A stronger transcription regulatory circuit of HIV-1C drives rapid establishment of latency with implications for the direct involvement of Tat

Sutanuka Chakraborty[1,2] , Manisha Kabi [1,3] and Udaykumar Ranga [1,#]

[1] Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur P. O., Bangalore 560064, India

[2] Present address: Chemical Engineering Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India

[3] Present address: Genome Architecture, Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain

Address of correspondence to: Prof. Udaykumar Ranga at Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur P. O., Bangalore 560064, India. E-mail: udaykumar@jncasr.ac.in

Keywords
gene expression, HIV-1C, latency, LTR, NF-κB, positive feedback, reporter virus, Tat, transcription factor binding sites, transcriptional silence, viral promoter

Abstract

The magnitude of the transcription factor binding site variation emerging in HIV-1C, especially the addition of NF-κB motifs by sequence duplication, makes the examination of transcriptional silence challenging. How can HIV-1 establish and maintain latency despite having a strong LTR? We constructed panels of sub-genomic reporter viral vectors with varying copy numbers of NF-κB motif (0 to 4 copies) and examined the profile of latency establishment in Jurkat cells. We found surprisingly that the stronger the viral promoter, the faster the latency establishment. Importantly, at the time of commitment to latency and subsequent points, Tat levels in the cell were not limiting. Using highly sensitive strategies, we demonstrate the presence of Tat in the latent cell, recruited to the latent LTR. Our data allude, for the first time, to Tat establishing a negative feedback loop during the late phases of viral infection, leading to the rapid silencing of the viral promoter.

Introduction

The post-integration HIV-1 latency is characterized by the presence of the transcriptionally silent but replication-competent provirus within the host cells, a serious challenge for HIV-1 eradication. A significant amount of controversy surrounds HIV-1 latency following the discovery of a latent HIV-1 reservoir in the resting $CD4^{+ve}$ T-cells (Chun TW et al., 1995, Finzi D et al., 1999) whether the external host cell parameters or the intrinsic proviral elements are deterministic in modulating HIV-1 latency. One school of thought considers the establishment of HIV-1 latency to be an 'epiphenomenon' or a rare side-effect that manifests during the transition of the infected and activated $CD4^{+ve}$ T-cells to the resting, memory phenotype (Siliciano RF and Greene WC., 2011; Eisele E and Siliciano RF., 2012). The 'epiphenomenon' hypothesis being deterministic suggests that extraneous host parameters such as the activation status of the cell, site of integration, transcriptional interference, and host epigenetic modifications play a critical role in HIV-1 latency (reviewed in Van Lint C et al., 2013; Archin NM et al., 2014). An alternate stochastic hypothesis has been proposed to explain the decision making in HIV-1 latency, based on experimental observations when uniform activation stimuli fail to reactivate latent provirus synchronously from all the patient-derived $CD4^{+ve}$ lymphocytes (Ho YC et al., 2013; Weinberger AD and Weinberger LS., 2013). The stochastic hypothesis proposes that the LTR-Tat positive feedback circuit functions as the master regulator to regulate viral latency, although the influence of the host environmental factors is not disregarded (Weinberger LS et al., 2005; Weinberger LS et al 2008., Razooky et al., 2015).

Transcriptional master circuits have been identified in prokaryotic and eukaryotic viruses functioning as integral molecular switches to toggle between transcriptional ON and OFF phases. In this context, the lysis-lysogeny decision in bacteriophage λ has been widely researched. In λ phage, fate selection between the lytic mode of replication or a lysogeny state was determined by the preferential expression of two different key viral proteins- the CI or the λ repressor and the Cro or the lytic activator, from a bi-directional promoter. The phenotypic switch in λ phage is characterized by two salient molecular features. First, the Cro and CI proteins form a mutual repression circuit; each of the two proteins mutually inhibits the expression of the other while activating its own synthesis, suggesting the involvement of both positive and negative feedback circuits (Arkin A et al., 1998). Second, the cooperativity of the CI repressor to form a protein octamer capable of binding the promoter-operator region is critical for the λ phage to establish a 'bistable' circuit manifesting lysis or lysogeny (Dodd IB et al., 2001). Transcriptional circuits in several latency-establishing eukaryotic viruses function through rate-versus-level trade-off where rapid up-regulation of a viral protein is essential for efficient viral replication, but the same molecule is cytotoxic at saturating levels. The immediate-early 2 (IE2) transactivator protein of CMV is a typical example of this phenomenon (Dwarakanath RS et al., 2001; Sanders RL et al., 2008; Stinski MF et al., 2008). The ICP0 and Rta proteins of Epstein Barr virus (EBV) and Herpes Simplex Virus-1 (HSV-1), respectively, exploit the phenomenon of cooperativity to control alternate replication fates (Ragoczy T et al., 2001; Sarisky RT et al., 1996; Cai W et al., 1993; Roizman B et al., 2005; Kent JR et al., 2003).

In the recent past, several groups have attempted to develop theoretical models to explain the Tat-feedback mediated latency decision in HIV-1. Weinberger's group, using a sub-genomic HIV-1 vector LTR-GFP-IRES-Tat (LGIT) in conjunction with in silico analyses demonstrated for the first time that the Tat-feedback circuit amplifies a huge thermal fluctuation (genetic noise) of Tat molecules during the initial stages of viral transcription that finally drives the phenotypic variability in the virus-productive vs. latent phenotypes (Weinberger LS et al., 2005). Importantly, the transcription regulatory circuit in HIV-1 appears to differ significantly from those of the λ phage or the eukaryotic viruses mentioned above. First, there is no evidence for a repressor molecule or negative feedback loop controlling the HIV-1 circuit while Tat functioning as only an integral component of a positive feedback circuit. Second, the Tat-feedback circuit seems to lack bistability such that Tat transactivates the LTR as a monomer

with no self-cooperativity to form multimers (H=1; Razooky BS et al., 2011). Tat was proposed to undergo post-translational modifications at specific sites to modulate latency and to account for Tat mono-stability, and the lack of self-cooperativity. It was proposed that enzymatic conversion of acetylated Tat (Tat_A) to a more stable deacetylated form (Tat_D) constitutes a feedback-resistor module with a predominant off state (Weinberger LS and Shenk T., 2006). Importantly, the models described above are majorly based on mathematical simulations supported by simple reaction parameters with minimal experimental validation. Of note, these studies exclusively modeled the HIV-1 subtype B system, although other genetic families of HIV-1 contain subtype-specific molecular features.

Collectively, the 'viral circuitry' hypothesis, using the HIV-1B model, could appreciate and integrate two paradoxical phenotypic outcomes. Firstly, the attribution of the Tat-feedback loop as the master regulatory circuit explained the resistance of viral gene expression to the environmental stimuli. Secondly, the robustness of the LTR-Tat positive feedback, permitted the highly sensitive viral response to minute fluctuations in the levels of Tat molecules, thereby allowing a switch between the active and latent states. Two distinct components that are the central components of the viral regulatory circuit function together to account for the above paradox- the weak and bursty promoter (5' LTR) and the strong Tat-positive feedback (reviewed in Pai A et al., 2017). While the LTR responds weakly to the extracellular cues, the Tat-feedback amplifies the weak response manifold. For instance, TNFα, a potent inducer of the NF-κB pathway can enhance the basal level promoter activity by only two folds, the positive Tat-feedback transactivates the LTR by >50 folds (Weinberger LS et al., 2005; Weinberger LS et al., 2008; Razooky BS et al., 2011; Karn J et al., 2012).

The examination of transcriptional silence is expected to be technically more challenging in HIV-1C as compared to that in the other subtypes of HIV-1, including HIV-1B, for two different reasons. First, HIV-1C contains several subtype-specific variations in nearly all types of TFBS present in the viral promoter, including that of NF-κB, Sp1, RBF-2, and other elements. Among these variations, the copy number difference of the NF-κB binding elements is the most striking one and unique to HIV-1C. While other subtypes of HIV-1 contain a single (HIV-1E) or two (all others including HIV-1B) NF-κB motifs in the viral enhancer, HIV-1C contains three or four of these motifs (Figure 1). Second, the additional NF-κB binding elements present in HIV-1C are genetically diverse (Bachu M et al., 2012a). Three different

kinds of NF-κB motifs, H, C, and F, may be found in the long-terminal repeat of HIV-1C (C-LTR). We demonstrated previously that the progressive acquisition of the additional NF-κB motifs enhances transcriptional strength of the viral promoter in HIV-1C and confers replication superiority over the canonical viral strains in natural infection, and under experimental conditions (Bachu M et al., 2012b). Given the positive correlation between the transcriptional strength of the viral promoter and the enhanced strength of transcriptional feedback, it is intriguing how viral latency is favoured in the variant viral strains containing a higher number of NF-κB motifs. In this backdrop, the present study is an attempt to examine the influence of variation in the number of NF-κB binding elements in C-LTR. Of note, the focus of the present study is only on the copy-number difference of the NF-κB binding sites, therefore, on the overall strength of transcription, and its influence on viral latency. The present study does not aim at examining the impact of genetic diversity of NF-κB binding motifs on viral latency.

Using sub-genomic HIV-1C reporter viruses that differed in the LTR-Tat transcriptional feedback architecture and using panels of LTR variant viral strains that varied in the copy-number of NF-κB motifs, we demonstrate for the first time that the enhanced transcriptional strength of the LTR leads to a rapid establishment of viral latency. Further, we explain the apparent paradox by demonstrating that a stronger transcriptional activity of the LTR leads to higher levels of cellular Tat protein which, above a certain threshold, possibly establishes negative feedback on viral transcription. Importantly, using indirect immunofluorescence and a highly sensitive proximity ligation assay, for the first time, we demonstrate the presence of Tat in cells harboring an active or a latent provirus. We also show the recruitment of Tat not only to the active but also the latent proviral LTR, albeit at a magnitude several folds lower. Our data, thus collectively allude to Tat playing a deterministic role in initiating transcriptional silence through negative feedback regulation.

RESULTS

The strength of Tat-transactivation is directly proportional to the number of functional NF-κB binding sites in the viral enhancer.

The present study is an attempt to examine the influence of variation in the number of NF-κB binding elements in the C-LTR. To this end, we employed two different Jurkat T cell models,

the autonomous Tat-feedback (ATF), and the tunable Tat-feedback (TTF) models to examine HIV-1C latency. The sub-genomic viral vectors encoding EGFP (or d2EGFP, a variant GFP with a shorter half-life), were pseudotyped with VSV-G envelope. The two experimental models differ from each other in the manner the LTR-Tat-feedback axis is regulated. Using these two experimental systems, we examined transcriptional activation and silencing as a function of the promoter strength (by varying the copies of NF-κB motifs ranging from 0 to 4 copies in the ATF model) or the Tat-feedback strength (by modulating the physiological concentration of Tat in the TTF model) or both from a panel of HIV-1C LTRs.

The 'Autonomous Tat-feedback' (ATF) model of HIV-1 comprises of the presence of only the LTR and Tat, with all the other viral factors being absent, thus retaining the natural functional association between the two major viral factors, as reported previously (Weinberger LS et al., 2005; Burnett JC et al., 2009). Several groups have adopted the ATF model to elucidate the mechanisms governing HIV latency. In the present study, we modified the LGIT sub-genomic reporter vector (LGIT in Weinberger LS et al., 2005) by substituting the Tat ORF and the 3' LTR of the parental pLGIT vector, both of HIV-1B origin, with the homologs of HIV-1C to construct cLGIT. In the pcLGIT (cLTR-EGFP-IRES-cTat) vector, the expression of EGFP and C-Tat are under the control of the C-LTR.

Given the natural propensity of HIV-1C to contain more copies of the NF-κB motif in the enhancer, three copies typically and up to four copies frequently (Bachu M et al., 2012a), we constructed a panel of cLGIT viral strains comprising of NF-κB copy-number variant LTRs (p911a and p911b series; Materials and Methods). Using the prototype C-LTR containing four functional NF-κB binding sites (FHHC), we introduced inactivating point mutations sequentially into the enhancer to reduce the number of functional NF-κB motifs progressively, from 4 copies to 0 copies (Figure 2A, left panel). The viral stocks of the panel pseudotyped with VSV-G envelope were generated in HEK293T cells, and the relative infectious units (RIU) of the stocks were determined in Jurkat cells using GFP fluorescence.

Next, Jurkat cells were infected with each viral strain of the panel independently at a ~0.5RIU; three days following infection, the cells were activated using a cocktail of global T-cell activators (40 ng/ml PMA + 40 ng/ml TNFα + 200 nM TSA + 2.5 mM HMBA); and 24 hours following activation, both GFP fluorescence, as well as Tat transcript levels of the control and activated cells, were determined using flow cytometry and Tat RT-PCR, respectively. We found the GFP MFI to be proportional to the number of functional NF-κB motifs in the LTR

(Figure 2A, central panel), although the percent of viral infectivity was comparable (Figure S1A). The LTR containing four NF-κB motifs (FHHC; 4-κB) demonstrated the highest fluorescence intensity with (82,917.51 ± 825.7 RFU) and without (12,365.13 ± 179.3 RFU) activation; while, the LTR in which all the four NF-κB motifs have been mutated (OOOO; 0-κB) demonstrated the lowest levels of the reporter expression with (22,190.38 ± 668.1 RFU) and without (6,083.36 ± 290.5 RFU) activation. The activity of the other three LTRs containing 3 (OHHC; 3-κB), 2 (OOHC; 2-κB), or 1 (OOOC; 1-κB) functional NF-κB motifs remained between the two extremes. The fold transactivation was directly proportional to the number of functional NF-κB motifs in the LTR with a linear correlation (r = 0.98) between the transcriptional activity and the functional NF-κB motifs in the LTR (Figure S1B). Similar to the GFP MFI profile, the level of Tat transcript expression (Figure 2A, right panel) and fold transactivation (Figure S1C) were directly proportional to the number of NF-κB copies in the LTR with or without activation. It is evident from the expression profile that a perfect correlation exists between the number of NF-κB motifs and the level of gene expression from the promoter. Importantly, the expression of EGFP can be used as a surrogate marker for the expression of Tat, since a perfect correlation exists between the two genes co-expressed from the viral promoter. In the subsequent assays, we routinely used the expression of EGFP as a measure of the transcriptional activity of the viral promoter with frequent confirmation of Tat expression.

Importantly, when the cell populations were divided into three categories based on the intensity of EGFP expression (GFP^{-ve}, GFP^{Low}, and GFP^{High}), it was the GFP^{High} fraction that displayed the most pronounced impact of the NF-κB site copy number difference on transactivation. The percentage of the GFP^{High} fraction on Day-4 following cell infection was directly proportional to the number of NF-κB motifs in the LTR, which was also reflected in the peak height of the GFP^{High} cluster in the stacked histogram profile (Figure 2B).

A stronger viral promoter establishes latency at a faster rate.

A major paradox in the transcriptional regulation of HIV-1C is that a virus that must establish latency tends to acquire a stronger promoter containing more NF-κB motifs, especially when other genetic families of HIV-1 do not employ such a strategy. To understand this paradox, we used the NF-κB copy number variant strains of the ATF panel to determine the kinetics of latency establishment. Using the experimental strategy depicted (Figure 2C, top panel), we

infected Jurkat cells at a low RIU (~0.1-0.2) to ensure a single integration event per cell. The cells were allowed to expand before inducing them with a cocktail of global activators, and the $EGFP^{+ve}$ cells were recovered by sorting. The kinetics of EGFP switch-off was subsequently monitored every four days for 16 days by flow cytometry. A representative strategy of cell gating and sorting is presented (Figure S2).

Using the experimental strategy described above, and the panel of NF-κB copy number variant viral strains containing 4 to 0 copies of the TFBS, we evaluated how the transcriptional strength of HIV-1 LTR would influence the kinetics of latency establishment over 16 days. The analysis found a profound impact of NF-κB motif copy number on the kinetics of HIV-1 latency establishment. Although latency establishment was evident for all the five LTRs examined, the rapidity of latency establishment unexpectedly was directly proportional to the number of NF-κB motifs in the viral enhancer (Figure 2C, panels: Total GFP^{+ve} MFI and Total GFP %). In other words, the stronger was the transcriptional activity of the LTR, the faster the latency was establishment. Based on the slope of GFP downregulation, the LTRs could be classified into two broader groups: The two strong promoters, the 3- and 4-κB LTRs, down-regulated the EGFP expression at a significantly faster rate than the other three not strong promoters 2-, 1- and 0-κB LTRs. In the present manuscript, we classify the LTRs into two groups, 'strong' and 'weak' based on the difference in the transcriptional strength, a categorization consistent with many other properties we analysed subsequently, although the 2-κB LTR occasionally occupied an intermediate position (see below). For instance, the EGFP intensity values (RFU) of 4-κB LTR reduced approximately 8-fold from a value of 30,631.64 ± 1,278.3 on D0 to 3,771.06 ± 245.2 on D16, whereas the corresponding values for the weakest 0-κB LTR were the modest and reduced by only two-folds during the same period from 4,455.11 ± 258.9 to 2,371.98 ± 59.3. Of note, although both 3- and 4-κB LTRs demonstrated a rapid GFP downregulation, the 3-κB LTR established viral latency at a faster rate, and the difference between the two promoters was highly reproducible and significant. It is not clear if this difference may have implications for the relative replication fitness of the two viral strains (see discussion).

The expression profile of the Tat transcripts determined using an RT-PCR on days 0, 8 and, 16 also correlated directly with the NF-κB copy number in the LTR, as expected (Figure 2C) and resembled that of the GFP MFI profile of the LTRs. A profound reduction in Tat expression was observed for all the viral promoters between days 0 and 8. The 4-κB LTR showed the highest level of Tat expression, 92.94 ± 5.4 at D0 that dropped to 12.02 ± 0.8 at D8 and

subsequently to 2.3 ± 0.01 at D16. The corresponding values for the 3-κB LTR are 71.76 ± 2.5, 12.8 ± 0.73, and 1.15 ± 0.1, respectively. Furthermore, using a Taqman qPCR, we confirmed a single integration event per cell in all the five stable cell pools, thus ruling out the possibility that the difference in the integration frequency influenced the outcome of the analyses (Figure S3). In summary, our data are suggestive that the enhanced strength of HIV-1C LTR due to the increase in the number of NF-κB sites could play a decisive role in regulating viral latency. A positive correlation between the Tat-transcript levels and the rapid rate of EGFP switch-off by the strong viral promoters is strongly indicative of the Tat-mediated positive feedback loop playing a critical role in establishing viral latency.

The kinetics of latency establishment is predominantly a function of the EGFP^{High} cells displaying a biphasic mode of transcriptional silence.

At the baseline of the above assay, all the variant viral strains were represented by nearly 100% GFP^{+ve} cells. Surprisingly, however, a marked difference in the intensity of EGFP among the LTR-variants was noted at D0 time point post-sorting (compare the Total GFP^{+ve} MFI and Total GFP^{+ve} % profiles, Figure 2C). This apparent paradox could be explained by analysing only the GFP^{High} cells but not the total GFP^{+ve} population (Figure 2C, bottom-right panel, and S4A). In the present assay, two different cell populations, GFP^{High} (GFP MFI> 10^4 RFU) and GFP^{Low} (GFP MFI ~ 10^2-10^4 RFU), were evident in the histogram profile of each NF-κB variant strain (Figure S4A). Importantly, the reduction in the total GFP MFI (Figure 2C, top-left panel) as well as the Tat-transcript levels (Figure 2C, top-right panel) corresponded perfectly only with the % GFP^{High} cells (Figure 2C, bottom-right panel and compare Figures S4A and B), but not with the % GFP^{Low} cells (Figure 2C, bottom-central panel). Thus, the GFP^{High} cells, not the total GFP^{+ve} cells, are decisive in regulating viral latency. Additionally, the profiles of latency establishment of the strong (3- and 4-κB) versus weak LTRs (0-, 1- and even 2-κB) were profoundly different. Firstly, on day 0, the strong LTRs produced the highest percentage of GFP^{High} cells as compared to the weak LTRs (Figure 2C, bottom-right panel). Secondly, the latency establishment of the strong LTRs appeared to have manifested in two distinct phases: a rapid reduction of EGFP expression between days 0 and 8 and a slower rate of decrease after D8; the bi-phase latency profile was either absent or not prominent with the weak LTRs. Thirdly, the rapid fall in EGFP expression of the GFP^{High} pool of the strong LTRs between days 0 and 8 synchronized with a significant rise in the GFP^{Low} cell pool peaking on

D8 (Figure 2C, bottom-central and -right panels, and S4A). These data collectively allude to the critical role the transcriptional strength of HIV-1 LTR plays in latency establishment. In summary, the GFP^{High} cell pool, not that of the GFP^{Low} cells, plays a decisive role in the population latency kinetics of the virus.

LTR-silencing in the GFP^{High} cells implicates Tat feedback

Given the apparent significance of the GFP^{High} phenotype for HIV-1 latency establishment, we investigated the phenomenon further by sorting only the GFP^{High} cells at a comparable level of fluorescence intensity (Figure 3A and Figure S2). At Day 0, the GFP MFI values were uniform among the variant viral strains of the panel and we monitored downregulation of the green fluorescence every four days for 24 days (Figure S4B; bottom panel). A clear distinction between the strong (4- and 3-κB) and weak (2-, 1-, and 0-κB) LTRs was evident in latency establishment (Figure 3A, left panel) or when the total GFP^{+ve} percentage was considered (Figure 3A, panel two from left), although the 2-κB LTR sometimes occupied an intermediary position. Importantly, the kinetics of latency establishment was explicit when the profile of GFP^{High} percentage was considered (Figure 3A, panel three from left). Nevertheless, the rate of latency establishment was significantly rapid for strong LTRs. The biphasic mode of latency establishment, rather than a gradual and monophasic mode, was evident in the GFP^{High} pool (Figure S4). The biphasic mode was also clearly manifested in a spike in the percentage increase of the GFP^{Low} cell profile on Day 8 (Figure 3A, right panel). The spike was the consequence of the GFP^{High} cells transiting to the GFP^{Low} compartment towards latency. Importantly, the spike of the GFP^{Low} cells on Day 8 was seen only in the case of the two strong LTRs, but not with the three weak LTRs. Of note, unlike the GFP^{High} cells of the strong LTRs, those of the three weak LTRs largely failed to transit into latency even by Day 24 (Figure 3A, right panel), suggesting that the transcriptional strength of the viral promoter is a critical parameter regulating latency establishment. We demonstrated above that a progressively increasing NF-κB site number in the LTR steadily enhances the transcriptional strength as well as the physiological concentration of Tat (Figure 2A, central and right panels). We, therefore, speculate that higher cellular Tat levels, an invariable outcome of the stronger positive transcriptional feedback, are necessary for the rapid silencing of the LTR as manifested by the GFP^{High} cells of the strong LTRs.

Of note, the process of latency establishment above was not complete with any of the LTRs regardless of the transcriptional strength (Figure 3A). The percent of GFP^{+ve} cells reached only the halfway mark after 24 days of sorting even for the strong LTRs that established latency at a faster rate. Importantly, the long half-life of the EGFP, ~ 48 h, used in these vectors as a surrogate marker for latency did not represent the actual dynamics of the LTR transcriptional activity faithfully. The cells were continued to be scored as GFP^{+ve} for a significant period even after the LTR was switched off, leading to a false positive scoring. To rectify this problem, we substituted EGFP in the reporter viral strains with d2EGFP characterized by a significantly shorter half-life (2 vs. 48 h; Li X et al., 1998). The viral strains of the new panel (cLdGIT, p911b series, see Materials and Methods, and Figure S5A) are analogous to the previous panel.

Using the new panel, we sorted the $d2EGFP^{High}$ cells as above to establish the profiles of latency (Figure 3B and see Figure S5B for the sorting strategy and Figure S5D for the stacked histograms). Several differences in the profiles of latency were readily evident between the cLGIT and cLdGIT panels. Unlike the cLGIT panel, the cLdGIT panel successfully established a near-complete viral latency (Figure 3B). All the members of the cLdGIT panel demonstrated latency establishment at a faster rate and the GFP MFI values reduced to the baseline within 96 h following sorting (Figure 3B, left panel). Although the substitution of EGFP with d2EGFP masked the differences in latency kinetics among the members of the cLdGIT panel to some extent, the overall pattern of latency establishment was consistent with that of the cLGIT panel. The percentage of the cells downregulating d2EGFP expression was directly proportional to the number of the NF-κB motifs in the viral promoter (Figure 3B). For instance, the time required for the loss of fluorescence in half of the cells (FL_{50}) was estimated to be 23.3, 22.1, 24.64, 32.9, and 48 h for the 4-, 3-, 2-, 1-, and 0-κB viral strains, respectively. Thus, a direct correlation between the transcriptional strength of the viral promoter and the rate of latency establishment was consistent between the cLdGIT and cLGIT panels (Figures 3A and B). The bi-phasic mode of latency establishment was also evident in the cLdGIT model. The spikes in the GFP^{Low} compartment was seen at 18 h for the two strong LTRs as compared to the weak LTRs where the peaks appeared at a later time point (Figure 3B, right panel) suggesting that strong LTRs established latency at a faster rate.

Collectively, our data are assertive that the transcriptional strength of the HIV-1 promoter is an important regulatory parameter for viral latency. Further, the latency kinetics in the Tat-

transactivated population (GFP^{High} cells), of two different models, cLGIT and cLdGIT vectors of the ATF panel, followed an NF-κB-site copy number-dependent transcriptional silencing.

A bimodal (ON or OFF) latency establishment in the pools of cloned cell lines

The observation that the transcriptional strength of the LTR and the feedback loop of Tat function synergistically to silence the viral promoter was drawn based on cell pools. Since individual cells in a pool are heterogeneous in several biological properties, including the site of proviral integration, we examined the nature of the latency profile in multiple cloned cell lines of all the five LTR variants. Jurkat cells were infected with the viral strains of cLGIT (ATF) panel, stimulated with the global activation cocktail, single GFP^{High} cells were sorted into individual wells of a 96-well culture plate, the sorted cells were allowed to expand for 3-4 weeks, and the EGFP-expression profiles were assessed by flow cytometry (Figure 3C; top-left panel). We recovered 16-25 clones from each NF-κB variant, and 16 clones from each variant were randomly selected for the latency analysis. Of note, since each cell line descended from a single parental cell, all the daughter cells derived from the parental cell are expected to have a common site of integration.

Based on the EGFP expression pattern, the clones could be categorized into three distinct types. The persisters, all the daughter cells descending from a single parental cell sustain expression of high-intensity EGFP throughout the observation period of 28 days and even beyond, comparable to that of the original parental cell, indicative of a provirus transcribing actively in all the daughter cells. The relaxers, all the daughter cells of the GFP^{High} parental cell, have switched-off EGFP expression entirely during the period of observation. And, the bimodallers, the distinctive feature of this clonal type was the simultaneous existence of both the phenotypes among the daughter cells, although all the cells in the cluster were derived from the same GFP^{High} parental cell. One subset of the cells maintained high GFP expression, whereas the other subset down-regulated the reporter gene completely with the minimal manifestation of an intermediate phenotype. Representative fluorescent images of each clonal phenotype are presented (Figure 3C, top-right panel).

Importantly, all the five viral strains of the panel displayed the three clonal phenotypes described above with the distinction that the proportion of the three phenotypes is directly correlated with the copy number of NF-κB sites in the LTR. Given the limitation of available cells for the flow analysis, we could determine the phenotype of the clonal cells only at D21

and D28, not earlier. We analysed 16 randomly selected clones for each of the five LTRs of the panel (Figure 3C, lower panel). The profile of the three phenotypes varied significantly among the members of the panel and appeared to associate with the transcriptional strength of the LTR. On D21, a larger proportion of cell lines representing the strong viral promoters (4- and 3-κB LTRs) transited to the OFF state as compared to those of weak promoters (1- and 0-κB LTRs) while the cells of 2-κB LTR occupied an intermediate position. On D28, the strong viral promoters established latency as compared to the other three. Despite the limitation of the small number of clonal cell lines used in the analysis, these data are broadly consistent with the results of cell pools (Figures 2C, 3A, 3B). Thus, cell pools and clonal cell populations, both the models demonstrated a direct correlation between the transcriptional strength of the LTR and the rate of latency establishment. Further, both the experimental models are also consistent with each other in demonstrating a bimodal, not a gradual, latency establishment.

A tunable regulatory circuit of HIV-1 transcription alludes to the direct role of Tat in latency establishment.

The stronger transcriptional activity of the LTR is expected to lead to a proportionately higher expression of Tat, which in turn should increase the transcriptional activity of the LTR further. As a consequence of the unique arrangement, the two principal regulatory elements collectively modulate viral gene expression. In this backdrop, the profile of latency kinetics observed using the ATF model above cannot be ascribed to the different functional activity of either of the elements alone. It was, therefore, necessary to employ a strategy where Tat transactivation alone becomes a variable factor while the transcriptional strength of the LTR remains constant. To this end, we constructed a new HIV-1-Jurkat cell line model, the 'Tunable Tat-feedback' (TTF) model, where the transactivation strength of Tat can be modulated independently while keeping the transcriptional strength of the LTR constant. Tat in the TTF model was engineered to possess two unique properties as compared to that in the ATF model (Figure 4A, left panel). First, Tat was fused with DsRed2-RFP (stated as RFP throughout the manuscript) to express as a fusion protein enabling the direct visualization of its expression. The new HIV-1 reporter vector, thus, co-expressed two different fluorescent proteins, d2EGFP and Tat-RFP, under the control of the LTR. Second, the Tat-RFP fusion protein was tagged with The C-terminal degradation domain (DD) of FK506 binding protein (Tat:RFP:DD). The DD-tag marks the Tat:RFP:DD fusion protein for rapid degradation through the proteasome pathway (Banaszynski LA et al., 2006). However, supplementing the culture medium with'Shield1', a

small molecule ligand that binds the DD motif, can rescue the protein from rapid processing in a dose-responsive manner. Thus, by changing the concentration of Shield1 in the culture medium, the intracellular concentration of Tat:RFP:DD can be fine-tuned by increasing the stability and thereby the half-life of the protein. The new sub-genomic HIV-1 reporter vector cLdGITRD (cLTR-d2EGFP-IRES-Tat:RFP:DD) permits fine-tuning the intracellular concentration of Tat without altering the transcriptional strength of the LTR.

We constructed a panel (cLdGITRD, the p913 series; Materials and Methods) of two LTR-variant viral strains consisting of 3 or 1 NF-κB motifs, representing the strong and weak LTRs, respectively (Figure 4A, left panel). A direct correlation between the Shield1 concentration in the medium, ranging from 0 to 5 μM, and the intensity of RFP expression was observed in HEK293T cells using the 3-κB viral reporter vector (Figure S6A). Importantly, the viruses could infect the target Jurkat cells, and a direct correlation was also established in the stable Jurkat cells between the Shield1 concentration and GFP fluorescence or Tat:RFP:DD expression (Figure 4A, middle and right panels, respectively) suggesting Shield1-dependent stabilization of the 'Tat:RFP:DD' cassette and the subsequent Tat-mediated LTR transactivation. Importantly, the effect of Shield1 concentration was directly manifested on the GFP^{High} population indicating Shield1 dose-dependent Tat transactivation and also confirming that the GFP^{High} phenotype represented the Tat-transactivated cells (Figure S6C). Of note, although we normalized the viral infection, the % GFP^{+ve} values demonstrated a dose-response proportional to the Shield1 concentration even though the d2EGFP itself does not contain the DD of FKBP (Figure S6B, left panel). The optimal fold activation of the d2EGFP expression (Figure S6B, central panel) and Tat transcript levels (Figure S6B, right panel) were found to be 1 μM and 2.5 μM, respectively. In the subsequent experiments, therefore, we used Shield1 in the range of 0 to 3 μM.

Importantly, the fusion of Tat with DsRed2-RFP offered the advantage of tracking the expression of Tat in real-time during latency establishment. To determine the kinetics of latency establishment in Jurkat cells, we used an experimental schematic as depicted (Figure 4B; top panel). Jurkat cells were infected with 3- or 1-κB viral strain at an RIU of ~0.1-0.2 in the presence of 1 μM Shield1 and expanded for a week in the presence of Shield1. Subsequently, the cells were activated with the global activators for 24 h, the GFP^{High} population (GFP MFI ~10^4 RFU) was sorted, the sorted cells were maintained separately at

four different concentrations of Shield1 (0, 0.5, 1.0 and 3.0 μM), and the levels of d2EGFP and RFP expression were monitored every 24 h by flow cytometry. The representative gating strategy for the GFP^{High} sort is presented (Figure S7A).

The TTF model of latency offered several essential insights (Figures 4B, 4C, and S7B). Importantly, the ability to visualize two different fluorescent proteins (d2EGFP and Tat:RFP:DD) co-expressed under the LTR and the fact that the half-lives of the two proteins were perceptibly different permitted to identify the different stages of the viral gene expression and latency, which we collectively refer to as the viral 'latency cycle' (Figures 4B and C). Although both the fluorescent proteins were expressed under the control of the same viral promoter, the expression of d2EGFP was perceptible earlier and at a higher intensity than that of the Tat:RFP:DD fusion protein. Before sorting, only less than 1% of cells were positive for RFP expression even though 10-20% of cells were GFP^{+ve}. The increased molecular size of the Tat:RFP:DD fusion protein, the slow maturation of DsRed2, and the compromised translation efficiency due to the IRES element, all may have contributed to the observed difference between the d2EGFP and RFP expression profile (Figure S7A). Our analysis, however, was primarily focused on the d2EGFP expression since the GFP^{High} cells directly represent the transcriptionally active and, more specifically, the transactivated viral promoter.

The profile of gene expression through the different phases of the latency cycle is remarkably different between the two viral promoters. The transiting of the cells through the successive phases of the latency cycle is illustrated explicitly when the 3-κB LTR profile is examined (Figure 4C, top panel). At Day 0 following the GFP^{High} sort (see experimental schematic, Figure 4B), the vast majority of cells (92.8%) were GFP^{+ve} Tat-RFP^{-ve} representing a transcriptionally active viral promoter (Figure 4C, Day 0). During the following 24 hours, the GFP^{+ve} Tat-RFP^{-ve} cells exited this compartment via two distinct and diagonally opposite routes (Figure 4C, top panel). While a significant proportion of these cells (approximately 15%) switched off GFP expression to directly return to the GFP^{-ve} Tat-RFP^{-ve} compartment, approximately 6.6% of cells up-regulated Tat-RFP expression from the 3-κB LTR to transit to the GFP^{+ve} $Tat^{-}RFP^{+ve}$ compartment alluding to a strong Tat-dependent transcriptional activity (Figure 4C, top panel; Day 1). At the subsequent time points, GFP^{+ve} Tat-RFP^{-ve} cells continued to vacate this compartment using both the exit routes to reach the GFP^{-ve} Tat-RFP^{-ve} compartment such that on Day 6, 84.3% of the viral strains re-established latency under the

strong viral promoter. Importantly, the cells in the GFP^{+ve} RFP^{+ve} compartment, unlike those of the GFP^{+ve} Tat-RFP^{-ve} compartment, appeared to move to latency only in one direction to the GFP^{-ve} Tat-RFP^{+ve} compartment. The relative proportion of the cells present in the GFP^{-ve} Tat-RFP^{+ve} compartment was significantly higher than that of the GFP^{+ve} Tat-RFP^{+ve} compartment at time points after Day 1 alluding to the unidirectional movement of these cells to latency. Importantly, the GFP^{-ve} Tat-RFP^{+ve} compartment is unique since this quadrant represents the proviruses that have 'recently' switched off transcription, with significant levels of physiological Tat still persistent in the system as indicated by the RFP^{+ve} phenotype. The proviruses of the GFP^{-ve} Tat-RFP^{+ve} compartment also transited to latency only in one direction.

In contrast, the 1-κB LTR predominantly displayed the Tat-independent transactivation (Figure 4C, bottom panel). Although approximately 4% of these cells expressed RFP at a Shield1 concentration of 3 μM, the Tat-RFP expression was delayed by 24 h, as compared to that of the 3-κB LTR, with the Tat-RFP expression reaching a peak only on D3. Importantly, despite the presence of Tat, these dual-positive cells of 1-κB LTR did not move forward to the GFP^{-ve} Tat-RFP^{+ve} compartment, unlike those of 3-κB LTR, but returned to the GFP^{+ve} Tat-RFP^{-ve} quadrant. The proviruses activated by Tat-independent transactivation primarily manifested the GFP^{+ve} Tat-RFP^{-ve} phenotype, and these viruses returned to latency by switching off the d2EGFP expression and typically not inducing Tat-RFP expression. While a large majority of 3-κB LTR viral strains and nearly all the viral strains of 1-κB LTR followed this route, a smaller proportion of proviruses of 3-κB LTR moved forward activated by Tat-dependent transactivation that manifested the GFP^{+ve} Tat-RFP^{+ve} phenotype. Approximately, 14% of the 3-κB LTR viral strains were activated by the Tat-dependent transactivation, and they followed a unidirectional trajectory to latency via the GFP^{+ve} Tat-RFP^{+ve} and GFP^{-ve} Tat-RFP^{+ve} compartments. In contrast, approximately, only 1% of 1-κB LTR viral strains could follow the Tat-dependent transactivation, while the reminder induced only by the Tat-independent activation and returning to latency directly from the GFP^{+ve} Tat-RFP^{-ve} compartment. Thus, the transcriptional strength of the viral promoter appears to play a critical role in not only regulating the activation of viral gene expression but also the latency kinetics and whether or not Tat-dependent transactivation is recruited to the LTR. Only the strong 3-κB LTR, but not the weak 1-κB LTR, could successfully undergo Tat-dependent transactivation. Individual trajectories of the percentages of the four distinct fluorescent populations are presented (Figure 4B; lower panel).

Interestingly, a clear demarcation in the profiles of the weak and strong LTRs is evident at the level of the Tat-independent transactivation – in the RFP negative cell populations (Figure 4B, lower-left and lower-right panels). The GFP^{+ve} $Tat\text{-}RFP^{-ve}$ cells of the 3-κB LTR down-regulated GFP at all the concentrations of Shield1 by D4. In contrast, latency establishment in the same population of the 1-κB LTR was incomplete, and nearly half of these cells remained GFP^{+ve} on D6. This was primarily because a subset of the GFP^{+ve} $Tat\text{-}RFP^{-ve}$ cells at the later time points (D1 and beyond) followed the Tat-dependent route to latency in the case of the strong 3-κB, but not the weak 1-κB promoter. Therefore, from the data of the TTF model, it appears that Tat-dependent transactivation can silence the promoter at a faster rate as compared to that of the Tat-independent pathway. Further, the kinetics of percent GFP^{+ve} to GFP^{-ve} transition, irrespective of the Tat-RFP expression, demonstrated an identical pattern of promoter silencing when compared with the ATF model (Figure S7B; left panel). At all the concentrations of Shield1, the strong 3-κB LTR switched off faster than the weak 1-κB LTR. Thus, the data obtained from the TTF model are strongly suggestive that the transcriptional strength of the HIV-1 LTR plays a critical role in controlling viral latency as a validation of the ATF model. A strong LTR is not only faster in establishing viral latency but also is rapid in revival kinetics from latency, whereas a weak viral promoter appears to be restricted in both the functions.

A sustained presence of Tat in the nucleus following the LTR switch-off

The latency kinetics of two different cellular models (ATF and TTF) alluded to the direct involvement of Tat in the transcriptional suppression of the viral promoter, in a concentration-dependent manner. Furthermore, we could detect the presence of Tat-RFP fusion protein in cells harboring a transcriptionally silent provirus (GFP^{-ve} $Tat\text{-}RFP^{+ve}$) containing a strong viral promoter (3-κB-LTR) (Figures 4B, 4C and S7). Therefore, it was necessary to evaluate the physiological levels and the relative distribution of Tat in cells concomitant with LTR-silencing. To this end, we tracked the expression pattern of the Tat protein in Jurkat cells using indirect immunofluorescence while the cells transited from the 'ON' to the 'OFF' state.

Jurkat cells infected with the J-cLdGIT-3-κB viral strain encoding d2EGFP were monitored at 4 - day intervals up to day 16 for d2EGFP expression using flow cytometry (Figure 5A, left panel). At Day 0, the GFP^{High} cells were sorted and subjected to indirect immunofluorescence for Tat at different points (D0, D4, D8, D12, D14, and D16) (Figures 5A, right panel). A high-

titer rabbit anti-Tat primary antibody (# ab43014, Abcam, Cambridge, UK) was used in the assay.

d2EGFP expression analysis by flow cytometry found a progressive downregulation of the fluorescence, and by D8 and D16, only 6.9% and 1% of the cells, respectively, remained positive (Figure 5A; left panel). The profile of d2EGFP expression of individual cells captured by confocal microscopy was perfectly consistent with that of the flow analysis; and, visible fluorescence could not be detected at D8 and beyond (Figure 5A). Immunofluorescence analysis of individual cells, as the provirus progressed towards the OFF state, permitted the monitoring of the LTR activity regulating d2EGFP expression and that of Tat (Figure 5A). Of note, Tat expression was found in two different compartments of the cells, nuclear and extra-nuclear, the latter mostly localized to the cell membrane. Fluorescent intensities of d2EGFP expression, as well as that of Tat-Alexa 568, were determined independently in the nuclear and the extra-nuclear compartments of 150 individual cells, at all the time points (Figure 5B). The threshold levels of fluorescent protein expression were determined by using uninfected Jurkat cell control for d2EGFP (n = 10) and no-primary antibody control for Tat (n = 10). Importantly, while the fluorescence of d2EGFP reduced progressively with time and fell below the GFP threshold by D12 representing the establishment of latency, the fluorescence of Tat, in either compartment, did not drop below the RFP threshold even at D16. The slopes of reduction in the Tat intensity, especially that of nuclear Tat, appear to be bi-phasic with a sharper fall during the initial phase (D0 to D4). The slopes of reduction of the Tat intensities during the initial phases (D0 to D4) of latency establishment were estimated to be -74.54 ± 16.8 and -37.28 ± 3.2 in the extra-nuclear and nuclear compartments, respectively. At the later time points (D8, D12, and D16), there was only a moderate reduction in the Tat levels in either of the compartments. The data are thus suggestive of a higher level of stability of Tat in the nucleus with possible implications for HIV latency. Importantly, the data of Tat-immunofluorescence are in perfect agreement with the results of the TTF model, where the few GFP^{-ve} $Tat\text{-}RFP^{+ve}$ cells at the later stages of promoter-silencing indicated sustained presence of low-levels of Tat molecules in the LTR-switched OFF cells (Figure 4C; top panel). In summary, immunofluorescence not only detected the presence of Tat in the latently infected cells as late as D16 post-sorting but also demonstrated a rapid loss of Tat from the extra-nuclear compartment while its relative stability in the nucleus.

The in situ proximity ligation assay (PLA) detects the presence of Tat in the latently infected cells.

In immunofluorescence, the over-all intensity of the Tat at D12, D14, and D16 in both the cellular compartments was only marginally above the background level. To increase the sensitivity and detect limited quantities of Tat in the 'LTR OFF' cells, we used the highly sensitive proximity ligation assay (PLA), which conjugates immunostaining with the rolling-circle replication and outperforms the traditional immune assays in sensitivity to detect trace amounts of endogenous proteins (Gustafsdottir SM et al., 2005; Soderberg O et al., 2006). We optimized Tat-PLA in HEK293T cells using a pair of anti-Tat primary antibodies raised in different hosts (rabbit and mouse). Since PLA does not work well in non-adherent cells, and our attempts to adapt the protocol to the Jurkat cells were not successful, we used HEK293/HEK293T cells in this assay. Using sub-genomic viral vectors encoding Tat representing HIV-1B (pLGIT or HIV-1C (pcLGIT), we optimized PLA (Figure S8). Both the Tat proteins could be detected as distinct white dots as opposed to sparse dots in single antibody controls (Figure S8). A dose-response in the intensity of PLA dots and plasmid concentration, as well as a good correlation between the PLA dot number and GFP MFI, are evident (Figure S8; bottom panel).

Using the optimized PLA protocol for Tat, we asked if the Tat protein could be detected in d2EGFP OFF cells. To this end, HEK293 cells stably infected with the 4-κB variant of the ATF-cLdGIT panel were sorted for the GFP^{High} cells. After a week of incubation following the enrichment, approximately 50% of the cells expressed d2EGFP, and the cell pool contained both active (GFP^{+ve}) and latent (GFP^{-ve}) cell clusters. The cells stained with either of the antibodies alone did not show any Tat-specific signals confirming the specificity of the assay (Figure 6A left panel, top two lanes). Tat-specific staining was evident only in the presence of both the antibodies not only in the GFP^{+ve} cells but also in the GFP^{-ve} cells (Figure 6A left panel; bottom two lanes). The mean Tat staining intensity was determined in a total of 85 GFP^{+ve} cells and 119 GFP^{-ve} cells comprising of three independent experiments (Figure 6A; right panel). These values were found to be 2.91 ± 2.5 and 2.34 ± 1.9 for GFP^{+ve} and GFP^{-ve} cells, respectively, although the difference was not significant statistically. The Tat-PLA data in HEK293 cells confirmed the presence of Tat in the latent cells at a concentration comparable to that of active viral transcription.

Differential occupancy of cellular complexes on active and silent LTRs of bimodal clones

Burnett JC et al. examined the differential occupancy of NF-κB factors (p50 and p65) at each of the two identical NF-κB motifs (I and II) in HIV-1B LTR by introducing inactivation mutations into each of these sites individually and the corresponding impact on the transcriptional activity (Burnett JC et al., 2009). A similar examination at the C-LTR has not been performed. A previous report from our laboratory demonstrated that NFAT1, 2, and 5 proteins could be recruited to the C-κB motif, the variant NF-κB motif unique for HIV-1C, with an affinity superior to that of the canonical H-κB site (Verma A et al., 2016). We attempted to compare the identity of the transcription factors and other host factors binding to the viral promoter between the active and suppressed states under identical experimental conditions.

The clonal cell lines that display the bimodal EGFP phenotype offer an excellent experimental model as these clonal lines demonstrate two contrasting phenotypes (GFP^{High} and GFP^{-ve} expression) despite an identical viral genotype, chromatin background and, host-cell activation (Figures 3C, S9, and S10). We selected two clones, 3c and 8c representing the strong 4-κB and 3-κB LTRs, respectively, characterized them, and then subjected them to the ChIP analysis. Importantly, a vast majority of the GFP^{-ve} cells representing the 3- or 4-κB LTRs were fully activated to the GFP^{High} phenotype, 95.6% (Figure S9A; right panel) and 90.4% (Figure S10A; right panel) GFP^{High} cells, respectively, following activation, thus confirming that the LTRs retain a potential for activation. The levels of proviral integration between the two subpopulations (GFP^{High} and GFP^{-ve}) of each bi-modal clone were comparable and close to ~1.0 (Figures S9B and S10B), ruling out the possibility of integration frequency differences underlying the bimodal phenotype. Importantly, the Tat transcript levels in the GFP^{+ve} subfractions of both the clonal cell lines were significantly higher compared to their GFP^{-ve} counterparts, approximately 112 folds for the 4-κB (Figure S9C) and 80 folds for the 3-κB clones (Figure S10C).

Having demonstrated the presence of Tat in the cells containing the latent provirus in both the TTF (Figure 4C), and ATF (Figure 5) models by flow cytometry and confocal imaging, respectively, we, next asked if Tat in these cells is recruited to the latent viral promoter. We sorted the GFP^{High} and GFP^{-ve} cell populations from the two clonal cell populations and using the chromatin immunoprecipitation assay, we examined for the presence of several essential host factors or epigenetic marks (Rel family members- p50, p65; NFAT1 and NFAT2; RNA polymerase Ser2 phosphorylation, and Histone 3 lysine 9 trimethylation), as well as Tat in the

chromatin preparations of the active and latent cells. The ChIP assays was performed by amplifying a 240 bp fragment spanning the enhancer-core promoter region in the LTR using a semi-quantitative PCR (Figures S9D and S10D; left panels) and also using a Taqman probe-based real-time PCR amplifying a 127 bp region spanning the NF-κB and Sp1 sites in the LTR (Figure 6B; bottom panel).

The comparative analysis of the nature of the host factors recruited between the active and latent promoters was highly reproducible and consistent between the 3- and 4-κB LTRs (Figures 6B; lower panels and S9D and S10D; left panels). While the transcription promoting host factors, p65 and NFAT2, and epigenetic marks RNA Pol II S2, were found associated with the active viral promoters at significantly higher levels, the transcription repressive factors, p50, NFAT1, and epigenetic marks H3K9Me3 were preferentially associated with the latent viral promoters. That the p50-p65 heterodimer is transcription-promoting, the presence of a significantly higher concentration of p65 at the active promoter is expected (Barbeau B et al., 1997; Chen-Park FE et al., 2002; Stroud JC et al., 2009). On the other hand, the preferential association of p50 with the latent promoter is suggestive of the formation of the p50 homodimer, a known transcription suppressor (Williams SA et al., 2006). Similarly, our data are also in agreement with the previous reports regarding the transcription suppressive and supportive functions of NFAT1 and NFAT2, respectively (Kinoshita S et al., 1997; Kinoshita S et al., 1998; Macián F et al., 1999).

The most crucial finding of the present study is the detection of the association of the Tat protein with the latent LTR. The Tat protein was found associated with the latent 4-κB and 3-κB promoters at levels 1.7- and 3-folds low, respectively, as compared to their active counterparts. The results were reproducible and consistent between the two strong viral promoters (Figures 6B; lower panels). The data were also consistent between the conventional PCR and the quantitative real-time PCR performed following immunoprecipitation. To the best of our knowledge, the present study is the first one to demonstrate the association of Tat with the latent LTR. The above ChIP data were generated using a commercial rabbit polyclonal anti-Tat antibody (Cat # ab43014, Abcam). The data were also reproducible (Figures S9D and S10D; right panels) when two additional anti-Tat antibodies targeting different epitopes in Tat (Cat # 7377 and # 4374, NIH AIDS reagent program, Maryland, USA) were used in the assay. All the three different anti-Tat antibodies furnished positive ChIP signals for Tat at both the latent viral promoters (3- and 4-κB), over and above the respective IgG-isotype controls.

Discussion

The significance of Tat recruitment to the latent LTR

The primary finding of the present work is the identification of a positive correlation between the transcriptional strength of the LTR and faster latency kinetics via the mediation of proportionately enhanced Tat concentration. Using a reporter viral vector encoding d2EGFP under the LTR and tagging Tat with DsRed2-RFP, we found that at the time of commitment towards latency and at subsequent time points, the intracellular concentration of Tat is not a limiting factor (Figure 5B), thus, ruling out the possibility that the limiting levels of Tat underlie latency establishment. Tat must be present in the nucleus to exert a positive or negative influence on the LTR and also must be recruited to the promoter. Using three different experimental strategies, the flow analysis of the Tat-RFP fusion protein (Figures 4 and S7), indirect immunofluorescence (Figure 5A), and a proximity ligation assay (Figure 6A), we successfully demonstrated the presence of Tat in the nucleus of the latent cell, through the successive stages of latency establishment. The presence of Tat could be detected in both the nuclear and cytoplasmic compartments by confocal microscopy (Figure 5A; right panel). By ChIP, we could ascertain the recruitment of Tat to the latent LTR using three different anti-Tat antibodies, although Tat levels in the latent nuclei were typically inferior to those of the active nuclei (Figures 6B, S9D and S10D). Furthermore, a weak but discernible signal of the Tat-transcripts was evident in the latent fractions of the bimodal clones of both the strong LTRs used in the assay (Figures S9C and S10C). All these data are indicative of Tat playing a direct role in promoting latency. Unfortunately, our attempts at adopting PLA to suspension cells were not successful. How is Tat recruited to the chromatin complex needs to be determined. A few studies previously showed the direct binding of Tat to extrachromosomal HIV-1 promoter and proviral DNA (Southgate CD et al., 1991; Dandekar DH et al., 2004); however, the tethering of Tat to the nascent TAR element, as a part of the paused RNA PolII complex proximal to the latent promoter (Barboric M et al., 2005), is a more likely possibility.

Based on several facts, the master regulator of the virus is well-positioned to be a potential candidate to impose a negative feedback on the LTR, in a temporal fashion; including the absence of a known transcription suppressor encoded by the virus, the ability of Tat to constitute the master regulatory circuit of the virus in combination with the LTR in the absence of other viral factors, the presence of Tat in the latent cell detected reproducibly and also

recruited to the latent promoter, and the identification of a positive correlation between the transcriptional strength of the LTR and the rate of latency establishment.

The negative feedback circuit represents one of the powerful and common strategies to regulate gene expression in biological systems.

Negative feedback circuits can rapidly switch off signalling cascades; therefore, this mode of gene regulation represents the most common strategy biological systems exploit to regulate gene expression (Chatterjee A et al, 2008). Molecules of biological significance, especially those controlling powerful signaling cascades such as cytokines and transcription factors, often attenuate their own production using negative feedback loops. The transcription factor NF-κB that controls the expression of numerous cellular factors regulating a wide variety of cellular processes, down-regulates self-expression by activating the inhibitor protein IκBα (Hoffmann A et al, 2002). Likewise, interleukin-2 (IL-2), the most potent cytokine that regulates T cell viability and proliferation, limits self-production by activating the expression of a FOXP3-mediated negative feedback loop (Smith KA and Popmihajlov Z, 2008). Given that the latency establishment is central for HIV-1 survival towards evading immune surveillance and minimizing cytotoxicity, an active molecular mechanism would be necessary to suppress gene expression from the LTR rapidly. The decision making to achieve such a critical phase of the viral life cycle must be an intrinsic characteristic of the MTRC of the virus and couldn't be left to stochastic phenomena or epiphenomena regulated by cellular events. That the MTRC of HIV-1 comprises of only two elements – the LTR and Tat, and that the latter is the only factor encoded by the virus, Tat is the viral factor best positioned to regulate viral transcription and transcriptional silence both, perhaps at different phases of the viral life cycle following integration. Data presented here using two different latency cell models are not only consistent with this critical biological function ascribed to Tat but also provide additional information on latency. In the present work, we examined the latency profile only in the context of HIV-1C, and its validity must be examined in other genetic families of HIV-1.

In the TTF model, the tagging of Tat with DsRed2-RFP permitted the simultaneous tracking of the LTR activity (d2EGFP expression) as well as monitoring the levels of Tat-expression in a temporal fashion (Figures 4, S6, S7). With the help of the TTF model, we could identify that the strong promoter (3-κB LTR) employs two independent modes of transcriptional silencing to latency - Tat-dependent and Tat-independent. The weak 1-κB LTR, in contrast, commissions

only the Tat-independent pathway to latency. In the Tat-independent route, the single GFP^{+ve} cells (GFP^{+ve}Tat-RFP^{-ve}) of 3-κB LTR returned to latency (GFP^{-ve}Tat-RFP^{-ve}) by directly switching off the expression of GFP (Figure 4C; lower panel). In contrast, in the Tat-dependent mode, the single GFP^{+ve} cells (GFP^{+ve}Tat-RFP^{-ve}) of 3-κB LTR, moved forward to the double-positive phase (GFP^{+ve}Tat-RFP^{+ve}) subsequently transiting to latency via the single positive phase (GFP^{-ve}Tat-RFP^{+ve}) to finally double-negative stage (GFP^{-ve} Tat-RFP^{-ve}; Figure 4C; upper panel). The GFP^{-ve}Tat-RFP^{+ve} cells are of great interest since these cells represent a phase immediately after transcription switch OFF and, therefore, the sustained presence of Tat in a promoter-silent condition, which has not been demonstrated previously. As mentioned above, the presence of Tat in latent cells (the GFP^{-ve}Tat-RFP^{+ve} population) is suggestive of the negative regulatory function of Tat contributing towards latency. The 1-κB promoter failed to take the Tat-dependent route presumably due to the absence of sufficient Tat concentration owing to the feedback strength being suboptimal. Thus, the TTF model directly revealed the significance of the Tat-feedback circuit in driving the Tat-dependent latency. The data of the ATF model are also consistent with Tat negatively regulating transactivation. In the ATF model, of the two panels encoding EGFP or d2EGFP, only the GFP^{High} cells (EGFP MFI $>10^4$ RFU; d2EGFP MFI $> 0.5 \times 10^4$ RFU), but not the GFP^{Low} cells, contributed significantly to both NF-κB dependent transcription as well as latency establishment (Figures 2B, 3, S4 and S5D).

The underlying mechanisms regulating HIV-1 latency remain enigmatic.

There have been several attempts to understand HIV-1 latency as this question contains direct relevance for clinical management and viral purging (Mbonye U and Karn J, 2017). The complexity of HIV-1 latency has led to two distinct schools of thought to explain the phenomenon - the hypothesis of 'epiphenomenon' where the host environmental factors including the epigenetic modifications play the deterministic role (Chun TW et al., 1997; Pierson T et al., 2000), and that of 'viral circuitry' where decision making is hardwired in the intrinsic Tat-LTR regulatory circuit (Razooky BS et al., 2015; Weinberger LS et al., 2005). The two models, which need not necessarily be mutually exclusive, have been supported by considerable experimental evidence but also have specific limitations.

Experimental evidence for epigenetic modifications controlling HIV-1 latency is available from studies using clonal cell populations typically harboring sub-genomic viral reporter vectors (Pearson R et al, 2008). The major limitation of this experimental model is the

prolonged periods required for the cells to establish latency. The majority of individual clonal populations reach 50% of latency on an average in 30 to 80 days, which probably is not representative of the kinetics of natural latency. Of note, the lifespan of productively infected T cells in natural infection is estimated to be quite short an average of only 2.2 days (Perelson AS et al, 1996; Simon V and Ho DD, 2003). Given the cytotoxic properties of the viral products and the immune response, viral gene expression is expected to drive viral evolution towards rapid, not prolonged, latency establishment, in natural infection. Thus, the validity of the prolonged latency profile determined using these experimental models is not clear. Additionally, it is also not understood how epigenetic silencing of an active viral promoter is ever achieved, especially in the presence of abundant quantities of Tat.

The contrasting model explaining HIV-1 latency based on the intrinsic and virus-driven stochastic phenomenon is also supported by compelling experimental evidence (Razooky BS et al., 2015; Weinberger LS et al, 2005). The 'feedback-resistor' module (Weinberger LS and Shenk T., 2006), on which the model is primarily based, considers a single type of chemical modification, acetylation, and deacetylation, of Tat, serving as the 'resistor' or dissipater of the positive transcription loop to ensure a stable latent state. The model doesn't take into account a plethora of other types of PTMs to which Tat is subjected. Whereas di-methylation of the lysine residues at positions 50 and 51 (Van DR et al, 2008) and the arginine residues at positions 52 and 53 (Xie B et al, 2007) can suppress Tat-transactivation, mono-methylation of the lysine residues shows the opposite effect (Pagans S et al, 2010). Importantly, methylated Tat is expected to have enhanced cellular stability with implications for latency (Sivakumaran H et al, 2009). In addition to acetylation, the phosphorylation of multiple serine and threonine residues can cooperatively enhance Tat transactivation (Endo-Munoz L et al, 2005; Ammosova T et al, 2006). Polyubiquitination of Tat can also enhance the stability of Tat, thereby augmenting its transactivation function (Bres V et al, 2003). Apart from the chemical modifications described above influencing Tat function, Tat is also known to be inactivated by the propensity of the protein to make dimer and multimer forms, although experimental evidence is scanty in this regard (Tosi G et al, 2000). Furthermore, the differential forward and reverse reaction kinetics of Tat acetylation have been evaluated only in HeLa cells, but not in cells of physiological relevance to HIV-1 infection (Ott M et al, 2004). This aspect, therefore, needs additional validation, especially in cells of a diverse lineage that serve as a natural target for the viral infection. Extending this work further, analyzing the feedback strength in terms of the noise autocorrelation function, the authors demonstrated that a stronger Tat feedback would

yield transcriptional pulses of longer durations leading to cell lysis while the weaker Tat feedback and the Tat-independent transcription would generate shorter transcriptional pulses leading to latency (Weinberger LS et al., 2008). The model, however, doesn't reconcile to the fact that a small proportion of T cells can still escape cell death following Tat-mediated transcription and establish a viral reservoir (Van Zyl G et al, 2018). In summary, despite significant experimental evidence, the question regarding the critical deterministic factor(s) regulating HIV-1 latency remains unresolved.

Subtype-associated molecular features may offer vital clues to HIV-1 latency.

Although the fundamental constitution of the HIV-1 promoter is highly conserved among the various genetic subtypes of HIV-1, within this underlying structural theme, there exist many subtype-specific molecular features that may modulate gene expression considerably from the viral promoter. Such differences are evident in the copy-number and nucleotide sequences of different TFBS especially those of USF, c-Myb, LEF-1, Ets1, NF-AT, Ap-1, NF-κB, and Sp1 binding sites, and regulatory elements such as the TATA box and the TAR element (Jeeninga RE et al, 2000; Mbondji-Wonje C et al, 2018; Montano MA et al, 1997). The critical roles that most of these regulatory elements play in positively regulating the basal and inducible levels of viral transcription have been well documented (Garcia JA et al, 1987; Pereira LA et al, 2000). Importantly, most of the TFBS, especially the AP-1, USF, NFAT, NF-κB, and Sp-1 motifs also play a critical role in regulating viral latency by recruiting chromatin-modifying complexes and transcription suppressing factors such as the histone deacetylases (HDACs) to the viral promoter (Bosque A and Planelles V, 2009; Chan JK et al, 2011; Colin L et al, 2009; Duverger A et al, 2013; Rohr O et al, 2003).

Of the various TFBS present in the viral promoter, those of NF-κB and Sp-1, both represented by multiple and tandem binding sites in the LTR, play a crucial role in regulating gene expression and latency (Baeuerle PA and Baltimore D, 1989; Doetzlhofer A et al, 1999; Suzuki T et al, 1998; Williams SA et al, 2006). The most striking feature in the HIV-1C LTR is the copy-number difference of NF-κB motifs, the sequence variation of the additional κB motifs (Bachu M et al, 2012b), and the associated sequence variation of the Sp1III site (Verma A et al, 2016). We demonstrated previously that NF-κB site duplication is unique for HIV-1C not recapitulated by any other HIV-1 genetic family. Importantly, in HIV-1C, a unique NF-κB motif (the C-κB element, GGGGCGTTCC) and a genetically distinct and subtype-specific

Sp1III site are located at the core of the promoter replacing the canonical H-κB motif (GGGACTTTCC) present in all the other HIV-1 subtypes at this location, in association with a Sp1III motif that shows subtype-specific variations (Verma A et al, 2016).

HIV-1C, the only subtype to demonstrate NF-κB copy-number variation, could serve as an ideal model to ask whether transcriptional strength can affect viral latency, a property not explored previously. The rapid expansion of the 4-κB viral strains in India in a short period of ten years, from 2% to 25 – 35%, is quite surprising (Bachu M et al., 2012b). It appears perplexing that a subfamily of HIV-1 evolves towards acquiring a stronger promoter by enhancing the copies of NF-κB binding sites, especially when other HIV-1 genetic subtypes do not adopt such an evolutionary strategy. The ATF model we used here demonstrated a perfect positive correlation between the number of NF-κB binding sites in the LTR and the viral transcriptional output in the form of GFP and Tat transcripts (Figures 2A, 2B and S1; r = 0.98 and 0.96 for GFP and Tat mRNA, respectively) suggesting that all the four NF-κB binding sites in the LTR are functional. In this backdrop, it remains intriguing why HIV-1C strains require enhanced transcriptional strength, and despite having a strong promoter, how do these viral strains establish and maintain latency. Importantly, a positive correlation between the transcriptional strength of the LTR and faster kinetics of latency establishment has not been demonstrated previously due probably to the absence of NF-κB copy-number variation in non-HIV-1C subtypes of HIV-1.

Reciprocal binding of host-factors at the active and latent promoters

Gene expression is the outcome of multiple layers of regulatory events consisting of the *cis*-acting TFBS and the *trans*-acting chromatin remodelers, viral factors, especially Tat, epigenetic marks, and a cross-talk between a wide array of proteins that ultimately bring about the diverse phenotypic outcomes. Numerous studies attempted to examine how the nature of the host factor complexes recruited at the LTR plays a critical role in accomplishing the dynamic switching between the active and latent states (Pearson R et al., 2008, Mahmoudi T et al., 2012). In an elegant analysis, Burnett et al. used PheB cells derived from the GFP^{Mid} parental Jurkat cells (analogous to GFP^{Dim} in Weinberger LS et al., 2005 and GFP^{Low} in the present study) and compared the nature of cellular complexes recruited between transcriptionally active and latent cells (Burnett JC et al., 2009). This study demonstrated a non-overlapping function of the two genetically identical NF-κB sites in regulating

transcriptional activation versus suppression. We employed a similar experimental strategy with the exception that we used the $EGFP^{High}$ clonal cell populations (the bimodallers, Figure 3c) that manifested a bimodal phenotype of GFP expression. Subsets of bimodal cell clones differ from each other in the ability to retain GFP expression, although the two daughter subsets are expected to share the same integration site, thus offering a significant technical advantage of normalizing the inherent differences in cellular parameters. A comparative ChIP analysis of GFP^{High} and GFP^{-ve} fractions of two selected clonal populations identified essential differences in the profile of host factor and Tat recruitment to the LTRs.

The preferential binding of p50 and p65 (RelA) at the latent and active promoters ascertained the repressive and inducing functions of p50-p50 homodimer and p50-RelA heterodimer respectively, of HIV-1 transcription (Figures 6B, S9D and S10D) (Stroud JC et al., 2009; Burnett JC et al., 2009; Williams SA et al., 2006). Unlike the NF-κB proteins, the impact of individual NFAT members on HIV-1 latency has not been examined in great detail. To the best of the knowledge, the present study is the first to demonstrate a reciprocal binding pattern of NFAT1 and NFAT2 at the active and latent promoters, respectively, in the context of clonal cells. Since NF-κB and NFAT factors share overlapping sites (Pessler F et al., 2004, Bates DL et al., 2008, Giffin MJ et al., 2003), NFAT may have a significant influence on latency in HIV-1C. Furthermore, the NF-κB sites in the C-LTR (F, H, and C-κB sites) are genetically different, adding to the multitude of possible combinations. Targeted inactivation of each κB site, one at a time, followed by ChIP, may provide meaningful insights into the contribution of each κB sequence to diverse signaling pathways and HIV-1C latency.

The key finding of the present study, however, is the detection of the association of the Tat protein with the latent LTR. The results were highly reproducible and consistent between the two strong viral promoters (Figure 6B). The data were also consistent between the conventional PCR and the quantitative real-time PCR performed following immunoprecipitation (Figures S9D and S10D). The data were reproducible when three different anti-Tat antibodies targeting different epitopes in Tat (Ab43014, Abcam, Cambridge, UK, # 7377 and # 4374, NIH AIDS reagent program, Maryland, USA) were used in the assay. The Tat protein was found associated with the active 4-κB and 3-κB promoters at 1.7- and 3-folds higher, respectively, as compared to their latent counterparts (Figure 6B). To the best of our knowledge, the present study is the

first one to demonstrate the association of Tat with the latent LTR, albeit at a lower intensity as compared to the active promoter.

A Tat-dependent negative feedback mechanism to establish latency?

Based on the present study, we propose a novel model for the transcriptional repression of HIV-1 through a Tat-negative feedback mechanism. The attenuation of Tat-positive feedback signaling has been proposed to cause the LTR silencing, triggered by extracellular cues (deterministic model) or limiting Tat levels probabilistically (stochastic model) (Weinberger LS et al., 2005; Weinberger LS and Shenk T., 2006; Weinberger LS et al., 2008; Burnett JC et al., 2009). In either case, Tat concentration gradually falls below a threshold insufficient for self-renewal or successful transcriptional elongation.

Our data allude to a concentration-dependent inter-conversion of the active form of Tat to a repressive form, the latter competing with the former, strengthening a negative-feedback circuit leading to the rapid silencing of the promoter (Figure 7). We propose that the autonomous Tat-feedback loop initially favors the steady accumulation of Tat molecules to enhance transcription. Subsequently, at a point when Tat intracellular concentration surpasses a specific threshold level, Tat switches to the suppression mode down-regulating transcription depending on differential PTM modifications of Tat itself. The strong promoters (3- and 4-κB LTRs) characterized by a stronger Tat-feedback, can initiate a rapid transcriptional silence as compared to the weak promoters (2-, 1- and 0-κB LTRs). Tat is subjected to multiple post-translational modifications, of which some enhance while the others attenuate the transactivation properties of the viral factor as comprehensively reviewed (Ott M et al, 2011).

Our data raises several important questions related to HIV-1C latency, which were beyond the scope of the present study. Is the LTR of HIV-1C likely to continue to acquire additional copies of NF-κB and/or other transcription factor binding sites to augment transcriptional strength further? Of note, unpublished data from our laboratory (Bhange D et al, 2020) demonstrate a recent trend of emergence of at least 10 different types of TFBS variant HIV strains in India. Further, how the variant NF-κB motifs unique for HIV-1C modulate viral latency? Answers to these questions will shed light on the mechanism of HIV-1 latency and likely to help design novel therapeutic strategies to purge HIV infection.

Materials and Methods

Design and construction of HIV-1C reporter vector panels

Autonomous Tat-feedback (ATF) model: The pLGIT reporter vector (HIV-1 LTR-EGFP-IRES-Tat; Weinberger LS et al., 2005) was a kind gift from Dr. David Schaffer (University of California, USA), in which the two different elements of the vector, the 3'LTR, and Tat, were of HIV-1B origin (NL4-3). We substituted these two elements with analogous counterparts of HIV-1C origin (Indie_C1- Genbank accession number AB023804) and referred to the vector as pcLGIT (cLTR-EGFP-IRES-cTat; Verma A et al., 2016). Using the pcLGIT backbone, we constructed a panel of five reporter vectors containing varying copies of functional NF-κB motifs, ranging from 0 to 4 (the p911a vector series). First, an LTR containing four tandem NF-κB motifs (FHHC-LTR; H- GGGACTTTCC, C- GGGGCGTTCC, F- GGGACTTTCT; variations among the κB-motifs underlined), the sequence configuration adopted from HIV-1C molecular clone BL42-02 (GenBank accession No. HQ202921) was generated in an overlap-PCR using the Indie_C1-LTR as a template. The amplified FHHC-LTR was inserted into pcLGIT vector, substituting the original 3'-LTR. Subsequently, using the overlap-PCR, inactivating point mutations were introduced sequentially into the 'FHHC' (4-κB) LTR, to generate the other members of the panel: OHHC (3-κB), OOHC (2-κB), OOOC (1- κB) and OOOO (0-κB) (Figure 2A). Of note, the inactivation mutations only introduced base substitutions, not deletions, keeping the length of the viral promoter constant among the variant viral vectors. The mutated κB-motif 'O' contains the sequence TCTACTTTTT (underlined bases represent inactivating mutations). All the members of the vector panel are genetically identical except for the differences in the copy number of the functional NF-κB motifs. The variant LTR fragments were cloned directionally between the XhoI and PmeI sites present on the outer primers- N1990 FP (5'-GCGTACCTCGAGTGGAAGGGTTAATTTACTCCAAGAAAAGGC-3') and N1991 RP (5'-TATGTCGTTTAAACCTGCTAGAGATTTTCCACACTACCAAAAGGGTCTGAG-3') thus, substituting the original 3'-LTR of pcLGIT. The 3' LTR sequences of all the panel members were sequence-confirmed, and the expression of EGFP was ascertained in HEK293T cells. A second panel of the five variant viral vectors, analogous to the p911a panel was also constructed using the pcLdGIT backbone (Verma A et al., 2016) where EGFP was substituted with d2EGFP, a variant form of the fluorescent protein characterized by the shorter half-life

(p911b vector series; Figure S5A). To generate the p911b panel, the variant LTRs of the p911a panel were transferred directionally to the pcLdGIT vector between the PmeI and XhoI sites, thus substituting the original 3'-LTR. The expression of d2EGFP from all the vectors of the p911b panel was verified using HEK293T cells.

Tunable Tat-feedback (TTF) model: In the TTF model, HIV-1C LTR regulates the co-expression of d2EGFP and Tat-RFP fusion protein from vector pcLdGITRD (cLTR-d2EGFP-IRES-cTat:RFP:DD). The 5' LTR in the pcLdGITRD vector transcribes a single transcript encoding d2EGFP and a 1,314 bp long fusion cassette separated by an IRES element. The fusion cassette is a combination of three different ORFs- (i) the cTat expression segment (BL4-3, GenBank accession number FJ765005.1) (ii) the ORF of DsRed2-RFP, and (iii) the FKBP destabilization domain (DD; Banaszynski LA et al., 2006). The three components of the 'Tat:RFP:DD' cassette were independently amplified using appropriate templates and primers, and, finally, using an overlap PCR, the fusion ORF was generated. The Tat ORF from the pcLdGIT-3-κB vector (p911b series) was replaced with the 'Tat:RFP:DD' ORF, thus, generating the pcLdGITRD-3-κB viral vector. pcLdGITRD-3-κB was subsequently used as the parental vector to construct the other member- pcLdGITRD-1κB of the panel p913 (Figure 4A) by cloning the respective 3'LTRs between PmeI and XhoI in the pcLdGITRD backbone. The d2EGFP expression from the two members of the panel p913 was confirmed in HEK293T cells.

Cell culture

Jurkat cells were maintained in RPMI 1640 medium (R4130, Sigma-Aldrich, St. Louis, USA) supplemented with 10% fetal bovine serum (RM10435, HiMedia Laboratories, Mumbai, India), 2 mM glutamine (G8540, Sigma-Aldrich), 100 units/ml penicillin G (P3032, Sigma-Aldrich) and 100 g/ml streptomycin (S9137, Sigma-Aldrich). The human embryonic kidney cell lines HEK293 and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (D1152, Sigma-Aldrich) supplemented with 10% FBS. All the cells were incubated at 37^0C in the presence of 5% CO_2.

Generation of pseudotyped reporter virus and the estimation of relative infectious units (RIU)

Pseudotyped reporter viruses were generated in HEK293T cells. Each viral vector was transfected together with the 3rd generation lentiviral packaging vectors using the standard calcium phosphate protocol (Jordan M et al., 1996). Briefly, a plasmid DNA cocktail consisting of 10 μg of individual viral vector (NF-κB motif variants), 5 μg psPAX2 (#11348; NIH AIDS reagent program, Maryland, USA), 3.5 μg pHEF-VSVG (#4693; NIH AIDS Reagent program) and 1.5 μg pCMV-rev (#1443; NIH AIDS Reagent program) was transfected in a 100 mm dish seeded with HEK293T at 30% cell confluence. pCMV-RFP (0.2 μg) was used as an internal control for transfection. Six hours post-transfection, the medium was replenished with complete DMEM. Culture supernatants were harvested at 48 h post-transfection, filtered using 0.22 μ filter and stored in 1 ml aliquots in a deep freezer for future use.

The RIU was quantified in Jurkat T-cells by measuring GFP (EGFP or d2EGFP) expression by flow cytometry. Precisely, 3 x 10^4 Jurkat cells in each well of a 12-well tissue culture plate were infected with viral stocks serially diluted 2-fold (from 10 xd to 80 xd) in a total volume of 1 ml of 10% RPMI containing 25 μg/ml of DEAE-Dextran. Six hours post-infection, the cells were washed and replenished with 1 ml of complete RPMI. Post 48 h, the cells were activated with a combination of 40 ng/ml PMA (P8139, Sigma Aldrich), 40 ng/ml TNFα (T0157, Sigma-Aldrich), 200 nM TSA (T8552, Sigma Aldrich) and 2.5 mM HMBA (224235, Sigma-Aldrich) for 18 h, following which the percent GFP^{+ve} cells were analysed using a flow cytometer (BD FACSAria III sorter, BD biosciences, New Jersey, USA). Following this, titration curves were constructed and analysed for 5-10% infectivity of the cells by regression analysis, which would correspond to ~0.05-0.1 RIU. For the TTF model, cells were maintained in 1μM Shield1 throughout the procedure.

Generation of kinetic profiles of latency establishment

Autonomous Tat-feedback (ATF) model: Approximately one million Jurkat cells were independently infected with the NF-κB variant, cLGIT viral strains at low infectivity (RIU ~ 0.1-0.2), in a 35-mm culture-ware containing 1 ml of complete RPMI supplemented with 25 μg/ml of DEAE-Dextran. After 6 h of infection, the cells were washed once with 1X PBS to remove excess DEAE, and the infected cells were maintained in 1 ml of complete RPMI

medium under standard culture conditions for a week. An activation cocktail comprising of 40 ng/ml PMA, 40 ng/ml TNFα, 200 nM TSA, and 2.5 mM HMBA was used to treat the infected cells for 18 h following which either total GFP^{+ve} (GFP-MFI > 10^2 RFU) or GFP^{High} cells (GFP-MFI > 10^2 RFU) were sorted using a FACSAria III sorter (Figure S2). The sorted cells were maintained under standard experimental conditions while a small aliquot was collected every four days to monitor EGFP expression using the FACSAria III flow cytometer, for 16-24 days. Temporal kinetic profiles for % GFP^{+ve} cells and GFP-MFI were constructed and compared among the five NF-κB variants both for the total GFP^{+ve} as well as GFP^{High} cells. We used a similar protocol to generate the kinetic profiles of latency establishment for the cLdGIT NF-κB variant panel with the exception that the analysis of $d2EGFP^{High}$ expression was performed every 24 h following sorting, for 7 days.

Tunable Tat-feedback (TTF) model: Approximately one million Jurkat cells were pre-treated with 1 μM Shield1 and then independently infected with either the 3-κB or the 1-κB cLdGITRD viral variant, at low infectivity (RIU ~ 0.1-0.2) in a 35-mm culture-ware containing 1 ml of complete RPMI medium supplemented with 25 μg/ml of DEAE-Dextran and 1 μM Shield1. After 6 h of infection, the cells were washed once with 1X PBS to remove excess DEAE, and the infected cells were maintained in 1 ml of complete RPMI medium containing 1 μM Shield1 under standard culture conditions for a week. Shield1 was replenished in the medium every 24 h. The infected cells were treated with the activation cocktail described above for 18 h following which GFP^{High} (GFP-MFI > 0.5 X 10^4 RFU) cells were sorted using FACSAria III sorter. The sorted cells were divided into four separate fractions and maintained at four different concentrations of Shield1 (0, 0.5, 1 and 3 μM) under standard experimental conditions, while a small aliquot from all the fractions was collected every 24 h to monitor the expression of d2EGFP using FACSAria III flow cytometer, for 7 days. Temporal kinetic profiles of % GFP^{+ve} cells and GFP-MFI were constructed and compared among the five NF-κB variants at all concentrations of Shield1.

The analysis of the Tat-transcripts in stable Jurkat cells

We quantitated Tat transcript levels using a real-time PCR as a surrogate marker of the transcriptional status of the LTR during latency establishment, or latency reversal. Total mRNA was extracted from 0.5 x 10^6 cells using a single-step RNA isolation reagent- TRI

reagent (T9424, Sigma-Aldrich) at specified time points. Using random hexamer primers, 250-1,000 ng of extracted RNA was converted to cDNA in a reaction volume of 20 µl using the Tetro cDNA synthesis kit (BIO-65043, Bioline, London, UK). The cDNA was then amplified using an SYBR green RT-PCR kit (06924204001, Roche Products, Mumbai, India) for a 139 bp region in Tat exon-1 using the primers- N1783 (5'-GGAATCATCCAGGAAGTCAGCCCGAAAC-3') and N1784 (5'-CTTCGTCGCTGTCTCCGCTTCTTCCTG-3'). The GAPDH RT-PCR was employed as an internal control (primer pair N2232: 5'-GAGCTGAACGGGAAGCTCACTG-3' and N2233: 5'- GCTTCACCACCTTCTTGATGTCA-3'). The relative gene expression was calculated using the ΔΔCt method.

Flow cytometry analysis

All the flow cytometry data were analysed using FCS Express 4 and 6 versions (De Novo Software, Los Angeles, CA).

Indirect immunofluorescence of Tat

Immunofluorescence staining of Tat was performed at multiple time points during the establishment of viral latency in stable J-cLdGIT-3-κB cells characterized by strong GFP fluorescence (MFI range 5 x 10^3 to 50 x 10^3). The sorted GFP^{High} cells were considered as the D0 sample, and Tat-IF was performed subsequently at an interval of every 4 days. Approximately 3 x 10^6 cells were collected in a 1.5 ml vial, washed once with 1X PBS, and fixed with 2% paraformaldehyde in PBS for 10 min at room temperature with mild rocking. Fixed cells were re-washed with 1X PBS followed by permeabilization with 0.2% Triton-X-100 in PBS for 10 min with gentle and intermittent vortexing. Fixed and permeabilized cells were then washed again with 1X PBS and blocked with 4% BSA in PBS for 30 min at room temperature with mild rocking. The blocked cells were incubated with rabbit anti-Tat antibody (ab43014, Abcam, Cambridge, UK) at 1: 250 dilution for 1h at room temperature followed by two PBS washes. This was followed by the incubation with 1: 500 dilution of Goat anti-rabbit Alexa Fluor 568 (A-11010, Molecular Probes, Thermo Fisher Scientific, Massachusetts, USA) for 20 min in the dark at room temperature followed by a PBS wash. The nucleus was stained with 4 µg/ml of DAPI for 20 min in the dark at room temperature. Cells were washed twice

and mounted on coverslips with 70% glycerol for confocal imaging. Images were acquired with a Zeiss LSM 880 confocal laser scanning microscope with Airyscan using a Plan Apochromat X63/1.4- oil immersion objective and analyzed using the ZEN 2.1 software. For imaging single cells, a 4X higher zoom was applied.

The proximity ligation assay

We used an in situ proximity ligation assay (PLA) to detect Tat in HEK293 cells infected with the cLdGIT-4-κB reporter virus. The assay was performed using a commercial kit (Duolink In Situ Red Starter kit Mouse/Rabbit, #DUO92101, Sigma-Aldrich) following the instructions of the manufacturer. Briefly, a heterogeneous population of HEK293 cells harboring both active and latent cLdGIT-4-κB virus, marked by the presence or absence of green fluorescence, respectively, were seeded on glass coverslips and allowed to grow to 60-70% confluence. The evenly distributed cells on the coverslip were fixed with 4% paraformaldehyde for 20 min at room temperature, permeabilized with 0.1% Triton-X-100 for 10 min at room temperature and washed thrice with 1X PBS. This was followed by blocking for one hour using the reagent supplied in the kit. The blocked cells were then treated with the rabbit polyclonal anti-Tat antibody at 1: 250 dilution (Catalog no. ab43014, Abcam) in combination with the mouse monoclonal anti-Tat antibody at 1: 250 dilution (Catalog no. 7377, NIH AIDS reagent program, Maryland, USA). The cells were incubated with a pair of probes (the PLA probe Anti-Mouse MINUS; DUO92004 and PLA probe Anti-Rabbit PLUS; DUO92002) in a 40 µl reaction volume, for one hour at 37^0C followed by washing twice with 500 µl of wash buffer A for 5 min each time. The ligation and amplification reactions were performed as per manufacturer's instructions using the Duolink In Situ Detection reagents Red (Catalog no. DUO92008). The DAPI-supplemented mounting medium (Catalog no. DUO82040, supplied in the PLA kit) was used for mounting the cells. Imaging of the cells was performed using a Zeiss LSM 880 confocal laser scanning microscope with Airyscan fitted with a Plan Apochromat 63X/1.4 oil immersion objective. Signal intensities of the PLA positive spots were quantitated manually using the Image J software.

The primary antibody pair, the rabbit polyclonal anti-Tat (ab43014, Abcam) and mouse monoclonal anti-Tat (7377, AIDS reagents program) antibodies, was validated for Tat specificity before performing PLA in stable HEK293 cells harboring the cLdGIT-4-κB provirus. Approximately 0.5 x 10^6 HEK293 cells in each well of an 8- micro chambered glass

slide (80826, ibidi, Grafelfing, Germany) were transfected with either 800 ng of pcLGIT vector (B-Tat) or 200, 400 or 800 ng of pcLGIT vector (C-Tat). After 48 hours of transfection, Tat-PLA was performed as detailed above. Confocal images were captured using the same model of a confocal microscope and identical parameters as mentioned above. Image J software was used to measure GFP intensity (AU) and manual quantitation of Tat-PLA spots.

Chromatin immunoprecipitation assay

We used a chromatin preparation equivalent of 2 x 10^6 cells (either GFP^{High} or GFP^{-ve}) for each immunoprecipitation assay, as described previously (Verma A et al., 2016). Briefly, 2 x 10^6 Jurkat cells collected in a 1.5 ml vial were washed with 1X PBS, resuspended in 1 ml of RPMI supplemented with 1% formaldehyde and incubated with gentle agitation for 10 min at room temperature. The cross-linking reaction was quenched by incubating the cells with 0.125 M glycine for 5 min with mild agitation at room temperature followed by centrifugation at 3,000 rpm for 5 min at 4^0C with a subsequent PBS wash (containing 0.01X protease inhibitor cocktail or PIC; #11836170001, Roche Applied Science, *Indianapolis*, USA). Following the complete removal of PBS, the cells were resuspended in 100 µl of ice-chilled lysis buffer (1% SDS, 50 mM Tris buffer, pH 8.0, 10 mM EDTA) and incubated on ice for 20 min with occasional mixing of the lysate using a wide-bore tip. The lysate in each vial was subjected to 22 cycles of sonication at the high mode, using 30-second-ON followed by a 30-second-OFF pulse scheme in the Bioruptor plus sonicator (UCD-300, Diagenode, Liege, Belgium) containing pre-chilled water. The sonicated lysate was centrifuged at 12,000 rpm for 10 min at 4^0C to remove any cellular debris; the clear supernatant was transferred to a fresh 1.5 ml vial and stored at -80^0C until use. One-tenth of the lysate (10 µl) was used to confirm the shearing of chromatin to generate 200- 500 bp fragment sizes. Each IP comprised of 100 µl of lysate and 2 µg of an antigen-specific antibody against p50 (ab7971, Abcam, Cambridge, UK), p65 (ab7970, Abcam), NFAT1 (ab2722, Abcam), NFAT2 (ab2796, Abcam), HIV-1 Tat (ab43014, Abcam or #7377, NIH AIDS reagent program or #4374, NIH AIDS reagent program), RNA Pol II CTD phospho S2 (ab5095, Abcam), or H3K9 Tri Meth (ab8898, Abcam). The ChIPed DNA was amplified using the primer pair N1054 FP (5'-GATCTGAGCC(T/C)GGGAGCTCTCTG-3') and N1056 RP (5'-TCTGAGGGATCTCTAGTTACCAGAGTC-3') spanning a 240 bp sequence within the enhancer-core promoter region in the LTR. The amplified DNA fragments were subjected to agarose gel electrophoresis, and the band intensities were normalized using

the percent-input method to compare differential recruitment of each transcription factor at the active vs. latent promoter. To enhance the sensitivity of the assay, TaqMan qPCR was performed using the ChIP-DNA and the primer-probe combination- N2493 FP, N2215 RP, and N2492 Hex (refer to Table 2.2). The final data were evaluated using the percent input method.

Acknowledgements

We thank Prof. Tapas Kumar Kundu (JNCASR, India) and Dr. Ravi Manjithaya (JNCASR, India) for intellectual discussions. We thank Dr. Uttara Chakraborty, S.L. Swaroopa Yalla and Dr. Narendra Nala of the flow cell at JNCASR, Suma B.S of the Confocal Imaging Facility and Anitha G. of the Sequencing Facility at JNCASR, India. We thank Neelakshi Varma and Surabhi Jirapure for initial help in establishing the latency models. Several reagents were obtained through the AIDS Research and Reference Reagent Program. This work was supported by grants to U.R. from the Department of Biotechnology (DBT), Government of India (DBT grant no. BT/PR7359/29/651/2012); National Institute of Health (NIH), USA (Grant No. NIDA 5RO1DA041751-02), and intramural funds from JNCASR.

Author contributions:

S.C., conception and design, acquisition of data, analysis and interpretation of data, drafting or revising the article, and providing essential unpublished data; **M.K.**, acquisition of data, analysis and interpretation of data, and providing essential unpublished data; **U.R.**, conception and design, fund acquisition, validation, writing, reviewing and editing the article.

Competing Interests

We declare that no competing interests exist.

References

Ammosova, T., Berro, R., Jerebtsova, M., Jackson, A., Charles, S., Klase, Z., Southerland, W., Gordeuk, V.R., Kashanchi, F. and Nekhai, S., 2006. Phosphorylation of HIV-1 Tat by CDK2 in HIV-1 transcription. *Retrovirology*, *3*(1), p.78.

Archin, N.M., Sung, J.M., Garrido, C., Soriano-Sarabia, N. and Margolis, D.M., 2014. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nature Reviews Microbiology*, *12*(11), p.750.

Arkin, A., Ross, J. and McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells. *Genetics*, *149*(4), pp.1633-1648.

Bachu, M., Mukthey, A.B., Murali, R.V., Cheedarla, N., Mahadevan, A., Shankar, S.K., Satish, K.S., Kundu, T.K. and Ranga, U., 2012a. Sequence insertions in the HIV type 1 subtype c viral promoter predominantly generate an additional NF-κB binding site. *AIDS research and human retroviruses*, *28*(10), pp.1362-1368.

Bachu, M., Yalla, S., Asokan, M., Verma, A., Neogi, U., Sharma, S., Murali, R.V., Mukthey, A.B., Bhatt, R., Chatterjee, S. and Rajan, R.E., 2012b. Multiple NF-κB sites in HIV-1 subtype C long terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. *Journal of Biological Chemistry*, *287*(53), pp.44714-44735.

Baeuerle, P.A. and Baltimore, D., 1989. A 65-kappaD subunit of active NF-kappaB is required for inhibition of NF-kappaB by I kappaB. *Genes & development*, *3*(11), pp.1689-1698.

Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.L. and Wandless, T.J., 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, *126*(5), pp.995-1004.

Barbeau, B., Bernier, R., Dumais, N., Briand, G., Olivier, M., Faure, R., Posner, B.I. and Tremblay, M., 1997. Activation of HIV-1 long terminal repeat transcription and virus replication via NF-κB-dependent and-independent pathways by potent phosphotyrosine phosphatase inhibitors, the peroxovanadium compounds. *Journal of Biological Chemistry*, *272*(20), pp.12968-12977.

Barboric, M. and Peterlin, B.M., 2005. A new paradigm in eukaryotic biology: HIV Tat and the control of transcriptional elongation. *PLoS biology*, *3*(2).

Bates, D.L., Barthel, K.K., Wu, Y., Kalhor, R., Stroud, J.C., Giffin, M.J. and Chen, L., 2008. Crystal structure of NFAT bound to the HIV-1 LTR tandem κB enhancer element. *Structure*, *16*(5), pp.684-694.

Bosque, A. and Planelles, V., 2009. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood*, *113*(1), pp.58-65.

Brès, V., Kiernan, R.E., Linares, L.K., Chable-Bessia, C., Plechakova, O., Tréand, C., Emiliani, S., Peloponese, J.M., Jeang, K.T., Coux, O. and Scheffner, M., 2003. A non-proteolytic role for ubiquitin in Tat-mediated transactivation of the HIV-1 promoter. *Nature cell biology*, *5*(8), pp.754-761.

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 pathogens*, *5*(1), p.e1000260.

Cai, W.E.I.Z.H.O.N.G., Astor, T.L., Liptak, L.M., Cho, C., Coen, D.M. and Schaffer, P.A., 1993. The herpes simplex virus type 1 regulatory protein ICP0 enhances virus replication during acute infection and reactivation from latency. *Journal of virology*, *67*(12), pp.7501-7512.

Chan, J.K., Bhattacharyya, D., Lassen, K.G., Ruelas, D. and Greene, W.C., 2013. Calcium/calcineurin synergizes with prostratin to promote NF-κB dependent activation of latent HIV. *PLoS One*, *8*(10).

Chatterjee, A., Kaznessis, Y.N. and Hu, W.S., 2008. Tweaking biological switches through a better understanding of bistability behavior. *Current opinion in biotechnology*, *19*(5), pp.475-481.

Chen-Park, F.E., Huang, D.B., Noro, B., Thanos, D. and Ghosh, G., 2002. The κB DNA sequence from the HIV long terminal repeat functions as an allosteric regulator of HIV transcription. *Journal of Biological Chemistry*, *277*(27), pp.24701-24708.

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. *Nature medicine*, *1*(12), p.1284.

Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C. and Kuo, Y.H., 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*, *387*(6629), pp.183-188.

Colin, L. and Van Lint, C., 2009. Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. *Retrovirology*, *6*(1), p.111.

Dandekar, D.H., Ganesh, K.N. and Mitra, D., 2004. HIV-1 Tat directly binds to NFκB enhancer sequence: role in viral and cellular gene expression. *Nucleic acids research*, *32*(4), pp.1270-1278.

Dodd, I.B., Perkins, A.J., Tsemitsidis, D. and Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of P RM and efficient switching from lysogeny. *Genes & development*, *15*(22), pp.3013-3022.

Doetzlhofer, A., Rotheneder, H., Lagger, G., Koranda, M., Kurtev, V., Brosch, G., Wintersberger, E. and Seiser, C., 1999. Histone deacetylase 1 can repress transcription by binding to Sp1. *Molecular and cellular biology*, *19*(8), pp.5504-5511.

Duverger, A., Wolschendorf, F., Zhang, M., Wagner, F., Hatcher, B., Jones, J., Cron, R.Q., van der Sluis, R.M., Jeeninga, R.E., Berkhout, B. and Kutsch, O., 2013. An AP-1 binding site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. *Journal of virology*, *87*(4), pp.2264-2277.

Dwarakanath, R.S., Clark, C.L., McElroy, A.K. and Spector, D.H., 2001. The use of recombinant baculoviruses for sustained expression of human cytomegalovirus immediate early proteins in fibroblasts. *Virology*, *284*(2), pp.297-307.

Eisele, E. and Siliciano, R.F., 2012. Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity*, *37*(3), pp.377-388.

Endo-Munoz, L., Warby, T., Harrich, D. and McMillan, N.A., 2005. Phosphorylation of HIV Tat by PKR increases interaction with TAR RNA and enhances transcription. *Virology journal*, *2*(1), p.17.

Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C. and Quinn, T.C., 1999. Latent infection of CD4+ T cells provides a

mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature medicine*, *5*(5), p.512.

Garcia, J.A., Harrich, D., Pearson, L., Mitsuyasu, R. and Gaynor, R.B., 1988. Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. *The EMBO journal*, *7*(10), pp.3143-3147.

Giffin, M.J., Stroud, J.C., Bates, D.L., von Koenig, K.D., Hardin, J. and Chen, L., 2003. Structure of NFAT1 bound as a dimer to the HIV-1 LTR κB element. *Nature Structural and Molecular Biology*, *10*(10), p.800.

Gustafsdottir, S.M., Schallmeiner, E., Fredriksson, S., Gullberg, M., Söderberg, O., Jarvius, M., Jarvius, J., Howell, M. and Landegren, U., 2005. Proximity ligation assays for sensitive and specific protein analyses. *Analytical biochemistry*, *345*(1), pp.2-9.

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

Hoffmann, A., Levchenko, A., Scott, M.L. and Baltimore, D., 2002. The IκB-NF-κB signaling module: temporal control and selective gene activation. *Science*, *298*(5596), pp.1241-1245.

Jeeninga, R.E., Hoogenkamp, M., Armand-Ugon, M., de Baar, M., Verhoef, K. and Berkhout, B.E.N., 2000. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *Journal of virology*, *74*(8), pp.3740-3751.

Jordan, M., Schallhorn, A. and Wurm, F.M., 1996. Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic acids research*, *24*(4), pp.596-601.

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

Kent, J.R., Kang, W., Miller, C.G. and Fraser, N.W., 2003. Herpes simplex virus latency-associated transcript gene function. *Journal of neurovirology*, *9*(3), pp.285-290.

Kinoshita, S., Su, L., Amano, M., Timmerman, L.A., Kaneshima, H. and Nolan, G.P., 1997. The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. *Immunity*, *6*(3), pp.235-244.

Kinoshita, S., Chen, B.K., Kaneshima, H. and Nolan, G.P., 1998. Host control of HIV-1 parasitism in T cells by the nuclear factor of activated T cells. *Cell*, *95*(5), pp.595-604.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.C. and Kain, S.R., 1998. Generation of destabilized green fluorescent protein as a transcription reporter. *Journal of Biological Chemistry*, *273*(52), pp.34970-34975.

Macián, F. and Rao, A., 1999. Reciprocal modulatory interaction between human immunodeficiency virus type 1 Tat and transcription factor NFAT1. *Molecular and Cellular Biology*, *19*(5), pp.3645-3653.

Mahmoudi, T., 2012. The BAF complex and HIV latency. *Transcription*, *3*(4), pp.171-176.

Mbondji-Wonje, C., Dong, M., Wang, X., Zhao, J., Ragupathy, V., Sanchez, A.M., Denny, T.N. and Hewlett, I., 2018. Distinctive variation in the U3R region of the 5'Long Terminal Repeat from diverse HIV-1 strains. *PloS one*, *13*(4).

Mbonye, U. and Karn, J., 2017. The molecular basis for human immunodeficiency virus latency. *Annual review of virology*, *4*, pp.261-285.

Montano, M.A., Novitsky, V.A., Blackard, J.T., Cho, N.L., Katzenstein, D.A. and Essex, M., 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *Journal of Virology*, *71*(11), pp.8657-8665.

Ott, M., Dorr, A., Hetzer-Egger, C., Kaehlcke, K., Schnolzer, M., Henklein, P., Cole, P., Zhou, M.M. and Verdin, E., 2004, June. Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation. In *Reversible Protein Acetylation: Novartis Foundation Symposium 259* (pp. 182-196). John Wiley & Sons, Ltd.

Ott, M., Geyer, M. and Zhou, Q., 2011. The control of HIV transcription: keeping RNA polymerase II on track. *Cell host & microbe*, *10*(5), pp.426-435.

Pagans, S., Kauder, S.E., Kaehlcke, K., Sakane, N., Schroeder, S., Dormeyer, W., Trievel, R.C., Verdin, E., Schnolzer, M. and Ott, M., 2010. The Cellular lysine methyltransferase Set7/9-KMT7 binds HIV-1 TAR RNA, monomethylates the viral transactivator Tat, and enhances HIV transcription. *Cell host & microbe*, *7*(3), pp.234-244.

Pai, A. and Weinberger, L.S., 2017. Fate-Regulating Circuits in Viruses: From Discovery to New Therapy Targets. *Annual review of virology*, *4*, pp.469-490.

Pearson, R., Kim, Y.K., Hokello, J., Lassen, K., Friedman, J., Tyagi, M. and Karn, J., 2008. Epigenetic silencing of human immunodeficiency virus (HIV) transcription by formation of restrictive chromatin structures at the viral long terminal repeat drives the progressive entry of HIV into latency. *Journal of virology*, *82*(24), pp.12291-12303.

Pereira, L.A., Bentley, K., Peeters, A., Churchill, M.J. and Deacon, N.J., 2000. SURVEY AND SUMMARY A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic acids research*, *28*(3), pp.663-668.

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

Pessler, F. and Cron, R.Q., 2004. Reciprocal regulation of the nuclear factor of activated T cells and HIV-1. *Genes and immunity*, *5*(3), p.158.

Pierson, T., McArthur, J. and Siliciano, R.F., 2000. Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annual review of immunology*, *18*(1), pp.665-708.

Ragoczy, T. and Miller, G., 2001. Autostimulation of the Epstein-Barr virus BRLF1 promoter is mediated through consensus Sp1 and Sp3 binding sites. *Journal of virology*, *75*(11), pp.5240-5251.

Razooky, B.S. and Weinberger, L.S., 2011. Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. *Methods*, *53*(1), pp.68-77.

Razooky, B.S., Pai, A., Aull, K., Rouzine, I.M. and Weinberger, L.S., 2015. A hardwired HIV latency program. *Cell*, *160*(5), pp.990-1001.

Rohr, O., Marban, C., Aunis, D. and Schaeffer, E., 2003. Regulation of HIV-1 gene transcription: from lymphocytes to microglial cells. *Journal of leukocyte biology*, *74*(5), pp.736-749.

Roizman, B., Gu, H. and Mandel, G., 2005. The first 30 minutes in the life of a virus: unREST in the nucleus. *Cell Cycle*, *4*(8), pp.1019-1021.

Sanders, R.L., Clark, C.L., Morello, C.S. and Spector, D.H., 2008. Development of cell lines that provide tightly controlled temporal translation of the human cytomegalovirus IE2 proteins for complementation and functional analyses of growth-impaired and nonviable IE2 mutant viruses. *Journal of virology*, *82*(14), pp.7059-7077.

Sarisky, R.T., Gao, Z., Lieberman, P.M., Fixman, E.D., Hayward, G.S. and Hayward, S.D., 1996. A replication function associated with the activation domain of the Epstein-Barr virus Zta transactivator. *Journal of virology*, *70*(12), pp.8340-8347.

Sivakumaran, H., van der Horst, A., Fulcher, A.J., Apolloni, A., Lin, M.H., Jans, D.A. and Harrich, D., 2009. Arginine methylation increases the stability of human immunodeficiency virus type 1 Tat. *Journal of virology*, *83*(22), pp.11694-11703.

Siliciano, R.F. and Greene, W.C., 2011. HIV latency. *Cold Spring Harbor perspectives in medicine*, *1*(1), p.a007096.

Simon, V. and Ho, D.D., 2003. HIV-1 dynamics in vivo: implications for therapy. *Nature Reviews Microbiology*, *1*(3), pp.181-190.

Smith, K.A. and Popmihajlov, Z., 2008. The quantal theory of immunity and the interleukin-2-dependent negative feedback regulation of the immune response. *Immunological reviews*, *224*(1), pp.124-140.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G. and Landegren, U., 2006. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature methods*, *3*(12), pp.995-1000.

Southgate, C.D. and Green, M.R., 1991. The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function. *Genes & Development*, *5*(12b), pp.2496-2507.

Stinski, M.F. and Petrik, D.T., 2008. Functional roles of the human cytomegalovirus essential IE86 protein. In *Human Cytomegalovirus* (pp. 133-152). Springer Berlin Heidelberg.

Stroud, J.C., Oltman, A., Han, A., Bates, D.L. and Chen, L., 2009. Structural basis of HIV-1 activation by NF-κB—A higher-order complex of p50: Rela bound to the HIV-1 LTR. *Journal of molecular biology*, *393*(1), pp.98-112.

Suzuki, T., Yamamoto, T., Kurabayashi, M., Nagai, R., Yazaki, Y. and Horikoshi, M., 1998. Isolation and initial characterization of GBF, a novel DNA-binding zinc finger protein that binds to the GC-rich binding sites of the HIV-1 promoter. *The Journal of Biochemistry*, *124*(2), pp.389-395.

Tosi, G., Meazza, R., De Lerma Barbaro, A., D'Agostino, A., Mazza, S., Corradin, G., Albini, A., Noonan, D.M., Ferrini, S. and Accolla, R.S., 2000. Highly stable oligomerization forms of HIV-1 Tat detected by monoclonal antibodies and requirement of monomeric forms for the transactivating function on the HIV-1 LTR. *European journal of immunology*, *30*(4), pp.1120-1126.

Van Duyne R, Easley R, Wu W, Berro R, Pedati C, Klase Z, Kehn-Hall K, Flynn EK, Symer DE, Kashanchi F. Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. Retrovirology. 2008 Dec 1;5(1):40.

Van Lint, C., Bouchat, S. and Marcello, A., 2013. HIV-1 transcription and latency: an update. *Retrovirology*, *10*(1), p.67.

Van Zyl, G., Bale, M.J. and Kearney, M.F., 2018. HIV evolution and diversity in ART-treated patients. *Retrovirology*, *15*(1), p.14.

Verma, A., Rajagopalan, P., Lotke, R., Varghese, R., Selvam, D., Kundu, T.K. and Ranga, U., 2016. Functional Incompatibility between the Generic NF-κB Motif and a Subtype-Specific Sp1III Element Drives the Formation of the HIV-1 Subtype C Viral Promoter. *Journal of virology*, *90*(16), pp.7046-7065.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P. and Schaffer, D.V., 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*, *122*(2), pp.169-182.

Weinberger, L.S. and Shenk, T., 2006. An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. *PLoS biology*, *5*(1), p.e9.

Weinberger, L.S., Dar, R.D. and Simpson, M.L., 2008. Transient-mediated fate determination in a transcriptional circuit of HIV. *Nature genetics*, *40*(4), p.466.

Weinberger, A.D. and Weinberger, L.S., 2013. Stochastic fate selection in HIV-infected patients. *Cell*, *155*(3), pp.497-499.

Williams, S.A., Chen, L.F., Kwon, H., Ruiz-Jarabo, C.M., Verdin, E. and Greene, W.C., 2006. NF-κB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *The EMBO journal*, *25*(1), pp.139-149.

Xie, B., Invernizzi, C.F., Richard, S. and Wainberg, M.A., 2007. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat Interactions with both cyclin T1 and the Tat transactivation region. *Journal of virology*, *81*(8), pp.4226-4234.

Figures and legends

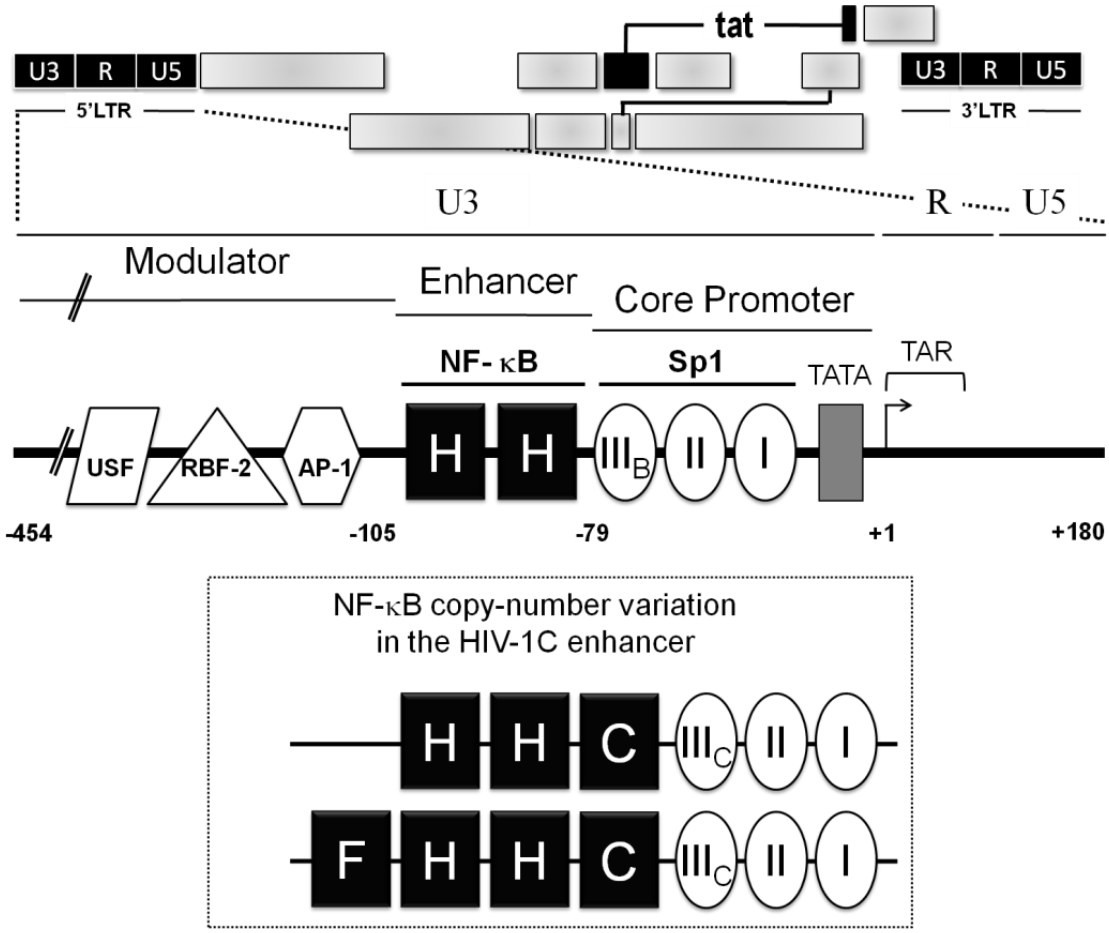
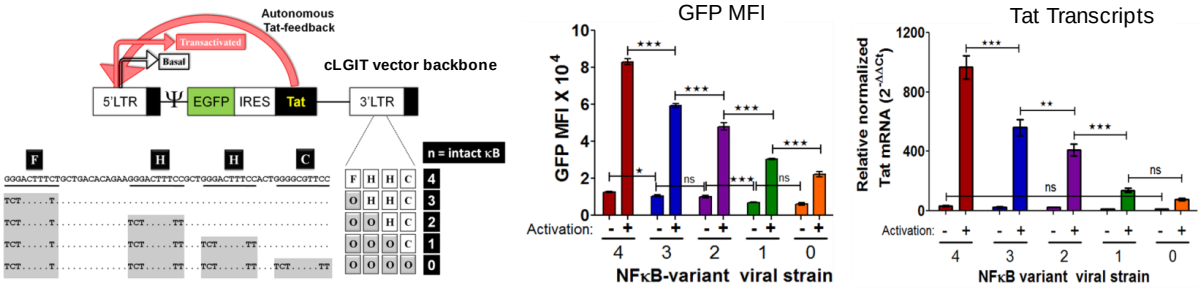
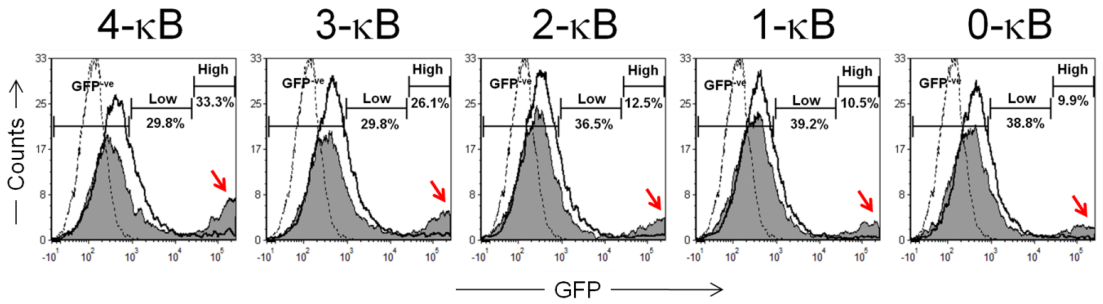


Figure 1: A schematic representation of NF-κB motif diversity in HIV-1C LTR. The canonical HIV-1B LTR containing two identical NF-κB motifs is presented at the top. The distinct regulatory regions (U3, R, U5, the Modulator, the Enhancer, and the Core promoter) and important transcription factor binding sites have been depicted. HIV-1C LTR not only contains more copies of the NF-κB motif (3 or 4 copies) but the additional motifs are also genetically variable (the bottom panel). The three genetically distinct NF-κB motifs present in the C-LTR are denoted as H (GGGACTTTCC), C (GGGGCGTTCC, differences underlined), and F (GGGACTTTCT). Note that the Sp1III site also contains subtype-specific variations as presented; B and C representing respective viral subtypes.

(A) cLGIT series of minimal HIV-1C vectors



(B) LTR transactivation



(C) The kinetics of latency establishment

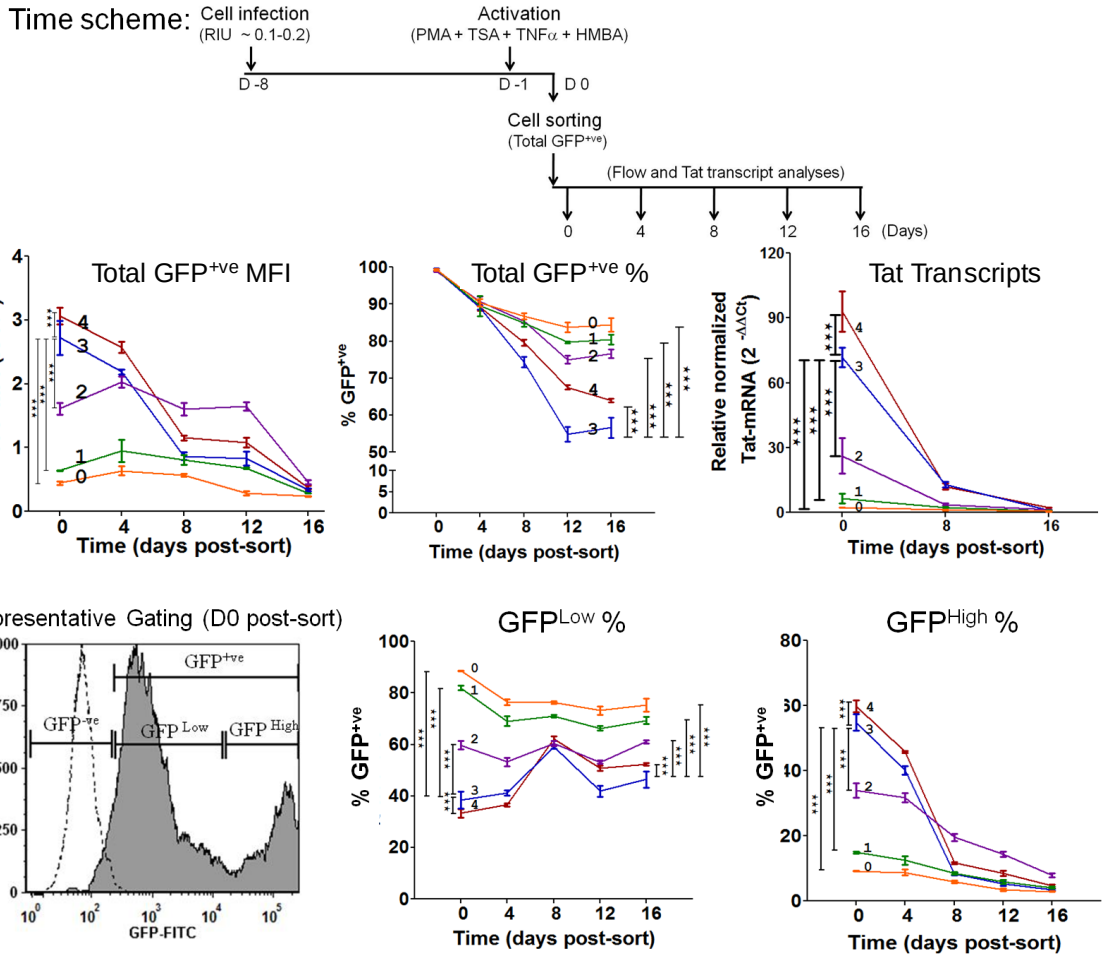
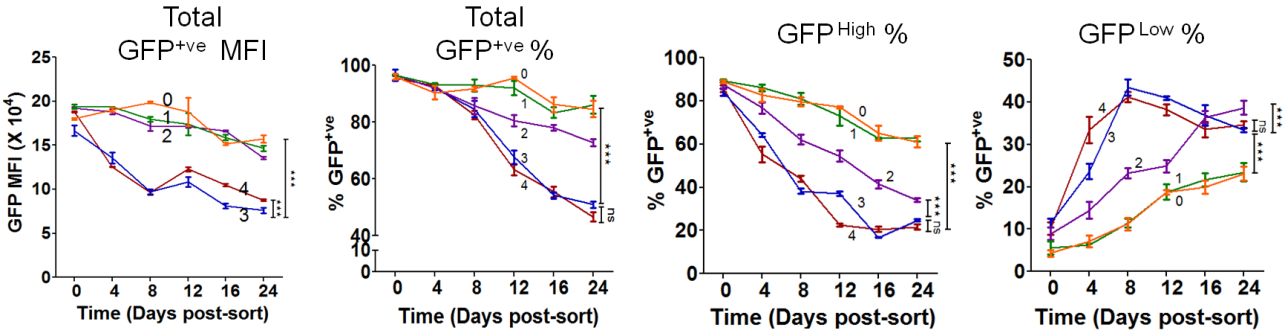
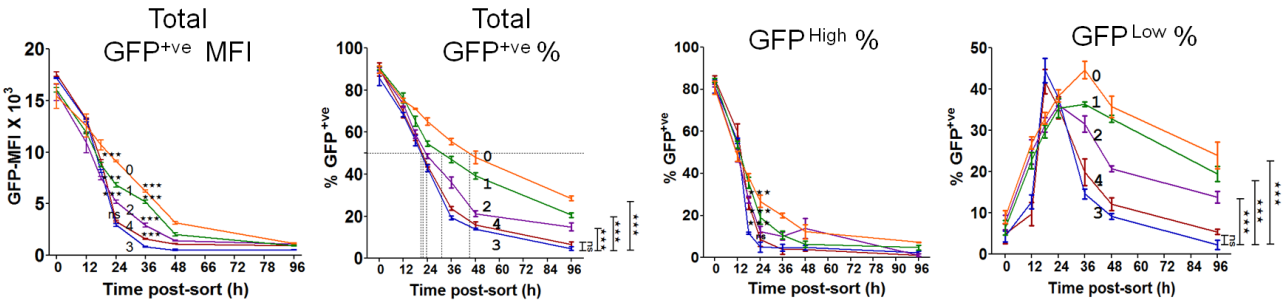


Figure 2: In the autonomous Tat-feedback (ATF) model, more NF-κB motifs cause stronger gene-expression as well as faster latency. (A) Viral gene expression is proportional to the number of NF-κB motifs. A schematic of the cLGIT sub-genomic viral vector panel of the autonomous feedback (ATF) model is presented (left panel). Note that both the 3' LTR and Tat are of HIV-1C origin. The four NF-κB motifs, the inactivating mutations introduced in the motifs, and the variant viral strains of the panel have been depicted. The three kinds of NF-κB motifs (H, C, and F) are genetically variant (Bachu M et al., 2012b). Viral gene-expression (GFP MFI; central panel) and Tat transcript expression (right panel) in Jurkat cells are indicated. One million Jurkat cells were infected at an RIU of ~ 0.5-0.6, independently with each LTR-variant strain. After 72 h of infection, half of the infected cells were activated with a cocktail of global T-cell activators (PMA+TNFα+TSA+HMBA) and 24 h post-activation, GFP expression (EGFP) was estimated for both the un-activated and activated fractions using a flow-cytometer (central panel). The mean GFP MFI values from experimental quadruplicates ± SD, representative of two independent experiments are presented. Total mRNA was extracted from 0.2 to 0.5 million cells of un-activated and activated populations, and Tat expression was evaluated in an RT-PCR using the ΔΔCt method (right panel). GAPDH was used as the reference gene. Mean values of the relative Tat expression from three independent experiments ± SEM are plotted. Two-way ANOVA with Bonferroni post-test correction was used for statistical analyses. **(B)** Representative GFP histograms from the above panel. In each panel, the black dotted histogram represents Jurkat cells not infected and not activated; the black hollow histogram represents cells infected but not activated, and the solid grey histogram represents cells infected and activated. The intensity ranges of GFP^{-ve}, GFP^{Low}, and GFP^{High} are marked. The red arrows mark the GFP^{High} (MFI > 10^4 RFU) population representing the Tat-transactivated population. **(C)** Stronger the promoter, faster the latency establishment. A schematic representation of the experimental protocol is depicted (top panel). One million Jurkat cells were infected with individual strains of the panel at a low infectious titer of ~0.1-0.2 RIU, allowed to relax for a week, treated with the cocktail of global T-cell activators, and 24 h later, all the GFP^{+ve} cells (harboring active provirus; GFP MFI >10^3 RFU) were sorted. The sorted, total GFP^{+ve} cells were then maintained in culture, and the EGFP expression was monitored by flow cytometry every four days (middle-left and -central panels) and that of Tat transcripts (middle-right panel) on days 0, 8 and 16. A representative, post-sort, stacked histogram profile from the 4-κB variant is shown depicting the gating strategy (bottom-left). Mean values from experimental triplicates ± SD, representative of three independent experiments are plotted. Two-way ANOVA with Bonferroni post-test correction was used for the statistical evaluation (*$p<0.05$, **$p<0.01$, ***$p<0.001$ and ns – non-significant).

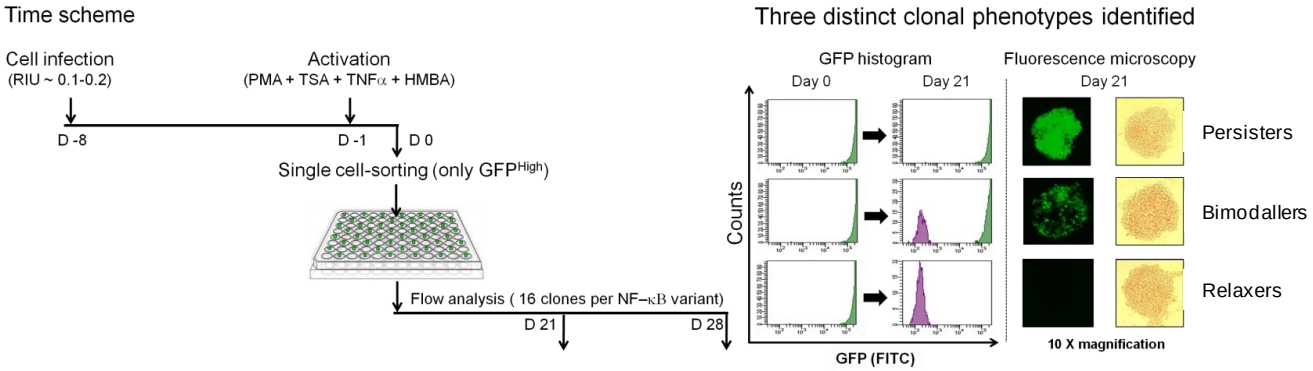
(A) Kinetics of latency with EGFP as the reporter



(B) Kinetics of latency with d2EGFP as the reporter



(C) Kinetics of latency in Jurkat-cLGIT clonal lines (EGFP reporter)



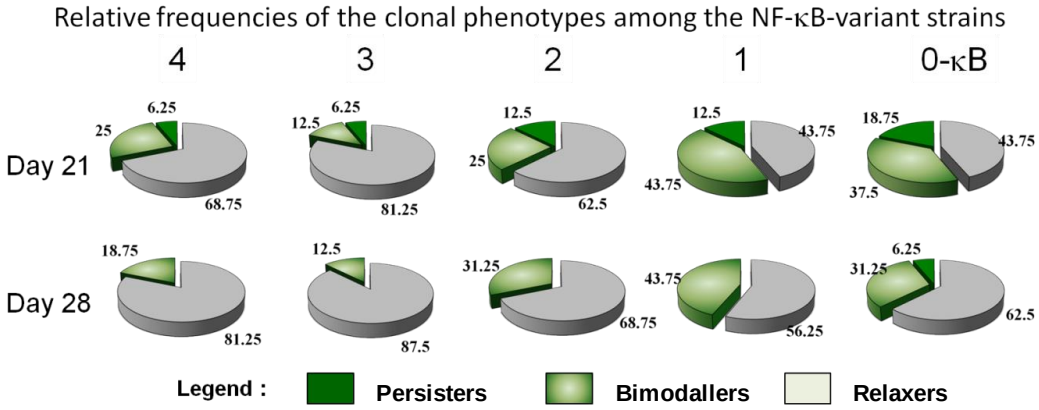
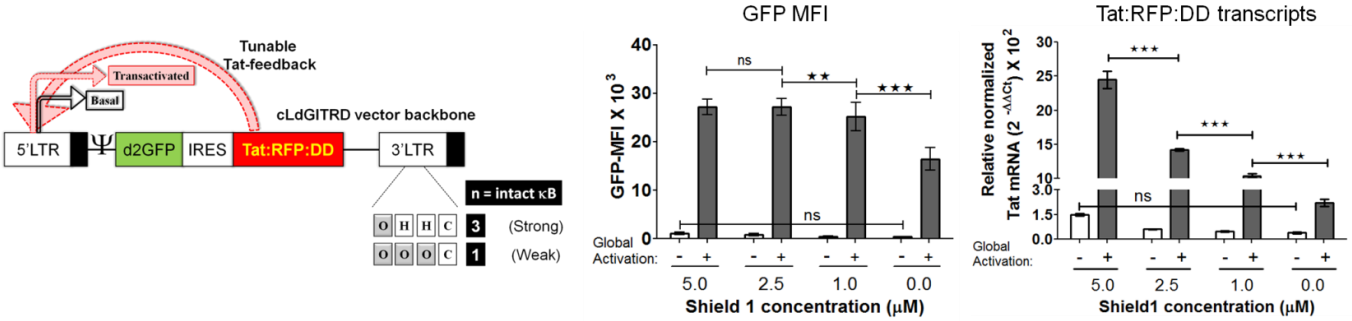
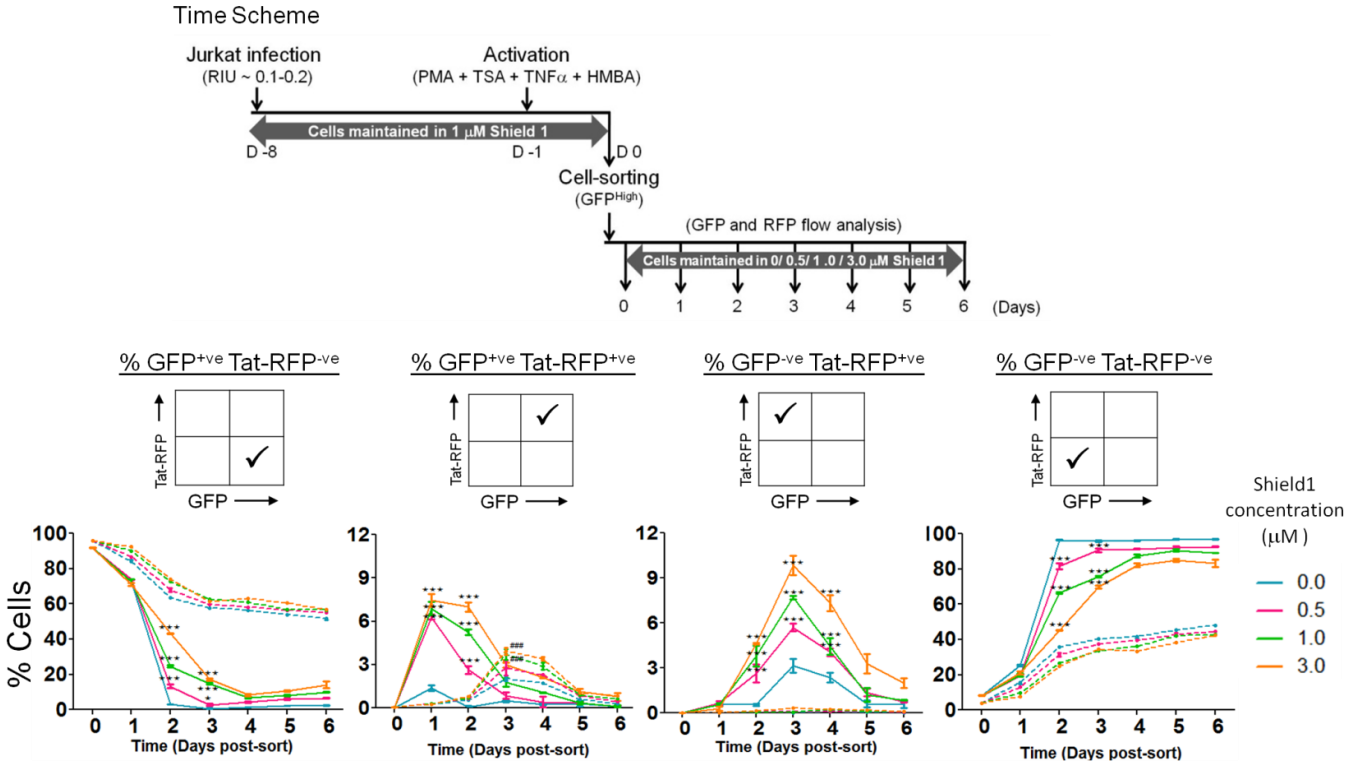


Figure 3: **The Tat-transactivated population delineates naturally between strong (3- and 4-κB) and weak (2, 1, and 0-κB) LTRs as the fast and the slow silencing promoters, respectively. (A and B)** The kinetics of latency establishment in the GFP^{High} population harboring either EGFP or d2EGFP reporter viral variants. Experimental layouts are detailed in Fig S4B, top panel, and S5C, respectively. Mean values from experimental triplicates ± SD, representative of two independent experiments are plotted. Two-way ANOVA with Bonferroni post-test correction was used for the statistical evaluation (*p<0.05, **p<0.01, ***p<0.001 and ns – non-significant). **(C)** The kinetics of latency establishment in J-cLGIT clonal cell populations of the ATF model. The experimental layout (top-left panel) for the clonal population is essentially similar to that of the non-clonal population kinetics. The $EGFP^{High}$ cells (GFP MFI $>10^4$ RFU) were single-cell sorted, expanded for three-four weeks, and the pattern of EGFP expression was monitored on days 21 and 28 post-sorting by flow cytometry and fluorescence microscopy. EGFP expression of 16 randomly selected clones, corresponding to each viral variant was measured. Based on the fluorescence profile, three distinct categories of clones- persisters (GFP-MFI $>10^4$ RFU), relaxers (GFP^{-ve}; MFI $<10^3$), and bimodallers were identified (top-right panel). The proportion of the above three phenotypes among the LTR-variants across time is indicated at Day 21 and 28 post-sorting (bottom panel).

(A) cLdGITRD series of minimal HIV-1C vectors



(B) The kinetics of latency establishment



(C) The two modes of latency establishment

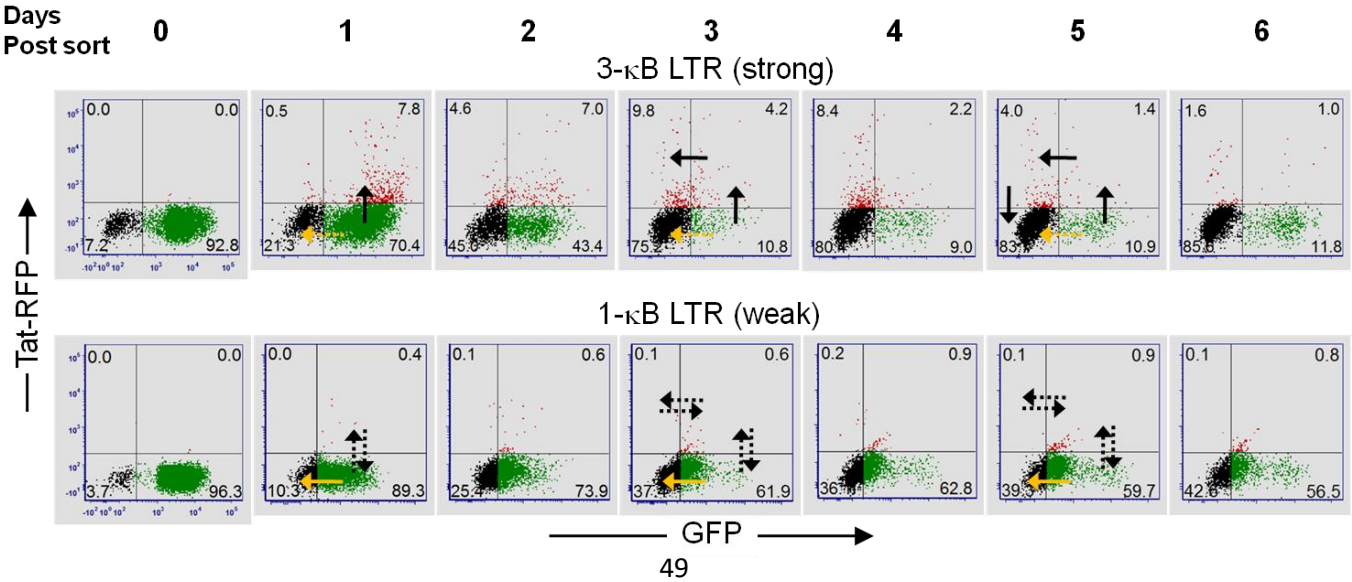
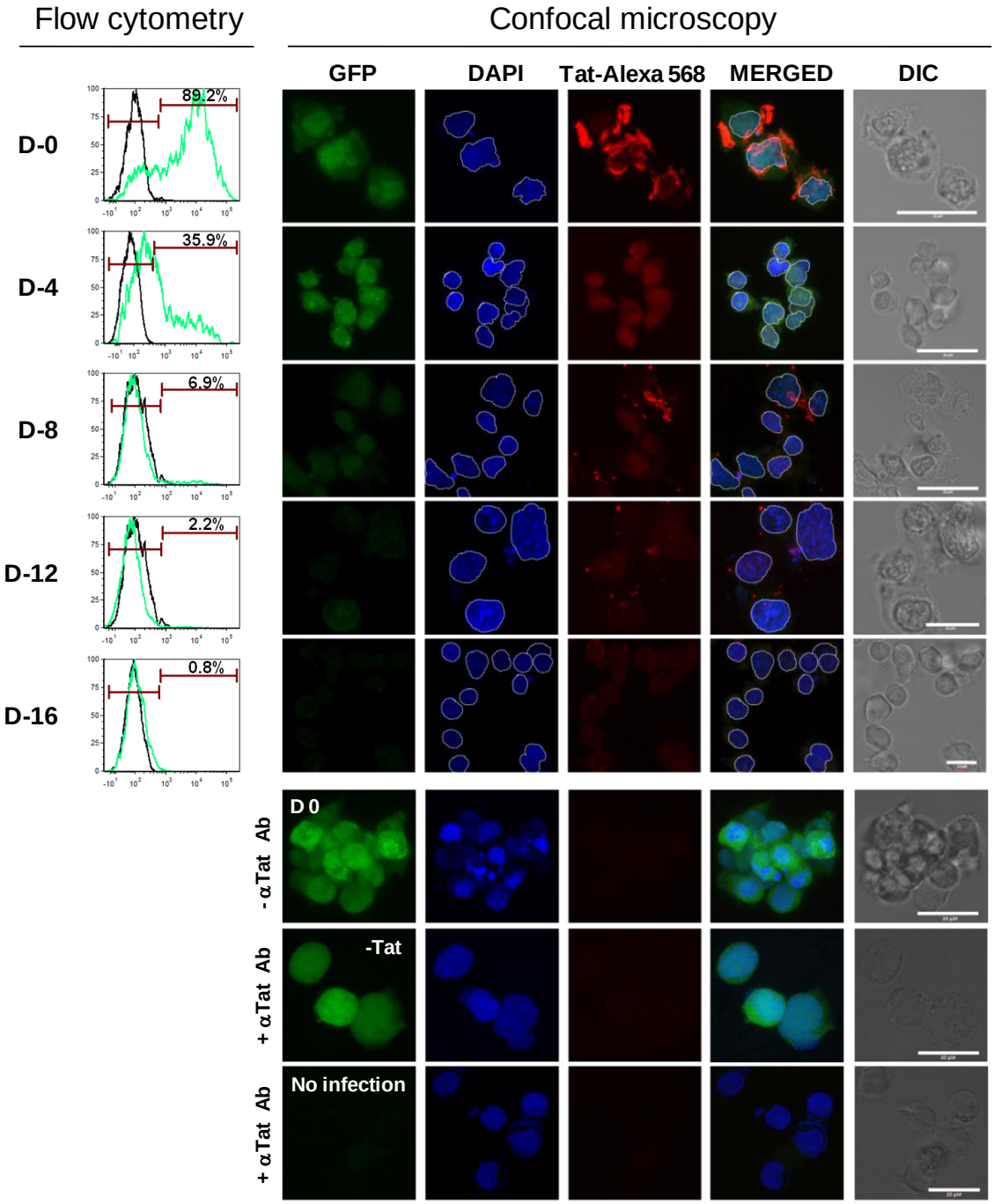


Figure 4: The tunable Tat-feedback (TTF) model identifies two distinct modes of latency establishment in the strong and weak LTRs. (A) A positive correlation between the viral gene-expression and progressively increasing Tat-feedback strength for a fixed LTR configuration. A schematic of the sub-genomic HIV-1 vector backbone- cLTR-d2EGFP-IRES-Tat:RFP:DD (cLdGITRD) representing the TTF model (left panel) is shown. After 24 h of Shield1 treatment and following cell activation for an additional 24 h, d2EGFP expression was measured using flow cytometry. The mean GFP MFI values from experimental quadruplicates ± SD, representing two independent experiments, are presented (central panel). The relative Tat transcript expression (right panel) was evaluated as in Fig 2A and 2C. The mean values of the relative Tat mRNA from three independent experiments ± SEM are plotted. Two-way ANOVA with Bonferroni post-test correction was used for the statistical evaluation (*$p<0.05$, **$p<0.01$, ***$p<0.001$ and ns – non-significant). **(B)** The strong (3-κB) and the weak (1-κB) LTRs followed the Tat-dependent and the Tat-independent silencing modes, respectively. The experimental scheme is depicted (top panel), and the protocol followed was similar to that as described in Fig 2C, 3A, and 3B. The sorted d2EGFP^{High} cells were divided into four separate fractions and maintained at different concentrations of Shield1 (0, 0.5, 1.0, or 3.0 μM). The cells were analyzed every 24 h for d2EGFP and RFP expression using flow cytometry. The individual kinetic curves of the four distinct populations are shown in the bottom panel. Mean values from experimental triplicates ± SD are plotted. Data represent three independent experiments. Two-way ANOVA with Bonferroni post-test was used for the statistical evaluation (*$p<0.05$, ***$p<0.001$ and ns - non-significant). **(C)** The temporal d2EGFP and RFP trajectories of the strong (3-κB; upper panel) and weak (1-κB; lower panel) LTRs at the 3 μM Shield1 concentration are presented. The black and the yellow arrows indicate the Tat-dependent and Tat-independent latency modes, respectively, while the solid and dotted arrows denote the dominant and less-dominant routes of transit for each LTR-variant, respectively.

(A) Indirect immunofluorescence of Tat



(B) Temporal quantitation of d2EGFP and Tat-Alexa 568 expression

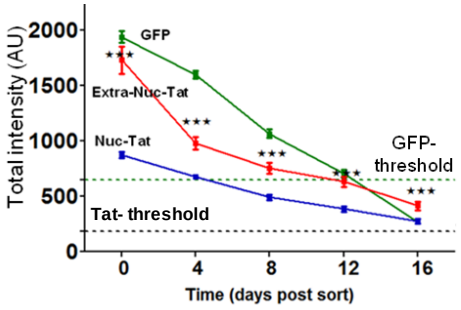


Figure 5: Persistent presence of Tat in latent cells. (A) The temporal profile of proviral Tat expression during LTR-silencing in the stable J-LdGIT-3-κB cell pool. The experimental strategy of latency establishment was similar to that described in Fig 3B. At defined time points, a small fraction of the sorted d2EGFP^{High} cells was flow-analyzed for GFP expression (left panel), while the remaining cells were subjected to indirect immunofluorescence of Tat using a rabbit anti-Tat primary antibody and an anti-rabbit Alexa-568 conjugated secondary antibody. DAPI was used to stain the nucleus. Representative confocal images of Tat and GFP-expression at indicated time points (right panel) are presented. Appropriate negative controls for Tat IF are presented in the bottom panels. The white dotted line demarcates the nucleus from the extra-nuclear compartments in each cell. Scale bar = 20μM. **(B)** Quantitative estimation of d2EGFP and Tat-Alexa 568 expression levels in the nuclear and extra-nuclear compartments (arbitrary units) at multiple time points. Data from 150 individual cells at each time point and three independent experiments are presented. The threshold intensity values for total cellular GFP and Tat-Alexa 568 intensities were obtained from uninfected Jurkat cells (Fig 5A; Lane-8) and infected, unstained cells (Fig 5A; Lane-6), respectively. Mean values ± SEM are plotted. One-way ANOVA was used for statistical evaluation (***$p<0.001$).

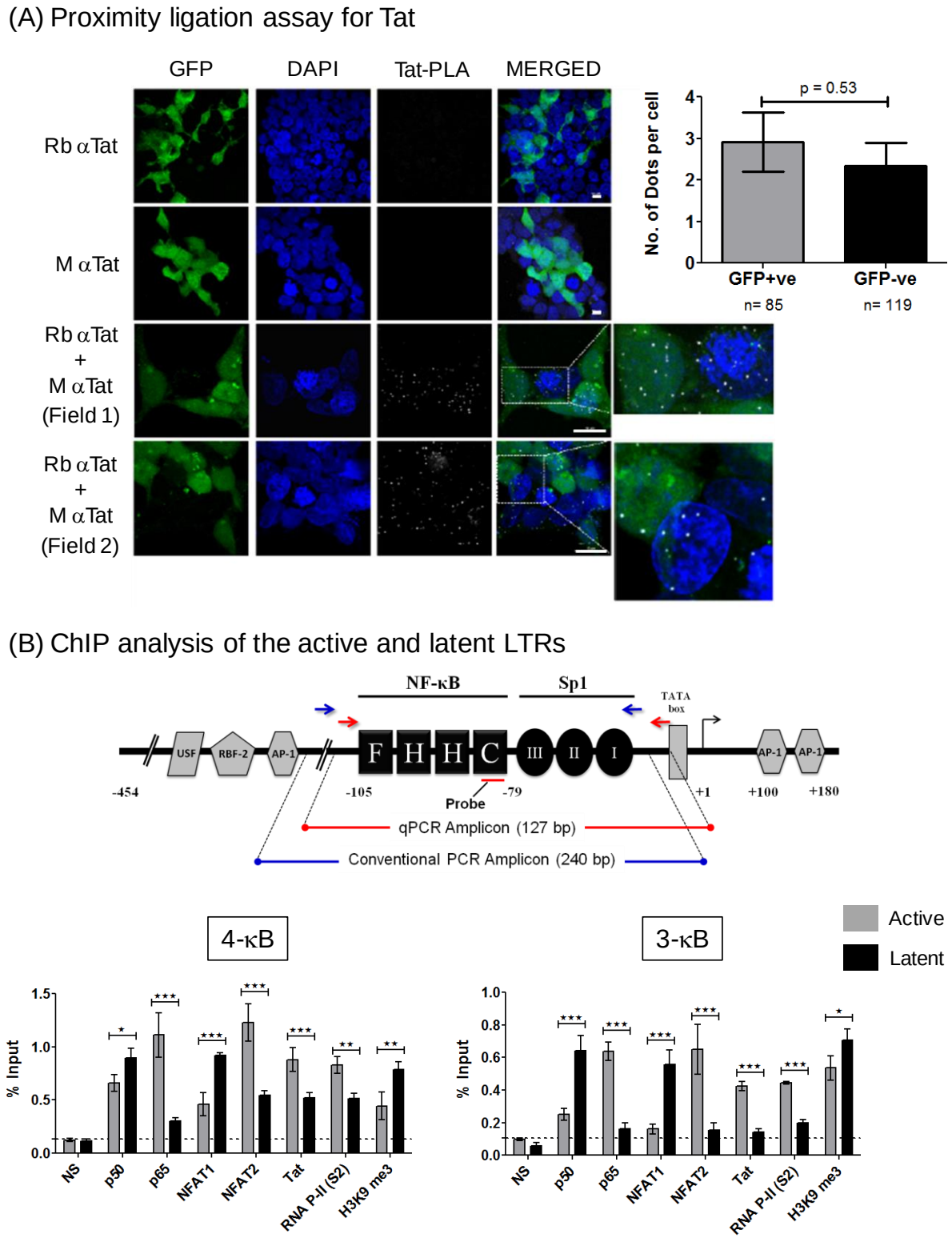


Figure 6: The active and latent LTRs recruit host-factors differentially. (A) The proximity ligation assay for Tat in the active vs. latent cells. HEK293 cells stably infected with the cLdGIT-4-κB strain of the ATF panel (~ 0.5 RIU) were sorted for GFP^{High} cells, allowed for proviral-LTR relaxation to arrive at a mixed population of GFP^{+ve} (active) and GFP^{-ve} (latent) cells and subjected to the PLA for Tat. Approximately 50,000 mixed GFP cells seeded in a well of an 8-well slide chamber were stained for Tat using a pair of anti-Tat primary antibodies raised in rabbit and mouse. The PLA was performed according to the manufacturer's protocol (Cat# DUO12901, Sigma-Aldrich). Representative confocal images depicting single antibody controls (Lanes-1 and 2) and Tat-PLA with both the antibodies (Lanes-3 and 4) are shown (left panel). Two sub-fields with distinct Tat-PLA dots (white) in both GFP^{+ve} and GFP^{-ve} cells have been enlarged for clarity. The number of Tat-PLA dots per cell was determined independently for GFP^{+ve} as well as GFP^{-ve} cells, and the mean values from three independent experiments

± SEM are plotted. A two-tailed, unpaired t-test was used for statistical analyses. **(B)** The ChIP analysis to assess the recruitment of cellular factors and Tat to the active and latent LTRs. Schematic representation of the LTR sequence and the position of the primer pairs for the conventional PCR (240 bp, blue) and qPCR (127 bp, red) is depicted (top-panel). ChIP-qPCR data of the active (GFP^{+ve}) and latent (GFP^{-ve}) promoters of two individual bimodal clones obtained from single-cell sorting (Fig 3C) representing the 4-κB (bottom-left panel) and 3-κB (bottom-right panel) proviruses are shown. The cell-lysate of two million cells (GFP^{+ve} or GFP^{-ve}) and 2 μg of the respective antibody were used for the individual IP reactions. One-tenth of the IP chromatin was used as the input control. A Taqman qPCR targeting the enhancer region was performed, and data for each IP was calculated using the percent-input method. The mean values from experimental quadruplicates ± SD are plotted. A two-tailed unpaired t-test was used for statistical analysis (*$p<0.05$, ***$p<0.001$, and ns- non-significant). Data from a conventional PCR-based ChIP assay, independently performed using identical experimental conditions as the qPCR-based ChIP, are presented in Fig S9 and S10.

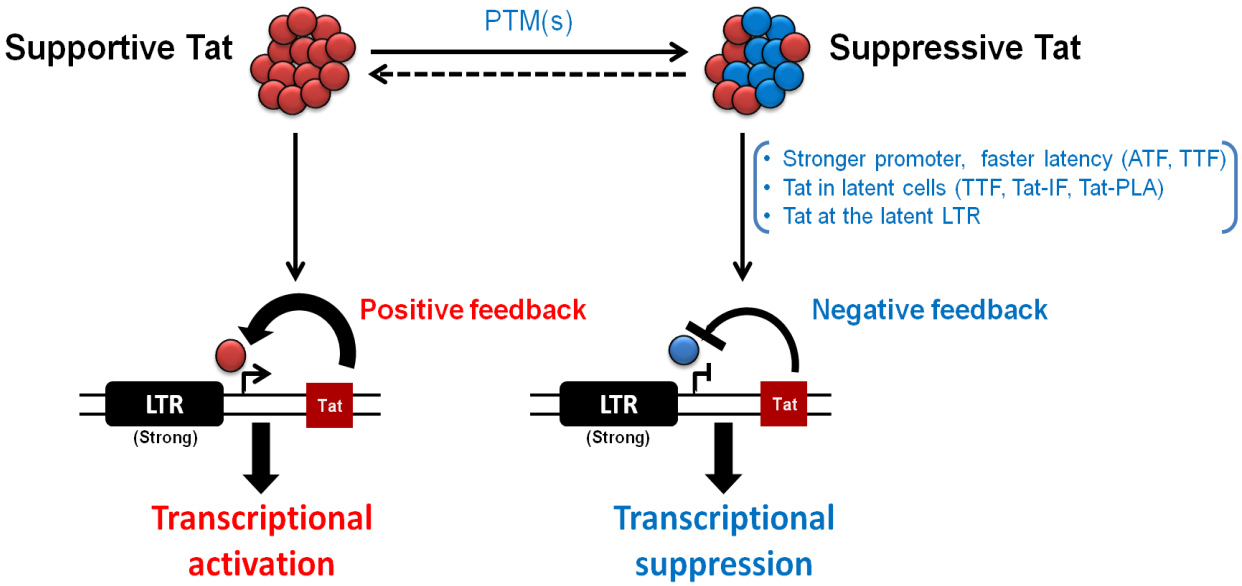


Figure 7: **A hypothetical model depicts that Tat controls its own PTM, thereby switching between being an activator or suppressor of viral transcription.** The model suggests that the diverse PTM(s) of Tat play a critical role in converting the transactivator between being an activator or suppressor of transactivation of the LTR. At the time of commitment to latency, the intracellular concentrations of Tat are not limiting. Tat may initiate negative feedback of viral transcription in a concentration-dependent manner by regulating the expression of host factors that control the PTM(s) of Tat. The negative feedback effect of Tat, therefore, follows its initial positive feedback on the viral promoter. A viral promoter characterized by stronger transcriptional activity, thus, mediates the establishment of latency at a faster rate by producing more quantities of Tat. The data presented in this work are consistent with this model.