Establishment and characterization of a cell model for HIV-1 latency with a Tat inducible expression system

Thesis submitted in partial fulfilment of the requirements of the
Five year BS-MS Dual Degree Program

Indian Institute of Science Education and Research, Pune

By

Surabhi Jirapure (20101024)
Biology

Under the guidance of

Prof. Udaykumar Ranga
Molecular Biology and Genetics Unit
JNCASR, Bengaluru

Certificate

This is to certify that this dissertation entitled **"Establishment and characterization of a cell model for HIV-1 latency with a Tat inducible expression system"** towards the partial fulfilment of the five year BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents original research carried out by **Ms. Surabhi Jirapure** at Molecular Biology and Genetics Unit, JNCASR under the supervision of **Prof. Udaykumar Ranga,** Molecular Biology and Genetics Unit, JNCASR during the academic year 2014-2015.

Prof. Udaykumar Ranga

Date: March 25th, 2015

Molecular Biology and Genetics Unit

JNCASR, Bengaluru

Declaration

I hereby declare that the matter embodied in the report entitled **"Establishment and characterization of a cell model for HIV-1 latency with a Tat inducible expression system"** are the results of the investigations carried out by me at the Molecular Biology and Genetics Unit, JNCASR, Bengaluru under the supervision of **Prof. Udaykumar Ranga** and the same has not been submitted elsewhere for any other degree.

Surabhi Jirapure

Date: 25th March 2015

Five year BS-MS Dual degree student

IISER Pune

Abstract

Highly Active Anti-Retroviral Therapy (HAART) reduces the viral load in the patients to undetectable titres but it is incapable of eradicating the virus from the patient's system. HIV is notorious for having the propensity to hide inside a patient's memory T cells and silencing its own transcription thus going into a state called latency. Several theories have been proposed and tested towards how this latency is established and maintained, one of which states that the viral protein Tat has a very important role to play. Studies have shown that Tat positive feedback loop supersedes the cell-driven silencing of the viral LTR transcription thus preventing the virus from staying in its latent phase. HIV-1C is known to have three functional NFκB TFBS but a recent study has shown that a strain of subtype C virus having an additional functional NFκB site is fast evolving. In this study, we separated Tat from the HIV-1C viral genome so as to have an extrinsic control over its expression which is dependent on the addition of Doxycycline. We generated isogenic variants of the virus which differ in the number of NFκB sites (0-5 sites). Using a dose response study, we found out the optimum concentration of doxycycline needed to drive Tat expression. Thus we developed a Tat inducible latency model which on addition of the viral LTR can be used to study the establishment and maintenance of latency in HIV-1C. Since we have 5 constructs varying in the number of NFκB sites, this model can further be used study how their reactivation kinetics vary.

List of Figures

List of Tables

Acknowledgements

I am greatly indebted to my guide, Prof. Udaykumar Ranga, JNCASR, Bengalure. He stirred curiosity in me and steered me at every step with his guidance, kind cooperation and good suggestions which have enabled me to progress so much in this project.

My project would not have been complete without Sutanuka Das (Ku), who helped me with all the technicalities involved. She has been a constant help and a great mentor who was my source of strength and support.

My gratitude to my lab mates, Anjali, Malini, Prabhu and Shilpee for teaching and assisting me all throughout the project.

I wish to thank Shambhu, Arun, Ria, Neelakshi, Disha, Siddharth, Deepak and Chaitra for being there whenever I needed urgent help and keeping a fun filled environment in the lab.

Finally I thank my friends and family for showing faith and providing confidence during my entire IISER life and my Bengaluru gang especially during this final year.

Introduction

The Human Immunodeficiency Virus (HIV), the etiological agent responsible for the acquired immunodeficiency syndrome (AIDS), was first observed in 1981 in the United States (Mandell et al., 2010). However, the virus was independently discovered by two research groups simultaneously in 1983: Robert Gallo and Luc Montagnier, and their findings were published in the same issue of the journal *Science*(Barre-Sinoussi et al., 1983; Gallo et al., 1983). By 1986, it was established that the virus Gallo and Montagnier had reported was the same and the Centers for Disease Control and Prevention (CDC) officially named it HIV.

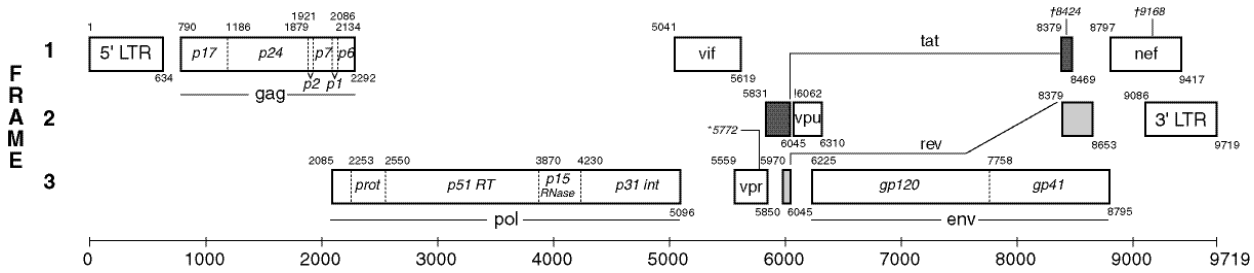
HIV belongs to the genus *lentivirus* of the family *Retroviridiae*. It uses RNA as its genetic material and infects mostly the T-helper cells and rarely the macrophages and microglial cells, thus progressively weakening the immune system. The replication cycle begins when the viral envelope fuses with the host cell membrane ($CD4^+$ T cells or macrophages) and the viral capsid enters the cell. The virally encoded enzyme reverse transcriptase converts the viral RNA into a double stranded complementary DNA (cDNA) molecule in the cytoplasm, which now enters the cell's nucleus and integrates into the host chromatin. Viral transcripts from the integrated provirus are then transcribed from the promoter at the 5' Long Terminal Repeat (LTR) to an mRNA which is then spliced into smaller pieces and these smaller mRNA molecules are transported from the nucleus into the cytoplasm. Following this, the spliced mRNAs are translated into regulatory and structural proteins in the cytoplasm. The polyproteins together with two copies of full length viral mRNA which actually represent the viral genome then assemble at the cell membrane and finally the immature virion begins to bud out from the host cell. As the virion buds out and is released from the cell surface, it undergoes maturation wherein the polyproteins are processed to result into a mature virion. This viral particle is now ready to infect new host cells.

i. Classification of HIV:

HIV is classified into two types based on their hosts of origin: HIV-1 in chimpanzees and HIV-2 in sooty mangabey. HIV-1 in itself is not a single virus but encompasses four groups M, N, O and P with M being the first lineage to be discovered and the one responsible for a global pandemic. Group M is further subdivided into 9 different subtypes, A, B, C, D, F, G, H, J and K(Santos and Soares, 2010) and over 40 circulating recombinant forms (CRFs) (Dhar et al., 2012). While subtype B of HIV-1M is most prevalent in the United States and Australia, subtype C is responsible for more than 90% of the infections in the Indian subcontinent (Santos and Soares, 2010).

ii. HIV-1 genome organization:

The HIV-1 genome is approximately 10 kb long, flanked by two LTRs and codes for 9 proteins. Three major structural genes are encoded by the Gag, Pol and Env polyproteins and six accessory proteins by Vif, Vpr, Nef; found in the viral particle. Tat and Rev are essential for gene regulation and Vpu assists in assembly of the virion at the host cell membrane (Fig 1).



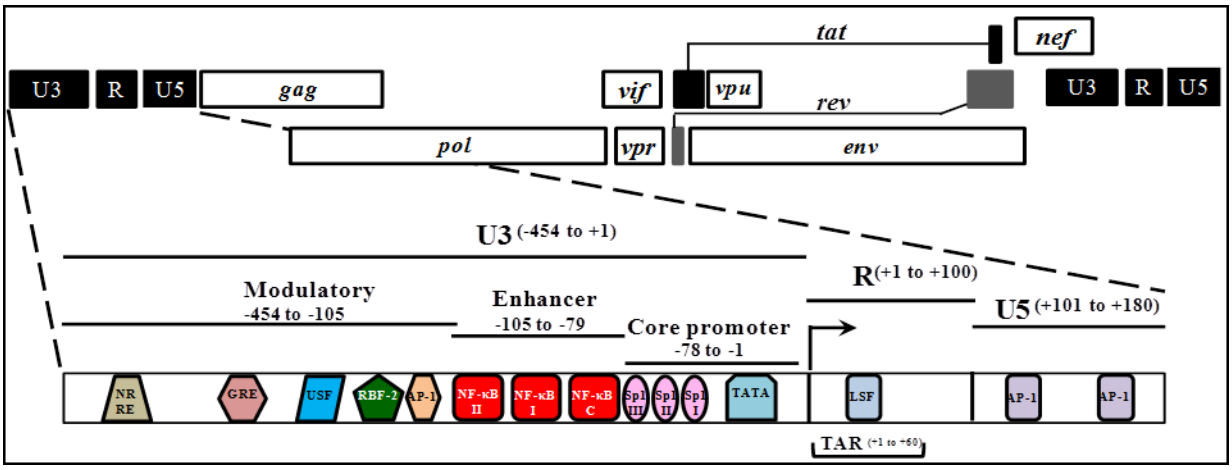
http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

Fig. 1: HIV-1 genome showing all the viral proteins

HIV-1promoter: An emphasis on subtype-C enhancer

The 10 kb genome of HIV-1 is flanked by two LTRs with 3' LTR functioning as the terminator signal for transcription and 5' LTR being the viral promoter. Each LTR, about 640 bp long is further segmented into three regions; U3, R and U5. During the multi-step process of reverse transcription, the viral progeny inherits only the U3 region of the

3’ LTR while the R region is inherited from the 5’ end resulting in a cDNA molecule with two identical LTRs at either end. The R region has a **T**rans **A**ctivator **R**esponse (TAR) element which forms a hairpin structure that interacts with the regulatory protein Tat and this interaction is essential for successful elongation of the full length viral transcript. The U3 region has a modulatory, enhancer and a core promoter region. A TATA box is located within the core promoter region. The U3 is the hotspot of several **T**ranscription **F**actor **B**inding **S**ites (TFBS) such as **S**pecificity **P**rotein **1** (SP1), **N**uclear **F**actor **kappa**-light-chain-enhancer of activated **B** cells (NFκB), **A**ctivator **P**rotein-**1** (AP-1), **N**uclear **F**actor of **a**ctivated **T** cells (NFAT) and **U**pstream **S**timulating **F**actor (USF).The HIV-1 LTR is one of the most conserved regions among individual subtype genomes and plays a key role in determining the virulence of a virus. Substantial sequence and copy number variation in the TFBS- AP-1, NFAT and USF, is noted across all HIV-1 subtypes. However, the enhancer region consisting of the NFκB sites show maximum polymorphism. Most subtypes have two NFκB sites; subtype C being an exception with three.



(Thesis, B Mahesh, JNCASR, 2012)

Fig 2: Representative HIV-1 subtype-C LTR

Subtype C enhancer is unique in terms of showing not only copy number differences but also genetic heterogeneity of the NFκB sites. Naturally, majority of the C isolates have three NFκB sites. However, a few variants with yet another additional κB has been

reported(G. Hunt, 2000). In the prototype C- enhancer, two canonical NFκB sites are genetically identical to those found in subtype B and are henceforth referred to as H-κB site (5' GGGACTTTCC 3') while the third κB site unique to subtype C differs in two base pairs and has been named C-κB site (5' GGG**G**C**G**TTCC 3'). The fourth site in the new subtype-C variant which has been named F-κB (fourth) differs from the H-κB site at one base pair (5' GGGACTTTC**T**3'). A recent finding from our lab has shown that the newly evolving subtype-C variant with four κB sites have a stronger replicative fitness as compared to the wildtype C variant containing three κB sites (Bachu et al., 2012).

HIV-1 Latency:

The treatment currently used for AIDS is referred to as **H**ighly **A**ctive **A**nti-**R**etroviral **T**herapy (HAART). It is a combination of three or more drugs that act on different viral targets thereby reducing the viral load below the detection limit. Nevertheless, the combinatorial therapy fails to render the patient a functional HIV cure, the reason being the ability of the virus to establish latent infections. Once integrated into the host chromosome, the proviral DNA can lie quiescent until activated a phenomenon termed "viral latency".

The cells in which the virus lies dormant are termed "latent reservoirs". Combination of anti-retrovirals creates a hostile environment for the virus and dissuades it from replicating. However, since these drugs only act on the actively replicating virus, they reduce the viral load below detection levels but fail to eradicate it.

Blankson et al., 2002, define a latent reservoir as "a cell type or anatomical site in association with which a replication-competent form of the virus accumulates and persists with more stable kinetic properties than the main pool of actively replicating virus". Latently infected resting CD4+ T cells form the major component of the latent HIV reservoir, others being macrophages (Igarashi et al., 2001), stem cells or progenitor cells (Carter et al., 2010). First described by the Siliciano lab in 1995 (Chun et al., 1995), these T cells are devoid of CD69, CD25, and HLA-DR activation markers (Han et al., 2007).

Several strategies have been proposed to eradicate the reservoir. One strategy is to supply more anti-retrovirals to eliminate replication. Another proposes stimulation of latently infected cells so they produce virions and are resultantly cleared by the immune system. Yet another strategy suggests the stimulation of the immune response to remove the latently infected cells (Pace et al., 2011) (not needed in brackets).

Mechanisms of establishment and maintenance of latency

The mechanism of viral latency establishment and maintenance is a daunting question and one that does not yet have a definite answer. It is however proposed as a multifactorial process with contributions from both the host cellular and the viral counterparts. It can be either an event of pre-integration (as a result of blocked viral life cycle before the virus can integrate into the host genome or unsuccessful import of the preintegration complex into the nucleus) or post-integration (as a result of failure in transcription from the integrated provirus (Colin and Lint, 2009). Broadly, these mechanisms can be classified into epigenetic and non-epigenetic factors.

Epigenetic factors include chromatin remodeling and epigenetic modification. Two nucleosomes nuc-0 and nuc-1 are positioned in the 5'LTR and the chromatin structure is remodeled by nuc-1 by ATP dependent remodeling complexes (Colin and Lint, 2009). In latent state, nuc-1 is deacetylatedby HDACs (Histone deacetylases). Absence of cellular transcription factors, transcriptional interference between the viral promoter and a neighboring cellular promoter, non-diving stage of the host cell are some of the non-epigenetic causes of post-integration latency.

Current models to study HIV latency

Several models are being used to understand the phenomenon with an aim to achieve a sterile cure for HIV. *In vivo* models currently used are SIV model in non-human primates like rhesus macaques and humanized mice transplanted with human tissues or stem cells. While both *in vivo* models have their own advantages, the foremost disadvantages are the time and cost to establish the model systems. On the other hand, *in vitro* models give a much faster readout and are also more suited for biochemical screening, but they

fail to accurately imitate the natural latent infections within a living system. The models and their research outcomes have been briefly described in the table.

Group	Cell type	Description	Result
Burke et al	Thymocytes	$CD4^+CD8^+$ thymocytes infected with HIV based reporter virus	NFκB signaling is important for HIV reactivation
Tyagi et al	Activated CD4+ T cells	Activated CD4+ cells infected with HIV derived vector and maintained in a feeder cell line	Epigenetic silencing and low P-TEFb levels result in HIV latency
Marini et al	Dendritic cells	AG-loaded dendritic cells cultured in IL-7	Reduced cell death of uninfected cells upon secondary stimulation
Bosque et al	Naïve CD4+	Activated naïve CD4+ cells primed and cultured in IL-2	Phenotypically varying latently infected cells suitable for drug screening
Yang et al	Primary cells	BCL2 transduced primary cells	Long lived cells suitable for drug screening
O'Doherty et al	Primary resting CD4+ T cells	Direct infection of primary cells with *wt* HIV	Latently infected cells reflecting *in vivo* latency pattern

Lewin et al	Primary resting CD4+ T cells	CCR7 stimulated primary cells	Enhanced HIV infection and cell viability

Table 1: *in vitro* models used for HIV latency studies

An ex-vivo HIV-1 latency model with inducible Tat expression:

We propose to develop and assess an ex-vivo model of HIV-1 latency in the immortalized CD4+ T cell line Jurkat containing an inducible expression system for the viral protein Tat. Tat is a regulatory protein which plays a crucial role in HIV-1 transcription by accelerating the elongation of full length viral transcripts thereby generating an autonomous positive circuit for viral gene expression. The present model attempts to uncouple the tat regulatory circuit from the viral promoter function and to control tat activity through an external molecular 'ON' or 'OFF' mode. The overall aim is to study subtype-C latency establishment and maintenance as a direct function of NFκB site polymorphism (Fig 3) with minimal Tat interference.

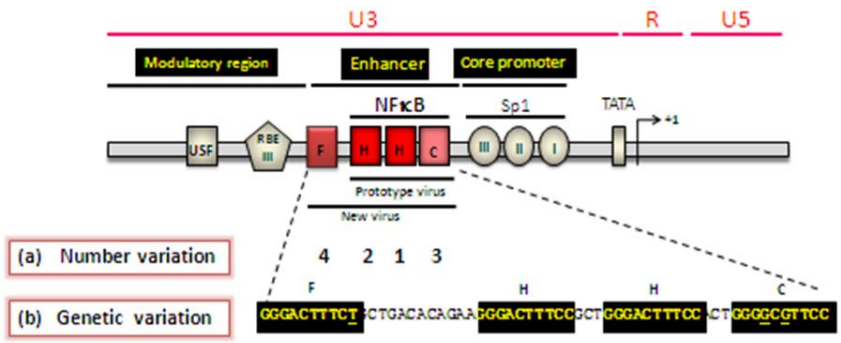


Fig. 3: Key variations in the subtype C enhancer with respect to NFκB sites

Tet-On system:

The Tet-On gene expression system is an inducible expression system where transcription of a gene is turned on in the presence of the antibiotic tetracycline or one

of its derivatives (e.g. doxycycline). In this system the rtTA (reverse tetracycline transactivator) protein on interaction with doxycycline undergoes a structural modification which favors its binding to the operator thereby driving transcription of the gene product. The presence and absence of Dox thus allows molecular ON or OFF switches for the transcription of the gene of interest.

The Tat inducible Jurkat cell line as a latency tool: To study the influence of NFκB site polymorphism in the establishment and maintenance of subtype C latency:

In 2009, Schaffer et al have used a lentiviral model of the Tat-mediated positive feedback loop (LTR-GFP-IRES-Tat, or LGIT)(Burnett et al., 2009) in Jurkat cells (immortalized human T lymphocytes) to study the contribution of individual TFBS (κB and SpI sites) in the subtype B promoter on stochastic expression of the tat protein that in turn governs the latency establishment and maintenance. Using the same model, our lab attempted to investigate the possible influence of NFκB site variations in subtype C latency. Towards this, isogenic viruses varying in κB site copy numbers were generated in the pcLGIT backbone by sequential disruption of individual κB sites using point mutations maintaining the length of the LTR region across the constructs. Preliminary results show that an inverse correlation exists between the number of κB sites and the rate of latency establishment in individual Jurkat populations singly integrated with isogenic κB viruses. Reactivation from latency however followed a direct response i.e, higher the number of κB sites, higher is the number of % reactivated cells and vice versa. However, establishment of latency requires a much longer time in this model. Moreover, presence of Tat in the lentivector and its positive feedback loop from the viral promoter results in continuous viral transcription and hampers latency establishment. The lentivector used in the current model lacks IRES and Tat regions of LGIT vector and allows subtype-C Tat to be monitored externally as an independent parameter to investigate its effect on latency.

The *in vitro* latency model we are trying to generate in the Jurkat cell line consists of the reporter protein (GFP) driven by the HIV-1 promoter whereas, subtype-C Tat is

introduced into an independent and inducible expression cassette based on the Tet-On system. The model has three-steps:

Step1: Selection of Jurkat cells transfected with rtTA3 (lentivector 1: CMV-rtTA3-IRES-Puro), using puromycin for selection. **(Parental line I)**

Step 2: Transfection of Parental line I with lentivector 2 (Tet-On-Tat-EF1α-ECFP) and selecting CFP^{low} cells to ensure minimal integration events and low Tat expression **(Parental line II)**. These will be validated for induced Tat expression from the Tet-ON promoter by varying the concentration of Doxycyline and an appropriate clone will be chosen.

Step 3: Parental line II will be infected with individual isogenic κB pseudoviruses (lentivector 3: LTR-GFP-LTR) containing GFP reporter under 5'LTR.

In figure 4, cells showing GFP expression have viruses in their active state and reduction in the percentage of GFP+ cells is indicative of cells having become latent. We see that virus with more functional κB sites (FHHC and OHHC) go towards latency much faster as compared to the virus with no functional κB sites (OOOO) which remains more or less constant in terms of its latency profile. The question we want to address is that latent state once established, how do these viruses maintain it? In the model that has been previously used (LTR-GFP-IRES-Tat, or LGIT), monitoring latency maintenance is not a practical option for the reason that Tat mediated feedback loop does not allow the cells to remain in their latent state. Tat not being under any extrinsic control prevents the LTR from being repressed.

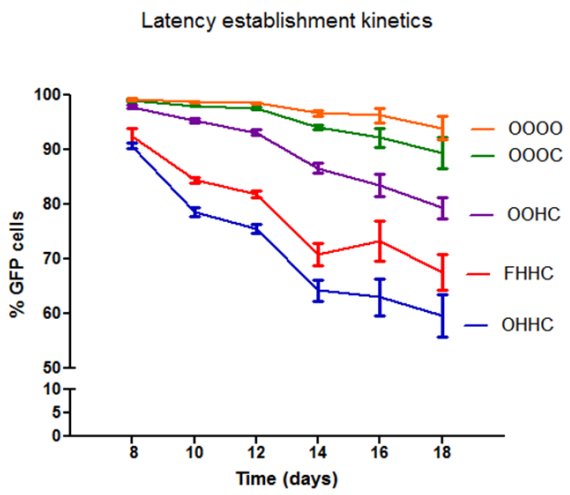


Fig 4: Latency establishment profiles of viruses varying in the number of functional NFκB sites

The principal aim of this project was to under the influence that NFκB site variation has on viral latency in the presence or absence of the viral protein Tat. To realize this aim, it was essential to decouple the Tat autonomous circuit from the LTR-GFP cassette in the previous lentivector (LTR-GFP-IRES-Tat) and control Tat expression as an independent parameter through an inducible gene expression system. We therefore, selected the Tet-ON inducible system to develop an *in vitro* latency model in Jurkats to extrinsically induce Tat by modulating doxycycline concentration in the medium. A recent study from Leor Weinberger's lab (Razooky et al., 2015) using a similar model shows that Tat positive feedback is in itself sufficient to silence HIV transcription and hence the establishment of latency independent of the cellular activation status. Our study will provide further understanding on the role of NFκB sites as an additional factor intrinsic to the virus on latency under conditions of normalized tat expression.

Materials and Methods

1. Cell culture

Jurkat, T-cell line, was cultured in RPMI 1640 medium (Cat. No. R4130, Sigma, St. Louis, USA), supplemented with 10% Fetal Bovine Serum (RM10435, HiMedia Laboratories), 2 mM glutamine (G8540, Sigma, St. Louis, USA), 100 units/ml penicillin G (P3032, Sigma, St. Louis, USA), and 100 µg/ml streptomycin (S9137, Sigma, St. Louis, USA). Parental line II low CFP Jurkat cells were cultured in RPMI 1640 medium (Cat. No. R4130, Sigma, St. Louis, USA), supplemented with 10% Tet system approved Fetal Bovine Serum (Cat. No. 631105, Clontech, USA). Doxycycline hyclate (D9891, Sigma, St. Louis, USA) used for dose and time response and further latency studies was used at 750 ng/ml. Human Embryonic Kidney (HEK) 293T cells were grown in DMEM medium (D1152, Sigma, St. Louis, USA).

2. Construction of lentivectors

i. Lentivector 2

Lentivector2 (pcDH-Tet-On Tat EF1α-ECFP) was constructed by replacing the sequence encoding GFP with that of CFP in the pcDHTet-On-Tat vector backbone. Lentivector 2 has subtype-C Tat under the Tet-On inducible promoter and an independent EF1α-ECFP cassette for the purpose of positive selection of successfully transduced cells. CFP recombinants were confirmed by XbaI linearization (a unique restriction site that selectively digests the recombinant but not the parent plasmid) and CFP expression was confirmed in HEK293T cells.

ii. Lentivector 3 (pcLG backbone)

The HIV-1 latency model vector pcLGIT has the elements- 5'LTR-GFP-IRES-C-Tat-3'LTR in the aforementioned order. The IRES and C-Tat regions were excised using BsrGI and BsiWI restriction enzymes and the ends were allowed to self-ligate owing to the presence of their compatible overhangs. The resulting pcLG clones were screened using enzymes AfeI and PmeI. pcLGIT releases an 800 bp AfeI-PmeI fragment while

pcLG being devoid of the AfeI site yields a single fragment 7.8 kb in length. Functionality of the selected clones was confirmed by checking GFP expression in HEK 293T cells. For tat transactivation assay, HEK 293T cells were co-transfected with the pcLG vector and c-Tat at 30% confluency. 24 hours post transfection, GFP intensity of the transactivated well was compared over and above the basal transcription under a fluorescence microscope (Olympus).

iii. Isogenic κB variants in lentivector 3 backbone

A panel of isogenic κB variant pseudovirus constructs in the pcLGIT backbone already existed in the lab that had different number of κB sites at the 3'LTRs in an otherwise isogenic backbone. The isogenic 3' LTRs in pcLGIT were derived from the subtype C molecular clone ***pINDIE***-C1 (***accession number***: AB023804). A similar panel of isogenic pseudoviruses was now engineered by replacing the 3' LTR in the pcLG backbone with the corresponding 3' LTR from the pcLGIT vector containing varying κB copy numbers. The NFκB site configuration of the isogenic LTRs is as follows: FHHC, OHHC, OOHC, OOOC and OOOO (where, F- 5' GGGACTTTCT 3'; O- 5'TCTACTTTTT 3'……) both the pcLGIT and pcLG vectors were digested using XhoI and PmeI and the resultant fragment from pcLGIT was inserted in the pcLG backbone. The pcLG recombinants were distinguished by digesting with NheI enzyme which is absent in the pcLG parent vector. Each isogenic pseudoviral construct was confirmed for GFP expression as well as transactivation following the same protocol as above.

3. Viral stock preparation and Virus titre determination

i. Viral Stocks

Viral stocks for all five pcLG constructs were made in HEK 293T cells using the standard calcium-phosphate protocol (Weinberger et al., 2005). Cells were seeded in a 100 mm dish at a density of $3x10^6$ cells and transfected with pcLG recombinant, c-Tat containing plasmid, psPAX2 and CMV-rev. 6 hours post transfection, medium was replaced with complete DMEM. Supernatant was harvested at 48 hours replenished with fresh DMEM media. Collected supernatant was centrifuged at 3000 rpm at room

temperature (RT) using table top centrifuge (Hareaus) and stored at -80℃ in multiple aliquots of 1 ml. 72 hours after transfection, the second harvesting was done.

ii. ELISA

The p24 levels of the viral stocks were measured using a p24 ELISA kit (4^{th} generation p24 ELISA kit, J. Mitra and Co. Pvt Ltd., New Delhi, India), as per the manufacturer's instructions.

iii. MOI

0.5×10^6 cells were taken and virus added at varying p24 equivalent concentrations (10 fold series i.e. 100, 10, 1, 0.1, 0.01 ng/ml). 6 hours post infection, virus was removed and 48 hours post infection CFP was measured using flow sorter. MOI was determined as (% CFP^+)/100.

4. Generation of parental line II

rtTA3-Jurkat stable cell line (Parental line I) which has Puromycin antibiotic resistance was previously generated in the laboratory. It was further transduced with lentivirus 2 (generated in HEK-293T by co-transfection of the lentivector 2 that contains the Tet-ON Tat cassette together with the packaging plasmids that carry the viral envelope and auxillary genes) to create Parental line II. 0.5×10^6 rtTA3-Jurkat cells were infected with lentivirus 2 at p24 equivalent of 0.01ng/ml. 7 days post infection, low CFP^+ cells were sorted using a flow sorter (BD FACS AriaIII Cell Sorter, BD biosciences). From this low CFP population, clonal cell lines were generated by sorting single cells into a 96 well plate and scaled up gradually. These clonal cell lines were recovered and frozen.

i. Selection of Parental II clones for Doxycycline dose response

0.5×10^6 cells from four Parental II clonal cell lines selected randomly were added to a 12 well culture plate and kept in 10% Tet system FBS-RPMI. 24 hours post this, Dox was added to the cells to achieve a final concentration of 1000 ng/ml. 24 hours post this, RNA was isolated form each dosage using Trizol (T9424, Sigma). RNA was quantified using NanoDrop ND 1000 and 2 µg was used to synthesize cDNA using a

commercial kit (BIO-65042, TetrocDNA synthesis kit, Bioline, London, UK). cDNA obtained was diluted 5 times with DEPC treated water and used for real time PCR. The real-time PCR was performed using a commercial kit (SensiFAST SYBR Mastermix kit, London, UK). Tat gene was amplified using primer pair N1783 (5'-GGAATCATCCAGGAAGTCAGCCCGAAAC-3') and N1784 (5'-CTTCGTCGCTGTCTCCGCTTCTTCCTG-3'). For normalization, transcript of GAPDH gene was amplified using primers N2232 (5'-GAGCTGAACGGGAAGCTCACTG-3') and N2233 (5'-GCTTCACCACCTTCTTGATGTCATC-3'). All the Real-time PCR analyses were performed using the Bio-rad CFX96touch real time PCR machine. Graphs were plotted using CFX Manager Software version 1.6.

ii. Doxycycline dose response in clone- 2c

1 x 10^6 cells from the selected Parental II clonal cell line were added per well in a 6 well plate with condition as mentioned in the previous section. Doxycycline was added in varying doses to get the final concentration of 0, 250, 500, 750, 1000 ng/ml. Tat induction was measured via real time PCR as described previously.

iii. Doxycycline dependent LTR transactivation as a function of time

To confirm Dox dependent LTR transactivation, 1000 ng/ml dox added to 3m. These cells were then transfected with Neat and 10x diluted pcLG virus. Every 24 hours GFP was monitored by flow sorter. Cells with no Dox added were used for control.

Results and Discussion

The Tat inducible latency model

The *in vitro* latency model we are trying to generate in the Jurkat cell line consists of the reporter protein (GFP) driven by the HIV-1 promoter whereas, subtype-C Tat is introduced into an independent and inducible expression cassette based on the Tet-On system (5).

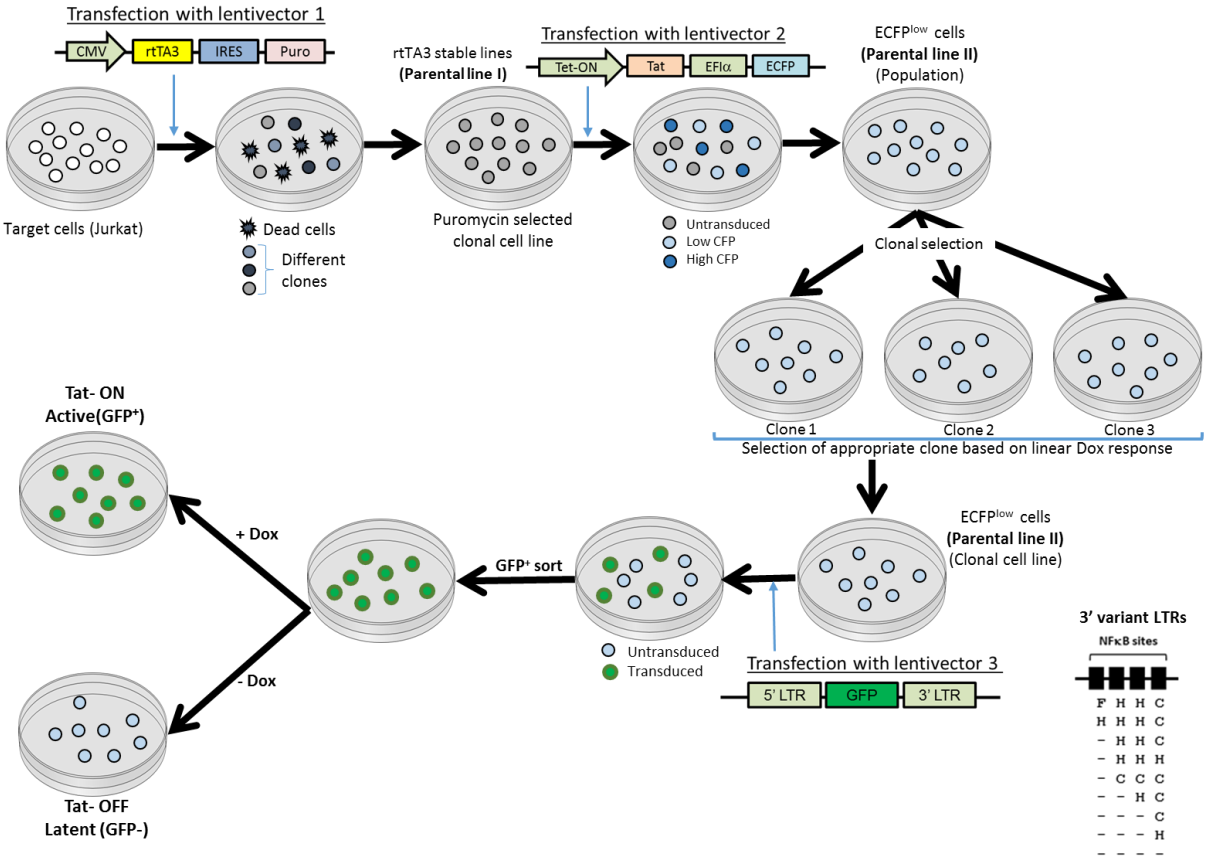


Fig 5: Tat inducible latency model

Lentivectors 1 and 2

Lentivector 1 has rtTA3 gene under CMV promoter and Puromycin gene for selection based on antibiotic resistance and was previously generated in the lab. Lentivector 2 has c-Tat under Tet-ON inducible promoter and an independent EF1α-ECFP cassette for selection of successfully transduced cells (Fig 6). CFP was chosen for the reason

that spectral overlap between GFP and CFP could be avoided using bandpass filters in the FACS ARIA III. Low CFP cells were chosen to ensure single integration events to mimic the viral integration in its natural state.

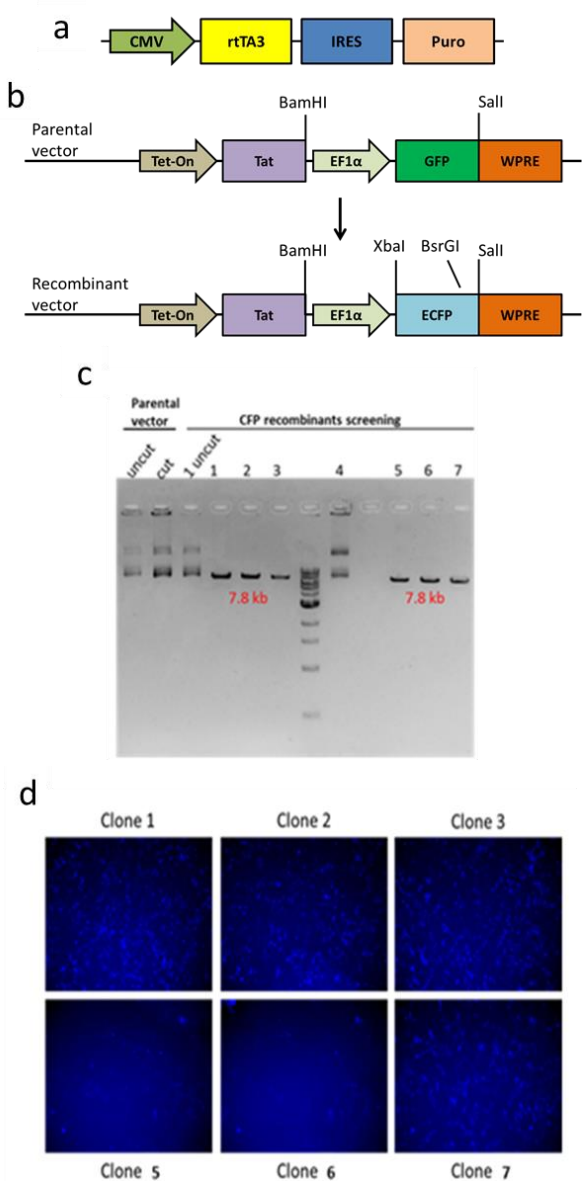


Fig 6: Lentivectors 1 and 2. (a) Lentivector 1. (b) Cloning strategy of lentivector 2. (c) 1% agarose gel for screening CFP recombinant clones by digesting with XbaI. (d) CFP expression in HEK293T cells.

1. Generation of lentivirus 2:

VSVG pseudotype virus was generated in HEK 293T cells as described and estimation of p24 in viral stocks was done using ELISA. Based on this 100, 10, 1, 0.1, 0.01 ng/ml p24 equivalent viral stocks were added to 1 x 10^6 cells of parental I cell line for multiplicity of infection (MOI) estimation using flow cytometry (Fig 7). Figure 7 shows images of infected cells under microscope as imaged using fluorescence.

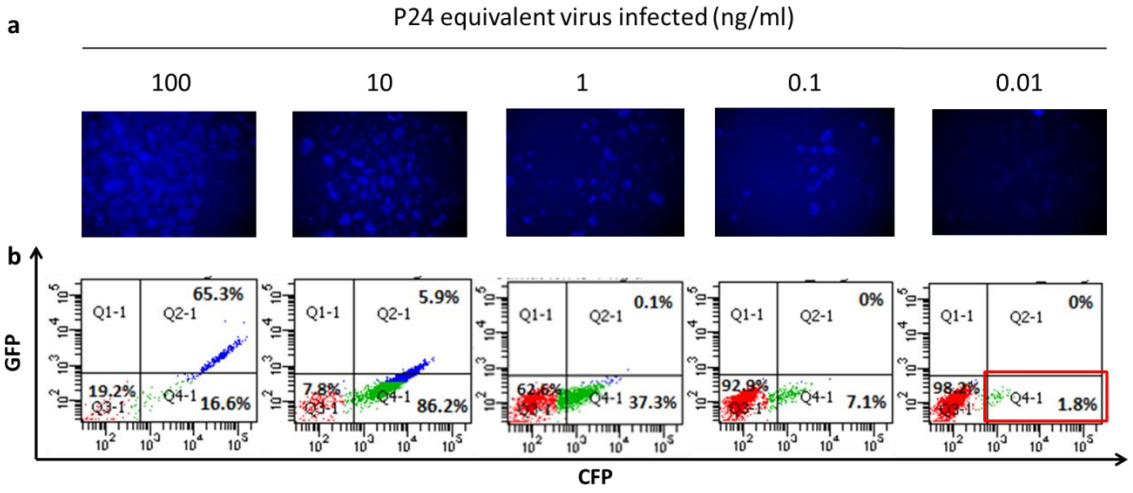


Fig 7: Titre estimation of lentivirus 2 by calculating multiplicity of infection in Jurkat rtTA3 (Parental line I). (a) Images taken under DAPI channel showing Jurkat rtTA3 infected with lentivirus 2. (b) GFP vs CFP flow cytometry plots. Quadrants Q1-1 to Q4-1: CFP^-GFP^+, CFP^+GFP^+, CFP^-GFP^-, CFP^+GFP^- respectively. Quadrant highlighted in red (CFP^+GFP^-) corresponding to 0.01 ng/ml p24 equivalent virus indicated the population sorted for generating Parental line II.

2. Generation of Parental II cell line

To account for single integration events of virus in Parental I line, lowest concentration of viral stock that shows cells which fluoresce visibly and using MOI calculation, 0.01 ng/ml pseudovirus viral concentration was decided. On infection with 0.01 ng/ml virus, an MOI of ~0.02 is observed, i.e. 2 out of every 100 cells are successfully infected.

Parental I cells infected with 0.02 MOI and showing low CFP fluorescence were sorted using FACS ARIA III and maintained in culture for 3 weeks. This is a population of cells

with lentivirus integration at random sites. To obtain a homogenous population with respect to the site of integration, cells were singly sorted into 4 96 well plates using FACS. Clonal cell lines thus generated were kept in 96 well plates for 2 weeks following which they were scaled up to 48, 24, 12, 6 well plates, maintain for a week at each step. 12 clonal cell lines were selected at random from these and frozen for further analysis.

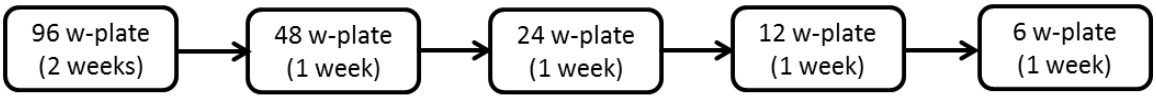


Fig 8: Generation of Parental line II. Representative strategy for scaling up of Parental line II clonal lines

3. Generation of lentivector and lentivirus 3

Lentivector 3 has GFP flanked by 3' and 5' HIV-1C LTRs. We have engineered isogenic variants of this vector which vary in the number and sequence of NFκB sites by replacing the 3' LTR in pcLG with the corresponding 3' LTR from pcLGIT vectors. Variant LTRs were generated from the FHHC variant by inactivating the F κB site by point mutations to maintain the LTR length across the isogenic pseudoviruses. OHHC configuration is representative of the prototype subtype C enhancer region constituting three functional NF κB sites. The other variants OOHC, OOOC and OOOO have two, one and zero functional NF κB sites respectively. Lentiviral stocks of these vectors were generated and transactivation of Tat was confirmed in HEK 293T cells 72 hours post infection by GFP imaging (Fig 9). HEK 293T cells were transfected with c-Tat containing plasmid (or pcDNA 3.1 as filler DNA) 24 hours prior to the addition of viral stocks. As was expected, enhanced GFP expression was detected in the presence of Tat.

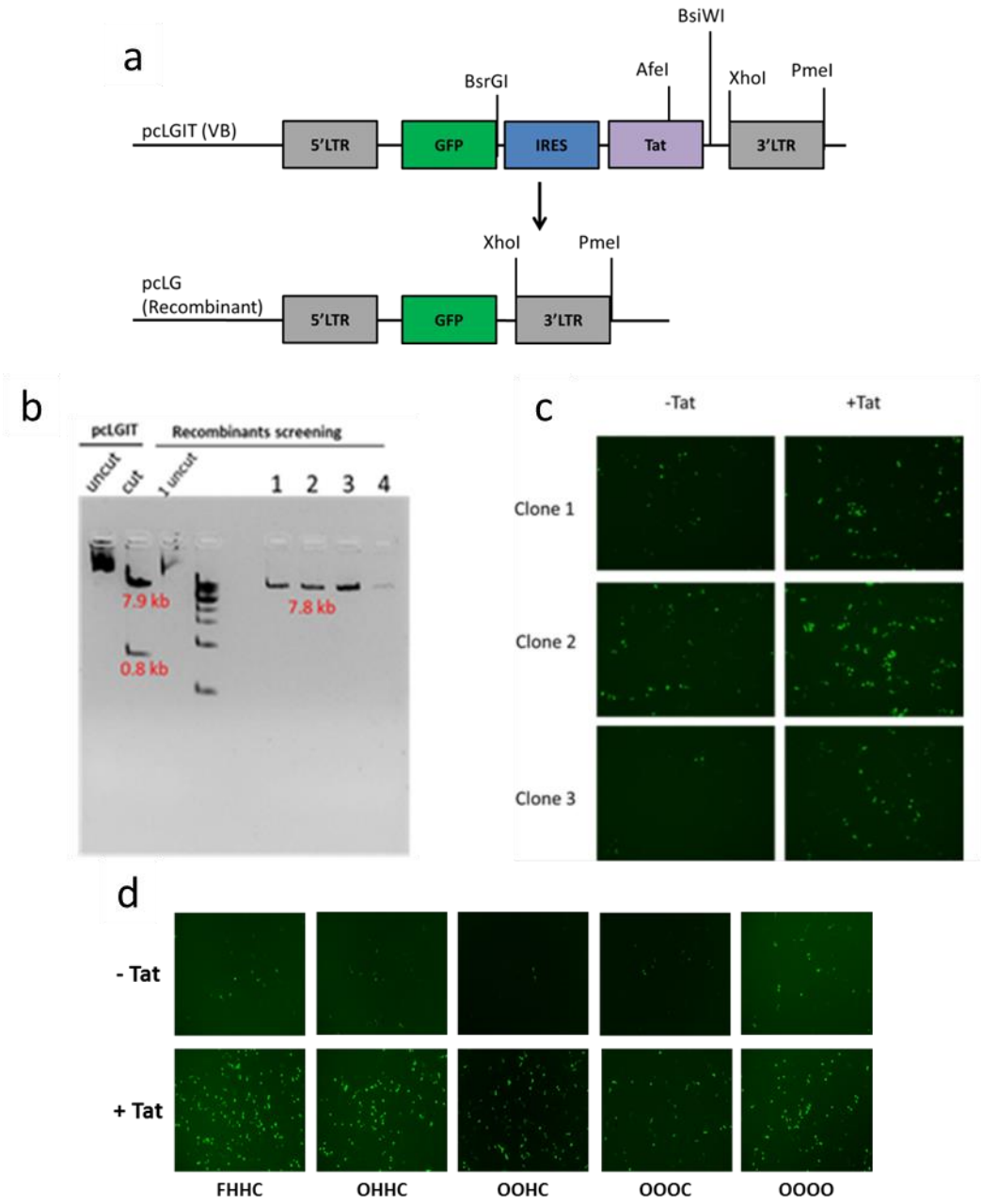


Fig 9: Generation and confirmation of lentivector and lentivirus 3. (a) Cloning of pcLG vector backbone. (b) 1% agarose gel for screening pcLG recombinants by digesting with AfeI + PmeI. (c) GFP expression from pcLG clones in HEK 293T in presence and absence of Tat. (d) GFP expression resultant of transactivation of isogenic variants of lentivirus 3 by Tat.

4. Validation of cell lines for dox dependent tat expression

As described previously, clonal cell lines were generated from the Parental line II population. The first attempt at generating these lines yielded a single clone, 3m. This clone was used to validate the dosage needed for latency establishment studies later.

i. Dox dose dependence of 3m

1 x 10^6 Jurkat rtTA3 low CFP 3m (referred to as low CFP 3m hereon) cells were taken per well in a 6 well plate for this assay. 24 hours after culturing cells in 10% Tet system FBS RPMI, varying doses of Dox were added to the wells to get the final concentrations of 0, 250, 500, 750 and 1000 ng/ml. Prior to performing this assay on clonal lines, it was done on the Parental II line population (0, 62.5, 125, 250, 500, 1000 ng/ml final Dox concentrations) to estimate the average fold induction of Tat, measured by the means of real time PCR using normalized GAPDH. While the population shows a relative Tat induction of roughly 140 folds, low CFP 3m shows only a 12 fold increase (Fig 10).

ii. Varying virus or Dox concentration to check for tat functionality (Pulse chase)

To rule out the possibility of Tat being dysfunctional, low CFP 3m cells were subjected to three different conditions involving varying the virus (cLG OHHC) or Dox concentration added. Percentage of cells showing GFP expression was measured using a flow sorter as a representative of Tat expression. All three reaction conditions show a negligible percentage of cells infected cells indicating a dysfunctional Tat or virus (Table 2).

iii. Other clones validated

Dox dose response studies previously done in the lab indicate that the expected relative Tat induction is anywhere between 100 and 200 fold (data not shown). As demonstrated from the previous result, low CFP 3m does not meet the expectation so we screened 3 more clones to carry out further experiments. Of the 12 clones obtained in the second round of Parental II clonal lines establishment, three clones (1a, 2c and 4b) were selected at random. 0.2 x 10^6 cells of each clone taken and treated with 1000 ng/ml Dox

for 24 or 48 hours along with untreated cells of each clone as controls. Real time analysis shows the relative fold induction of Tat of the four clones (Fig 11). While low CFP 3m shows the least induction (20 fold), low CFP 2c shows the highest Tat induction (200 fold).

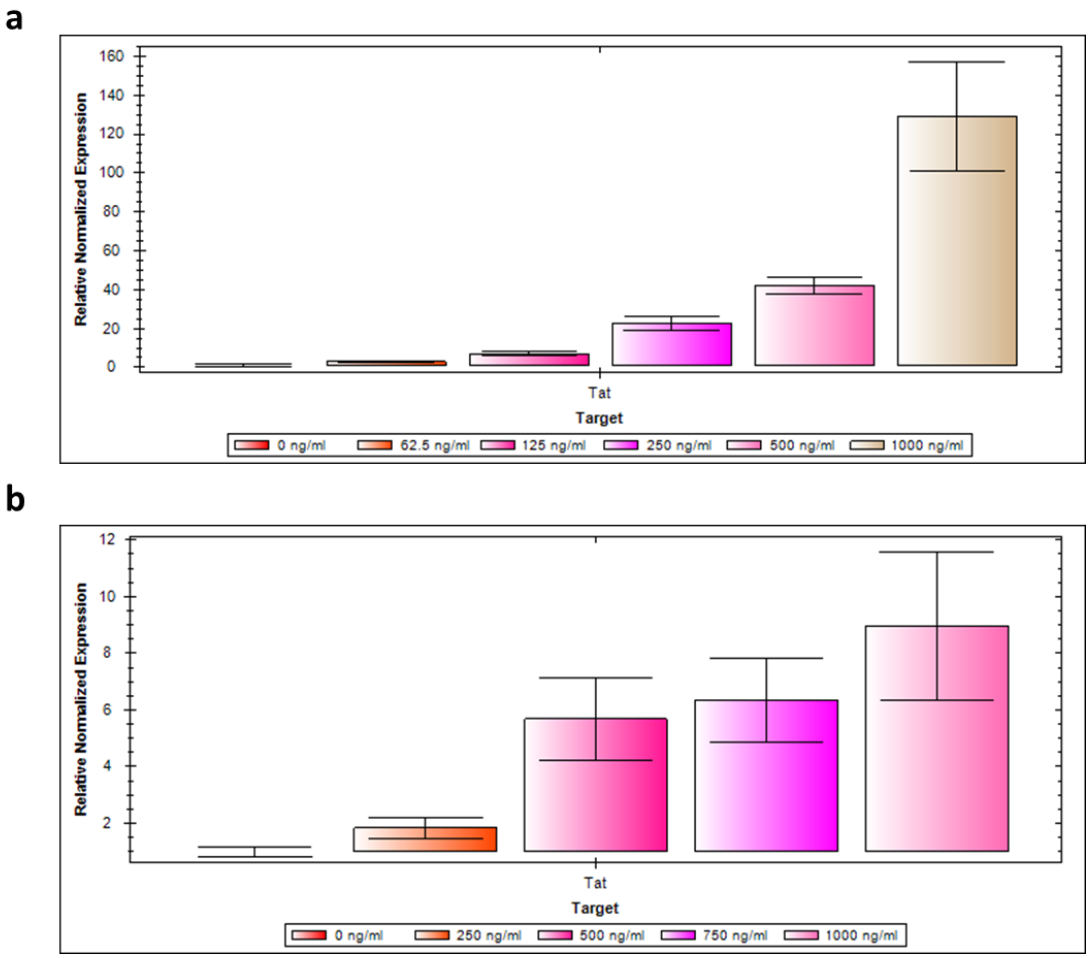


Fig 10: Real time analysis to measure fold induction of Tat in a dose response study of Parental II line. (a) Parental II line population. (b) Low CFP 3m.

Reaction condition	% CFP^+GFP^+ cells
750 ng/ml Dox + pure virus	0
1000 ng/ml Dox + 10x diluted virus	0.1
1000 ng/ml Dox + pure virus	0.1

Table 2: percentage of GFP cells as an indicator of Tat functionality on varying concentrations of virus or Dox

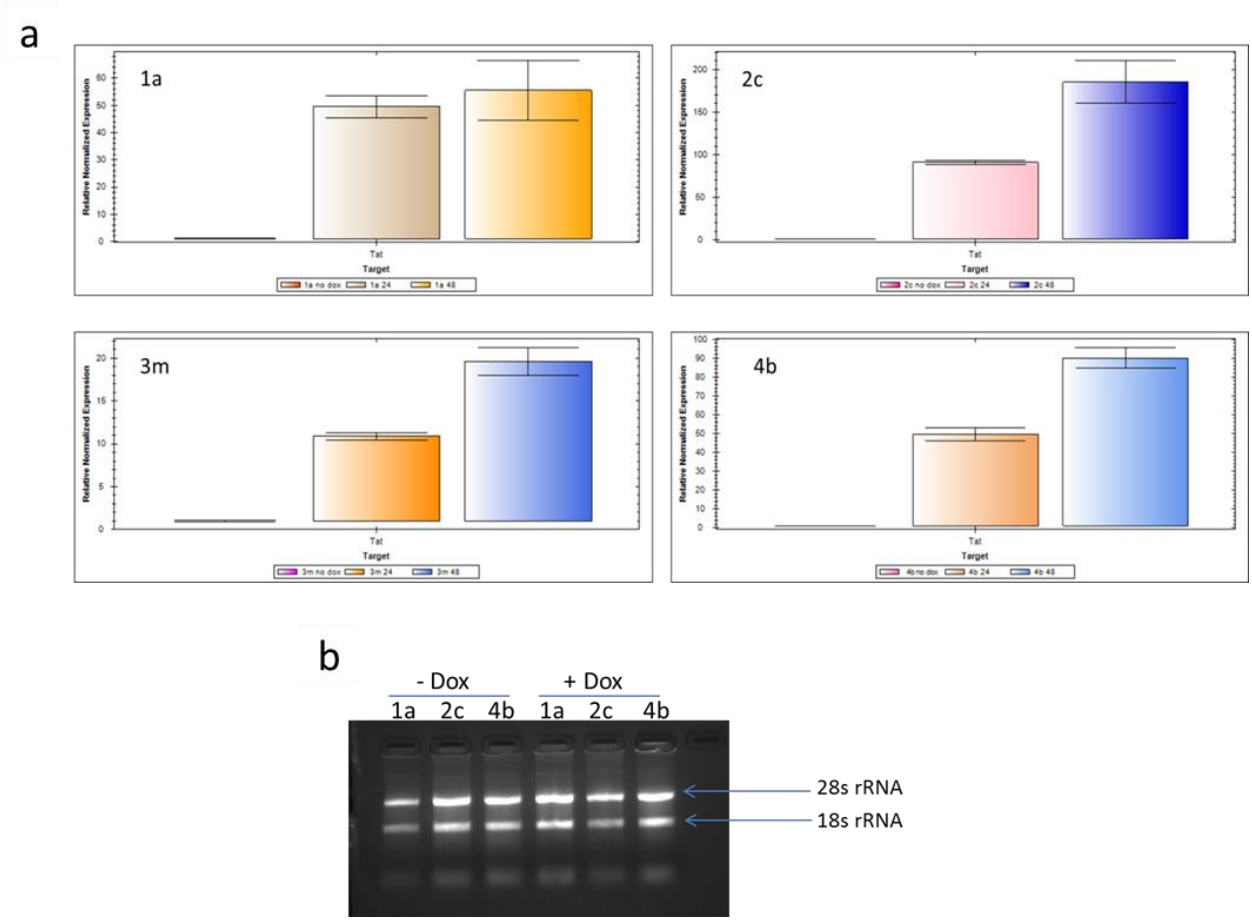


Fig 11: Low CFP clone screening based on relative fold induction of Tat. (a) Tat induction in 1a is 60 fold, 2c is 200 fold, 3m is 20 fold and 4b is 100 fold. Bars on the plots read; from left to right; no dox (untreated), dox 24 (cultured in Dox containing media for 24 hours) and dox 48 (cultured in Dox containing media for 48 hours). (b) 1.5% agarose gel made in DEPC treated water showing RNA extracted from the mentioned clones after 24 hours of treatment with Dox. 1 µg RNA loaded on the gel.

iv. Dox dependent GFP expression of low CFP clones

While the three Parental II clonal lines 1a, 2c and 4b were being tested for their relative Tat induction, another experiment was set up to confirm if Tat is functional in these clonal lines. 5 times diluted cLG virus having HIV1-B subtype LTR was added to each clone in the presence or absence of 1000 ng/ml Dox. Virus was removed from the medium after 6 hours and the media with or without Dox replenished. Taking this as the first time point, the clones were analyzed using FACS every 48 hours to check the percentage of CFP^+GFP^+ cells. All clones show 1-2% CFP^+GFP^+ cells while 3m shows a much lesser percentage.

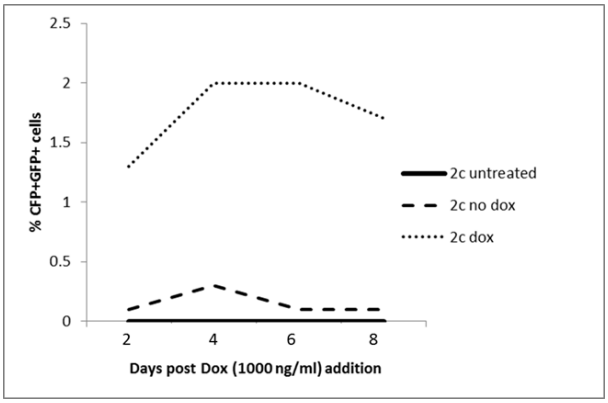


Fig 12: Dox dependent GFP expression as an indicator of Tat functionality. cLG OHHC virus was added to both '2c dox' and '2c no dox' and not added to '2c untreated'.

v. Dox dose dependence of 2c

1 x 10^6 low CFP 2c cells were taken per well in a 6 well plate for this assay. Other conditions were kept same as that of dose dependence of 3m. Varying doses of Dox were added to the wells to get the final concentrations of 0, 500, 750 and 1000 ng/ml.

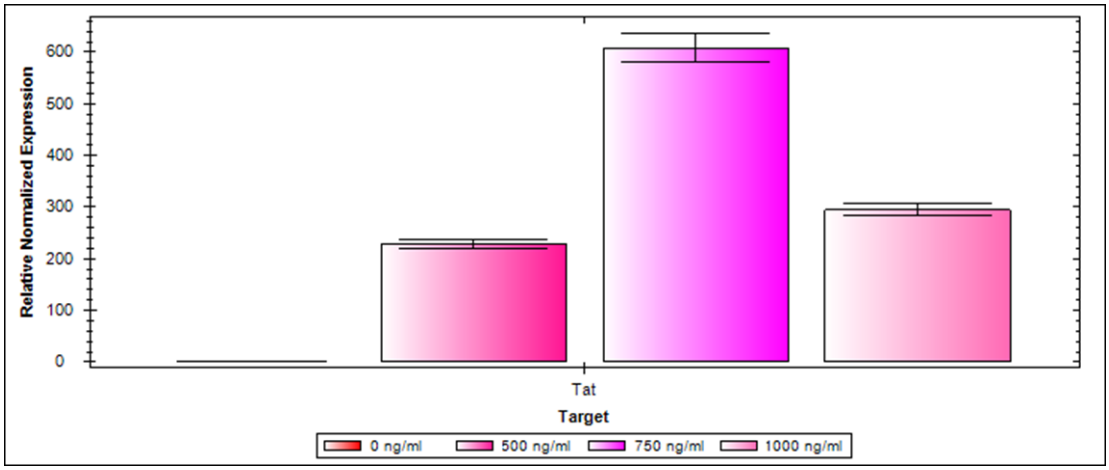


Fig 13: Dox dose dependent Tat expression in 2c.

The advantage of the current Tat inducible latency model from the rest is that, we can study κB site dependent viral latency independent of Tat as well as under normalized Tat expression. Tat expression can be modulated by changing the concentration of doxycycline added. Studies have shown Tat to control latency as an intrinsic viral parameter irrespective of the activation status of the cell. In a previous study from our lab, we have observed NFκB site numbers to directly influence latency establishment in a directly proportional manner. However, this was a cumulative effect of κB site number difference as well as Tat co-expression. We have also observed that Tat inhibits stable latency maintenance in some of the variant viruses. To rule out any influence of Tat on latency establishment or maintenance and obtain a cause and effect relationship between κB site number and viral latency, we intended to remove Tat from the circuit.

To validate the model, we did dox dependence and pulse chase studies on clone 3m. Both these experiments show that clone 3m does not give expected Tat induction (Fig 10 and Table 2). To select a Parental line II clone that gives a better Tat induction, we screened 3 more clones; 1a, 2c and 4b. Screening was done in two independent ways; by dox dependent GFP expression as measured using a flow sorter and an analysis wherein the Dox concentration was kept constant for all clones and relative Tat expression was measured using real time PCR. Both the experiments suggest that clone 2c gives the expected Tat induction.

Dox dose dependence study was then done on clone 2c and we see that between 500 ng/ml and 1000 ng/ml addition of Dox, Tat expression peaks at 600 fold induction at 750 ng/ml Dox and then drops down (Fig 13). While Tat drives the transcription of the virus and is essential for the viral replication, it is known to be toxic for the cell in excessive quantities. Hence, for future studies, 500 ng/ml Dox is selected.

When Tat was not controlled extrinsically, we have seen that the virus with 3 and 4 functional κB sites (OHHC and FHHC) achieve latency much faster than their null counterpart (OOOO) which remains more or less unaltered (Fig 4). Since we have eliminated the Tat positive feedback loop in this model, we expect to see a similar profile with the difference that the percentage of GFP positive cells will fall much faster.

The next step to this model will be infecting the selected low CFP clone (2c) with isogenic κB variant pseudoviruses and addressing questions related to viral latency. This model will enable us to understand the influence of three independent parameters i) cell activation status ii) κB site differences and iii) Tat on viral latency either in isolation or in synergy.

Bibliography

Bachu, M., Yalla, S., Asokan, M., Verma, A., Neogi, U., Sharma, S., Murali, R. V, Mukthey, A.B., Bhatt, R., Chatterjee, S., et al. (2012). Multiple NFκB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J. Biol. Chem. *287*, 44714–44735.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80-.). *220*, 868–871.

Blankson, J.N., Persaud, D., and Siliciano, R.F. (2002). THE CHALLENGE OF VIRAL RESERVOIRS IN HIV-1 INFECTION. Annu. Rev. Med. *53*, 557–593.

Burnett, J.C., Miller-Jensen, K., Shah, P.S., Arkin, A.P., and Schaffer, D. V (2009). Control of stochastic gene expression by host factors at the HIV promoter. PLoS Pathog. *5*, e1000260.

Carter, C.C., Onafuwa-Nuga, A., McNamara, L.A., Riddell, J. 4th, Bixby, D., Savona, M.R., and Collins, K.L. (2010). HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs. Nat. Med. *16*, 446–451.

Chun, T.W., Finzi, D., Margolick, J., Chadwick, K., Schwartz, D., and Siliciano, R.F. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nat. Med. *1*, 1284–1290.

Colin, L., and Lint, C. Van (2009). Molecular control of HIV-1 postintegration latency : implications for the development of new therapeutic strategies. *29*, 1–29.

Dhar, D.V., Amit, P., and Kumar, M.S. (2012). In-Silico Identification of New Genes in HIV-1 by ORF Prediction Method. *1*, 52–54.

G. Hunt, C.T. (2000). Occurrence of Additional NF- k B-Binding Motifs in the Long Terminal Repeat Region of South African HIV Type 1 Subtype C Isolates. *16*, 305–306.

Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., et al. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science *220*, 865–867.

Han, Y., Wind-Rotolo, M., Yang, H.-C., Siliciano, J.D., and Siliciano, R.F. (2007). Experimental approaches to the study of HIV-1 latency. Nat. Rev. Microbiol. *5*, 95–106.

Igarashi, T., Brown, C.R., Endo, Y., Buckler-White, A., Plishka, R., Bischofberger, N., Hirsch, V., and Martin, M.A. (2001). Macrophage are the principal reservoir and sustain

high virus loads in rhesus macaques after the depletion of CD4+ T cells by a highly pathogenic simian immunodeficiency virus/HIV type 1 chimera (SHIV): Implications for HIV-1 infections of humans. Proc. Natl. Acad. Sci. U. S. A. *98*, 658–663.

Mandell, G.L., Bennett, J.E., and Dolin, R. (2010). Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases: Contiene (Churchill Livingstone Elsevier).

Pace, M.J., Agosto, L., Graf, E.H., and O'Doherty, U. (2011). HIV reservoirs and latency models. Virology *411*, 344–354.

Razooky, B.S., Pai, A., Aull, K., Rouzine, I.M., and Weinberger, L.S. (2015). A Hardwired HIV Latency Program. Cell *160*, 990–1001.

Santos, A.F., and Soares, M. a (2010). HIV Genetic Diversity and Drug Resistance. Viruses *2*, 503–531.