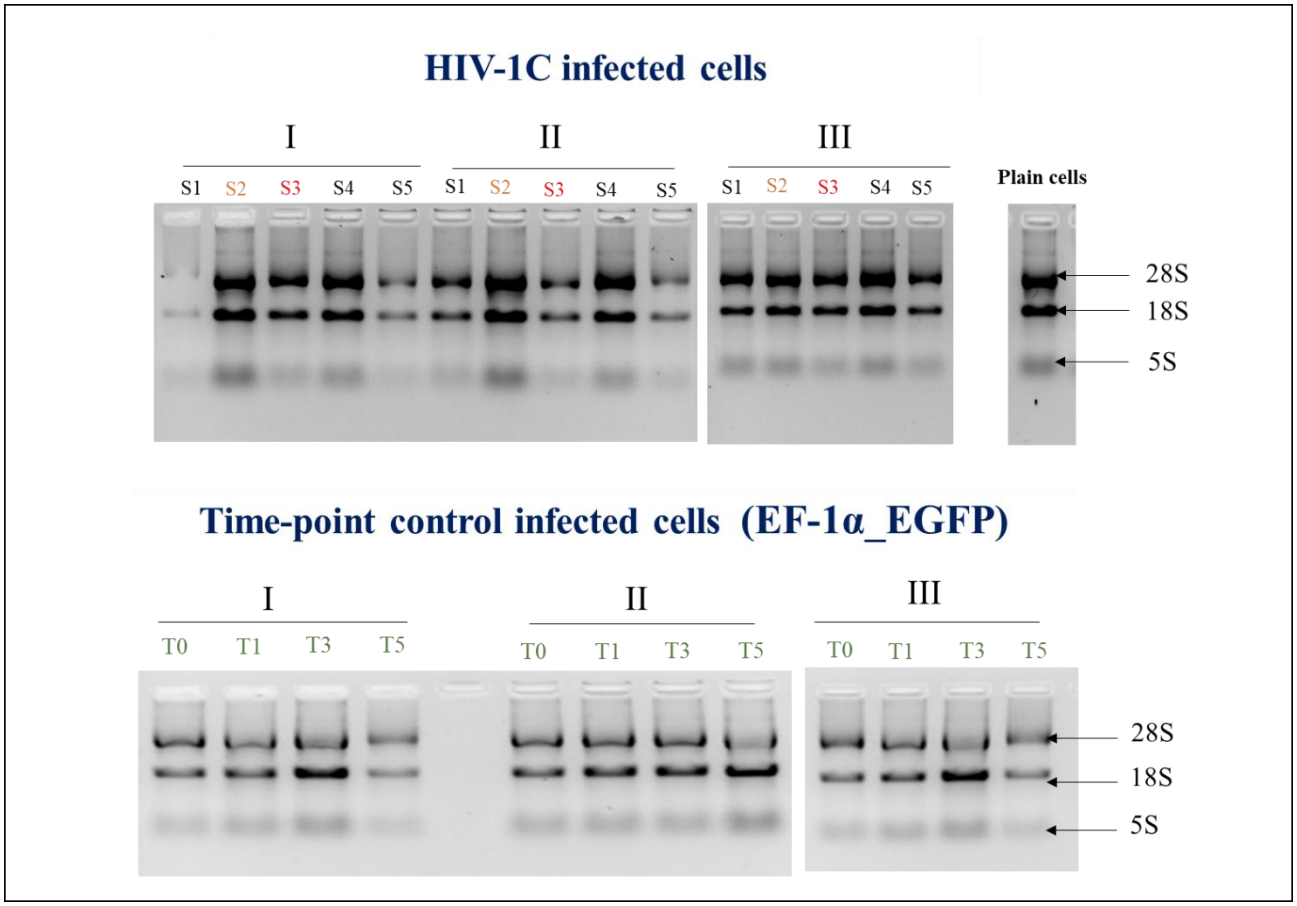

Figure 24 Non-denaturing RNA gel (A) for infected cells (B) for time-point control cells, showing sharp bands for all three types of rRNA.

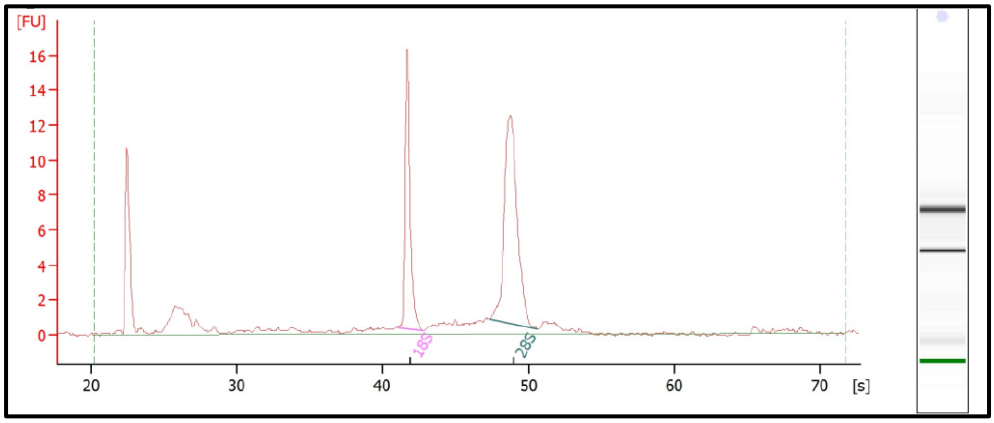

Figure 25 Electropherogram and Gel-like image Generated by Agilent 2100 Bioanalyzer system (representative result for S1.1 sample)

Sr. No.	Sample ID	RIN Value	Status
1	S1.1	9.2	passed
2	S2.1	9.5	passed
3	S3.1	9.7	passed
4	S4.1	9.7	passed
5	S5.1	9.6	passed
6	S1.2	9.7	passed
7	S2.2	9.5	passed
8	S3.2	9.7	passed
9	S4.2	9.7	passed
10	S5.2	9.5	passed
11	T1.1	9.9	passed
12	T2.1	9.9	passed
13	T3.1	9.3	passed
14	T5.1	9.9	passed
15	T1.2	9.2	passed
16	T2.2	9.7	passed
17	T3.2	9.7	passed
18	T5.2	9.9	passed
19	Plain cells	9.7	passed

Table 4 RIN Value for All the Samples Including HIV-1C and Time-point RNA Given for Sequencing (in duplicates)

3.4 Quantification of Tat transcript and Tat-RFP protein

As expected, Tat transcripts were significantly high in S2 population (active) compared with S1 (latent) (figure 26). We also checked Tat-RFP expression across all the time points by taking into consideration median fluorescent intensity (MFI) of RFP signal, as expected MFI was significantly high in S2 quadrant compared with S1. However high MFI in S3 (early latent population) quadrant signifies role of Tat in latency establishment (Figure 27).

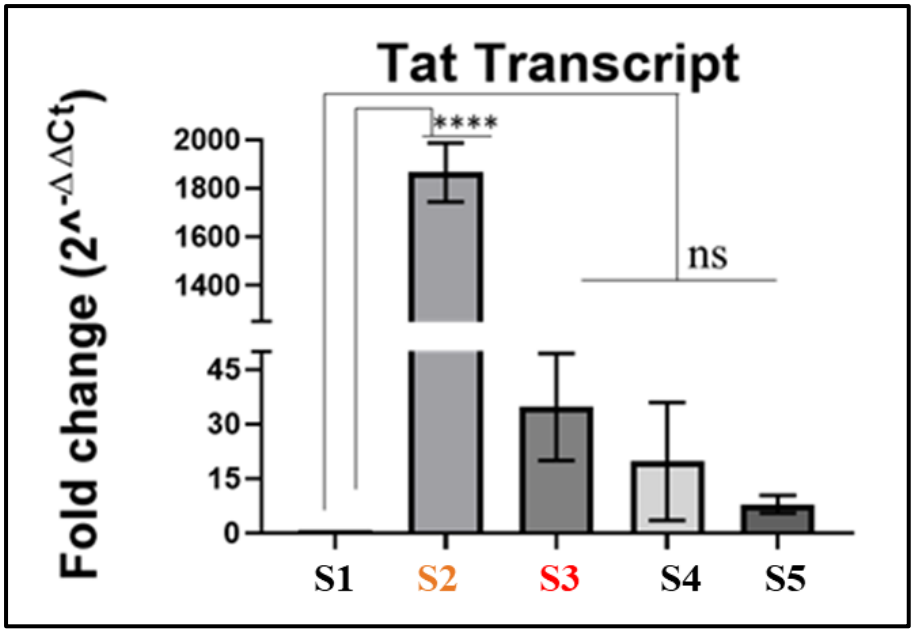

Figure 26 Tat Transcripts Quantification for all HIV-1C infected Jurkat Cells

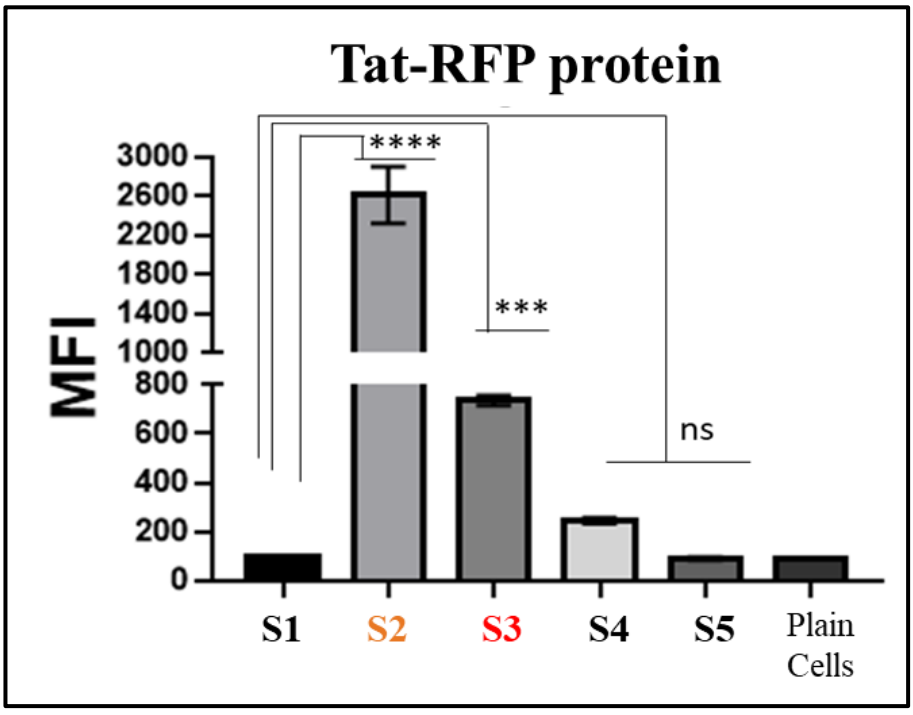

Figure 27 Tat-RFP Protein quantification for all time points from HIV-1C Infected samples

3.5 NGS Data Analysis

Quality of reads was fine for all the samples generated by FastQC software (Figure 28). Base calling in all the samples was also fine as observed by Per Base Sequence Content plots (figure 29).

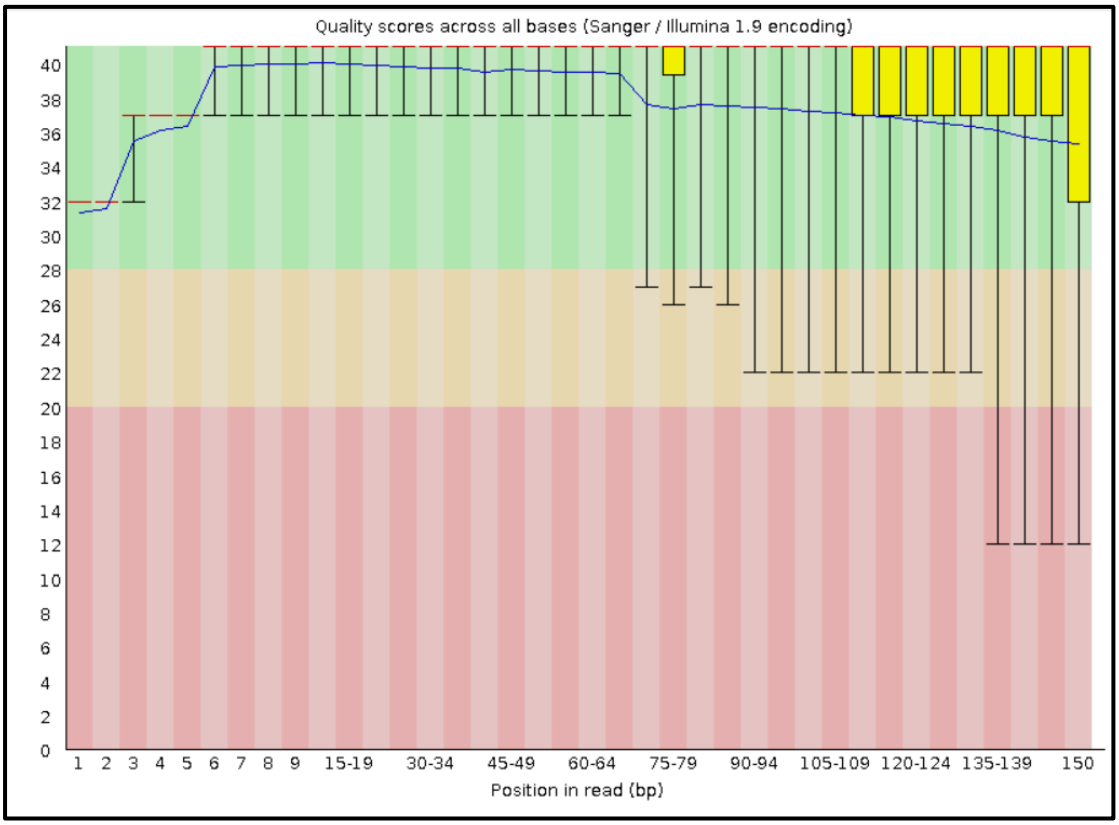

Figure 28 Per Base Sequence Quality by FastQC Software. Y-axis is quality score; green zone represents good quality score and pink represents low quality reads. (Representative image for S2.1 sample)

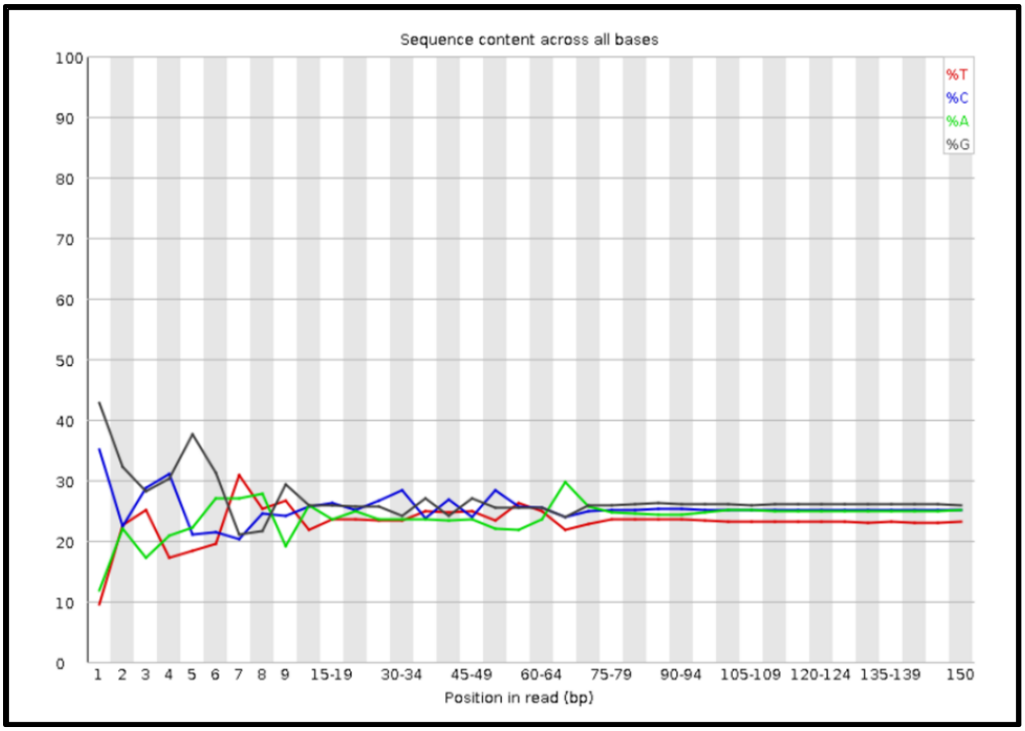

Figure 29 Per base Quality Index. Proportion of each base position in a file for which each of the four normal DNA bases has been called. (Representative image for S2.1 sample)

Correlation plot has shown good correlation between the duplicates, with R value between 0.95-0.98 (Figure 30). Comparative analysis from samples S1 to S3 as described in figure 31, gave us 85 differentially expressed genes. Data analysis for rest of time points is in process.

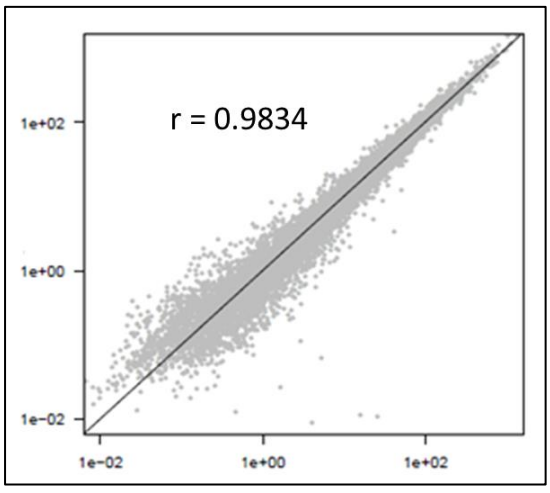

Figure 30 Correlation Plot for Duplicates (Representative image for S2 sample)

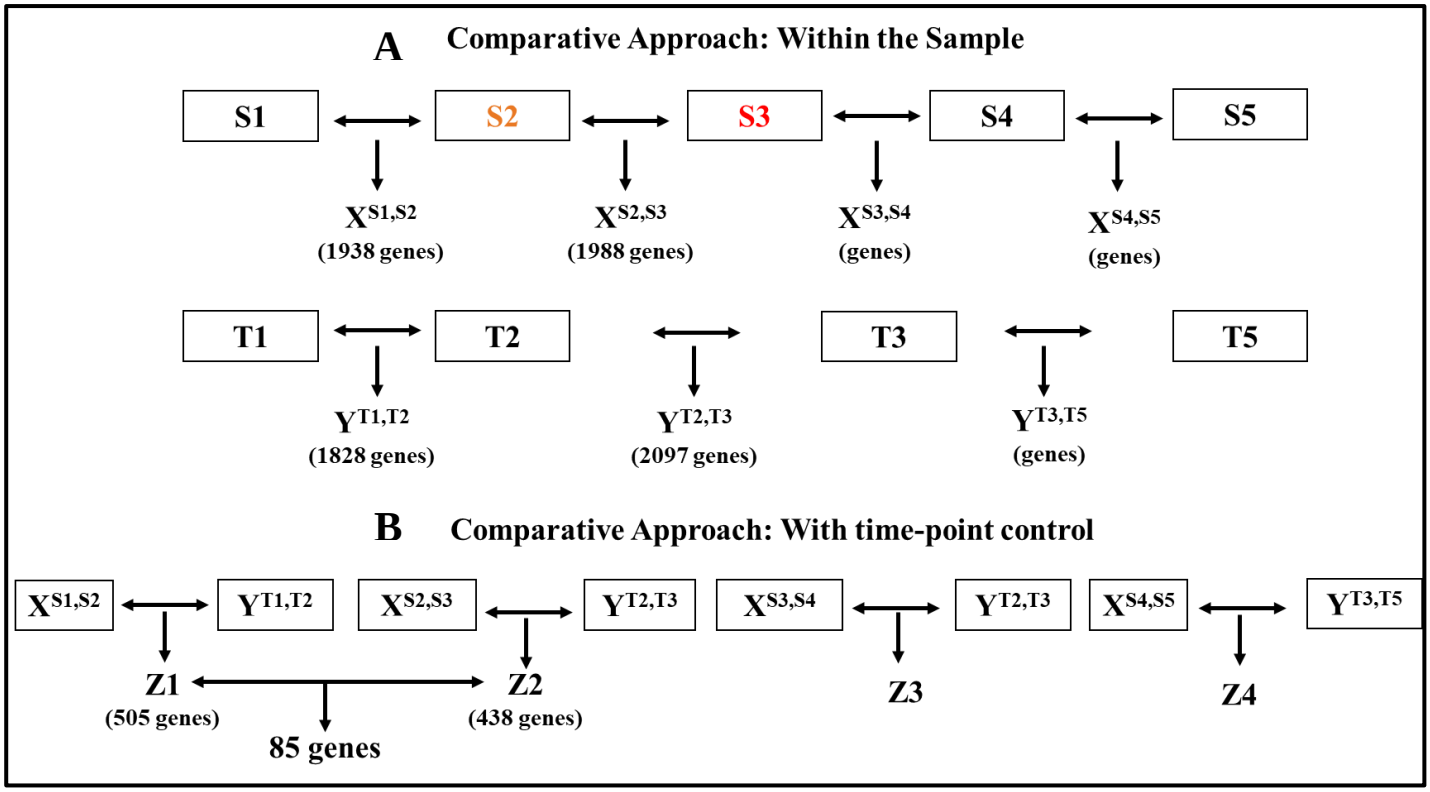

Figure 31 NGS Data Analysis: Comparative Approach (A) Within the Sample (B) With Time-point Control

Chapter 4

Discussion

Chapter 4: Discussion

The active state of HIV-1 is a well-understood process mediated by its master regulatory circuit, however, the mechanism of latency establishment is still unclear. The long-standing debate is whether the decision to establish latency is predominantly from the host cell or the virus. Some studies [11] demonstrated the entry of the virus into latency irrespective of the cellular state (active or rest), suggesting that an intrinsic viral circuit regulates latency. To understand the mechanism of latency, it is important to know the interplay between host factors and the viral transcription circuit. A few studies analysed the effect of HIV-1 replication on host gene expression using either NGS or microarray [27]. A different study examined the dynamics of cellular gene expression in HIV-1 LAI infected CD4+ cells (SUP-T1) using total RNA-seq at 12 and 24 h.p.i [28]. This study identified differential regulation of many novel poly(A)- and poly(A)+ early host transcripts in response to viral gene expression. However, from all the published reports, it is still not evident how HIV-1 transcription circuitry or Tat-feedback loop modulates host gene expression and vice-versa. The impact of the host transcriptome on the latent state of HIV-1 was investigated using single-cell RNA-seq to concluded that transcriptional signature in subset of T-cells influences viral latency [29]. In this backdrop, it is important to dissect different phases of viral infection to understand the complexity of latency establishment as well as its maintenance.

Preliminary work from our laboratory demonstrated the presence of Tat in latent Jurkat cell population suggesting the prominent effect of viral transcription circuitry on the host gene expression during latency cycle. In this pilot experiment, we examined how the pattern of host gene expression is modulated by the transcriptional activity of the virus through the complete the latency cycle. The preliminary RNA-seq data analysis identified the differential regulation of Our initial approximately 85 genes through the latency cycle. The data analysis is still in progress.

The strength of the experimental strategy used here is the ability to identify different and sequential stages of the complete HIV-1 latency cycle from latency to activation to latency. Using two different fluorescent proteins containing variable half-life stabilities and both the proteins co-expressed under the control of the viral promoter, we could identify a subset of cells immediately after the transcriptional switch off of the viral promoter, for the first time. Through the pilot work and the leads obtained, we realised a few technical limitations of the

present work. The global activation used to switch on the latent viral promoter showed a dramatic and profound influence on the cellular activation, thus, masking any possible signalling events specifically caused by the viral infection. Work is presently in progress to devise a novel molecular strategy to specifically activate the latent viral promoter without causing the cellular activation using the Tet-On regulatory circuit. Additionally, the work will also be extended to primary CD4+ T cells and full-length viral infection.

References

1 Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E., Hammer, S. M. 2008 The challenge of HIV-1 subtype diversity. *N Engl J Med*. **358**, 1590-1602. (10.1056/NEJMra0706737)

2 Frankel, A. D., Young, J. A. 1998 HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem*. **67**, 1-25. (10.1146/annurev.biochem.67.1.1)

3 Neogi, U., Sood, V., Ronsard, L., Singh, J., Lata, S., Ramachandran, V. G., Das, S., Wanchu, A., Banerjea, A. C. 2011 Genetic architecture of HIV-1 genes circulating in north India & their functional implications. *Indian J Med Res*. **134**, 769-778. (10.4103/0971-5916.92624)

4 Mbonye, U., Karn, J. 2017 The Molecular Basis for Human Immunodeficiency Virus Latency. *Annu Rev Virol*. **4**, 261-285. (10.1146/annurev-virology-101416-041646)

5 Cary, D. C., Fujinaga, K., Peterlin, B. M. 2016 Molecular mechanisms of HIV latency. *J Clin Invest*. **126**, 448-454. (10.1172/JCI80565)

6 Jean, M. J., Power, D., Kong, W., Huang, H., Santoso, N., Zhu, J. 2017 Identification of HIV-1 Tat-Associated Proteins Contributing to HIV-1 Transcription and Latency. *Viruses*. **9**, (10.3390/v9040067)

7 Hetzer, C., Dormeyer, W., Schnolzer, M., Ott, M. 2005 Decoding Tat: the biology of HIV Tat posttranslational modifications. *Microbes Infect*. **7**, 1364-1369. (10.1016/j.micinf.2005.06.003)

8 Kaehlcke, K., Dorr, A., Hetzer-Egger, C., Kiermer, V., Henklein, P., Schnoelzer, M., Loret, E., Cole, P. A., Verdin, E., Ott, M. 2003 Acetylation of Tat defines a cyclinT1-independent step in HIV transactivation. *Mol Cell*. **12**, 167-176.

9 Agbottah, E., Deng, L., Dannenberg, L. O., Pumfery, A., Kashanchi, F. 2006 Effect of SWI/SNF chromatin remodeling complex on HIV-1 Tat activated transcription. *Retrovirology*. **3**, 48. (10.1186/1742-4690-3-48)

10 Pagans, S., Pedal, A., North, B. J., Kaehlcke, K., Marshall, B. L., Dorr, A., Hetzer-Egger, C., Henklein, P., Frye, R., McBurney, M. W., *et al.* 2005 SIRT1 regulates HIV transcription via Tat deacetylation. *PLoS Biol*. **3**, e41. (10.1371/journal.pbio.0030041)

11 Chavez, L., Calvanese, V., Verdin, E. 2015 HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells. *PLoS Pathog*. **11**, e1004955. (10.1371/journal.ppat.1004955)

12 Siliciano, R. F., Greene, W. C. 2011 HIV latency. *Cold Spring Harb Perspect Med*. **1**, a007096. (10.1101/cshperspect.a007096)

13 Williams, S. A., Greene, W. C. 2007 Regulation of HIV-1 latency by T-cell activation. *Cytokine*. **39**, 63-74. (10.1016/j.cyto.2007.05.017)

14 Strahl, B. D., Allis, C. D. 2000 The language of covalent histone modifications. *Nature*. **403**, 41-45. (10.1038/47412)

15 Ishida, T., Hamano, A., Koiwa, T., Watanabe, T. 2006 5' long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals. *Retrovirology*. **3**, 69. (10.1186/1742-4690-3-69)

16 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. *J Virol*. **85**, 9078-9089. (10.1128/JVI.00836-11)

17 Palmer, A. C., Egan, J. B., Shearwin, K. E. 2011 Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription*. **2**, 9-14. (10.4161/trns.2.1.13511)

18 Lenasi, T., Contreras, X., Peterlin, B. M. 2008 Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe*. **4**, 123-133. (10.1016/j.chom.2008.05.016)

19 Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., *et al.* 2012 Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell*. **150**, 1147-1157. (10.1016/j.cell.2012.08.004)

20 Huang, H., Santoso, N., Power, D., Simpson, S., Dieringer, M., Miao, H., Gurova, K., Giam, C. Z., Elledge, S. J., Zhu, J. 2015 FACT Proteins, SUPT16H and SSRP1, Are Transcriptional Suppressors of HIV-1 and HTLV-1 That Facilitate Viral Latency. *J Biol Chem*. **290**, 27297-27310. (10.1074/jbc.M115.652339)

21 Clark, E., Nava, B., Caputi, M. 2017 Tat is a multifunctional viral protein that modulates cellular gene expression and functions. *Oncotarget*. **8**, 27569-27581. (10.18632/oncotarget.15174)

22 Verhoef, K., Bauer, M., Meyerhans, A., Berkhout, B. 1998 On the role of the second coding exon of the HIV-1 Tat protein in virus replication and MHC class I downregulation. *AIDS Res Hum Retroviruses*. **14**, 1553-1559. (10.1089/aid.1998.14.1553)

23 Peruzzi, F. 2006 The multiple functions of HIV-1 Tat: proliferation versus apoptosis. *Front Biosci*. **11**, 708-717.

24 Debaisieux, S., Rayne, F., Yezid, H., Beaumelle, B. 2012 The ins and outs of HIV-1 Tat. *Traffic*. **13**, 355-363. (10.1111/j.1600-0854.2011.01286.x)

25 Marban, C., Su, T., Ferrari, R., Li, B., Vatakis, D., Pellegrini, M., Zack, J. A., Rohr, O., Kurdistani, S. K. 2011 Genome-wide binding map of the HIV-1 Tat protein to the human genome. *PLoS One*. **6**, e26894. (10.1371/journal.pone.0026894)

26 Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T. 2017 Transcriptomics technologies. *PLoS Comput Biol*. **13**, e1005457. (10.1371/journal.pcbi.1005457)

27 Devadas, K., Biswas, S., Haleyurgirisetty, M., Wood, O., Ragupathy, V., Lee, S., Hewlett, I. 2016 Analysis of Host Gene Expression Profile in HIV-1 and HIV-2 Infected T-Cells. *PLoS One*. **11**, e0147421. (10.1371/journal.pone.0147421)

28 Peng, X., Sova, P., Green, R. R., Thomas, M. J., Korth, M. J., Proll, S., Xu, J., Cheng, Y., Yi, K., Chen, L., *et al.* 2014 Deep sequencing of HIV-infected cells: insights into nascent transcription and host-directed therapy. *J Virol*. **88**, 8768-8782. (10.1128/JVI.00768-14)

29 Bradley, T., Ferrari, G., Haynes, B. F., Margolis, D. M., Browne, E. P. 2018 Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. *Cell Rep*. **25**, 107-117 e103. (10.1016/j.celrep.2018.09.020)