

# Quantitative limits of host-driven HIV transcription and host gene control by the viral transactivator Tat

Chuan Li<sup>1,\*</sup>, Yuan Ma<sup>1</sup>, Yi Wang<sup>1</sup>, Huiming Yang<sup>1</sup>, Susana T. Valente<sup>1,2,\*</sup>

<sup>1</sup>Department of Immunology and Microbiology, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, Jupiter, FL 33458, United States

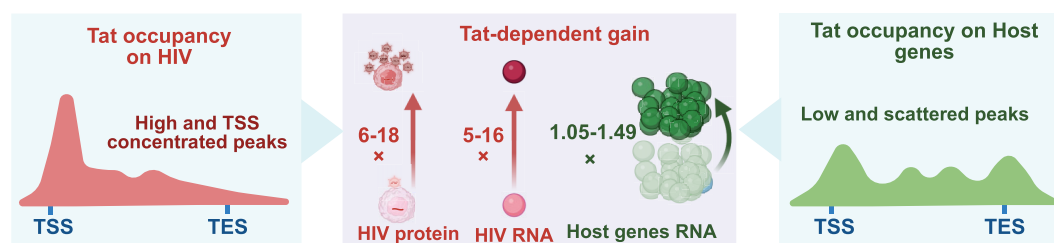
<sup>2</sup>Skaggs Graduate School of Chemical and Biological Sciences, The Scripps Research Institute, Jupiter, FL 33458, United States

\*To whom correspondence should be addressed. Email: [svalente@ufl.edu](mailto:svalente@ufl.edu)  
Correspondence may also be addressed to Chuan Li. Email: [lichuan@ufl.edu](mailto:lichuan@ufl.edu)

## Abstract

HIV transcription is amplified by the viral transactivator Tat, yet two quantitative issues remain unresolved in chromatin: the extent to which host machinery sustains transcription in the absence of Tat, and why Tat's widespread chromatin association yields limited host-gene output. We engineered an isogenic Tat-deficient derivative of the HIV<sub>GKO</sub> dual-reporter virus that preserves native proviral architecture, enabling longitudinal measurements across diverse integration sites in primary CD4<sup>+</sup> T cells and Jurkat cells. Across thousands of proviruses, host factors alone supported a restricted baseline, with Tat-null proviruses producing ~4%–15% of wild-type protein output and remaining constrained under strong stimulation. Chromatin immunoprecipitation-seq profiling revealed that Tat is dispensable for promoter-proximal RNAPII engagement and pausing but required for efficient CDK9 recruitment, Ser2 phosphorylation, and productive elongation, defining a ceiling for Tat-independent transcription. Tat deficiency reduced overall RNAPII occupancy without increased promoter-proximal accumulation. Genome-wide mapping using Tat-null controls revealed broad Tat association with active host loci and modest increases in elongation-factor occupancy; however, host-gene RNA gains were small (~1.05–1.49×) and poorly correlated with Tat binding. Instead, elongation-associated chromatin features better predicted responsiveness. Together, these findings define a framework separating host-driven transcription from Tat-dependent scaling and explain Tat's disproportionate potency at the provirus.

## Graphical abstract



## Introduction

Integrated HIV-1 proviruses operate at the interface between host and viral transcriptional control, where layered regulatory mechanisms determine whether proviral genomes remain silent or become transcriptionally active. Host transcription factors and chromatin regulators establish the initial conditions for RNA polymerase II (RNAPII) recruitment and promoter activity at the viral 5' long terminal repeat (5'LTR), whereas the viral transactivator Tat subsequently reprograms RNAPII dynamics to convert basal transcription into a highly processive and self-reinforcing transcriptional program. By binding the HIV transactivation response (TAR) RNA element, Tat recruits P-TEFb and the Super Elongation Complex, promotes RNAPII pause release, and drives robust transcriptional elongation [1–5]. Through additional interactions with histone acetyltransferases and chromatin-remodeling complexes, Tat reshapes nucleosome architecture

at the LTR to further stabilize productive transcription [6–13]. Together, these mechanisms create a hierarchical coupling between host-driven initiation and Tat-dependent elongation, providing a molecular basis for the exceptional specificity, dynamic range, and magnitude of HIV transcriptional amplification [14–16].

Before Tat accumulation, HIV transcription is governed exclusively by host regulatory pathways. Cellular transcription factors such as NF-κB, NFAT, and AP-1, initiate low-level transcription at the viral promoter, generating the first Tat transcripts and enabling the transition to Tat-dependent elongation [17–20]. Although inhibition of Tat activity by small molecules or dominant-negative variants strongly suppresses viral reactivation, it fails to eliminate residual HIV RNA expression [21–24], revealing that Tat-independent transcription persists even under conditions of profound elongation blockade [25]. Despite decades of investigation, the

Received: February 19, 2026. Revised: May 6, 2026. Accepted: June 4, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

quantitative magnitude and mechanistic basis of host-driven transcription from chromatin-integrated proviruses remain poorly defined, particularly in native genomic contexts where proviral chromatin architecture and integration-site diversity shape transcriptional potential.

Consequently, the physiological extent of Tat-independent HIV transcription has never been systemically quantified across diverse integration sites in native viral contexts. Most previous studies have relied largely on LTR-only reporter systems, which fail to recapitulate proviral chromatin organization and integration-site heterogeneity [26–29]. Consequently, the intrinsic capacity of host transcriptional machinery to drive proviral transcription, and the extent to which Tat selectively amplifies this activity relative to host genes, have remained unresolved. A fundamental quantitative question therefore persists: what is the maximal level of HIV transcription that host factors alone can sustain in a native chromatin context where Tat-TAR coupling is preserved?

Here, we address this gap by adapting a widely used dual-fluorescent HIV reporter construct, HIV<sub>GKO</sub> [30–33], to generate an isogenic Tat-null HIV<sub>GKO</sub> reporter that preserves the native LTR and proviral genome while eliminating Tat expression. By integrating this platform with RNAII-directed chromatin immunoprecipitation (ChIP)-seq and transcriptomic analysis, we quantitatively define the physiological baseline of host-driven HIV transcription across thousands of independent integration sites and dissect the mechanistic contribution of Tat to transcriptional initiation and elongation. We further assess the extent to which Tat binding to host chromatin is associated with functional transcriptional regulation. Together, our results reveal that host factors alone support only a limited, quantitatively constrained level of HIV transcription and establish Tat as a selective and disproportionately potent amplifier of transcriptional elongation.

## Materials and methods

### Cell lines and cell culture

HEK293T cells (ATCC) were cultured in high-glucose Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Atlas Biologicals, #FS-0500-AD) and penicillin-streptomycin-glutamine (PSG; penicillin 100 U/ml, streptomycin 100 µg/ml, L-glutamine 2 mM; Thermo Fisher, #10378016). Jurkat CD4<sup>+</sup> T cells and derived HIV-integrated populations were maintained in RPMI 1640 (Thermo Fisher, #11875093) supplemented with 10% FBS and PSG. Cells were routinely tested for mycoplasmas contamination and maintained at 37°C in a humidifier incubator with 5% CO<sub>2</sub>.

Primary naïve CD4<sup>+</sup> T cells were isolated from HIV-1-negative donors using magnetic negative selection (Naïve CD4<sup>+</sup> T Cell Isolation Kit II, Miltenyi Biotec, #130-094-131). Cells were differentiated into central memory-like CD4<sup>+</sup> T cells as previously described [34, 35]. Briefly, cells were activated with anti-CD3/CD28 Dynabeads (Thermo Fisher, #11131D, 1:1 cell-to-bead ratio) in the presence of anti-IL-4, anti-IL-12, and TGF-β1 [1 µg/ml, 2 µg/ml, and 10 ng/ml, respectively (Peprotech)]. After bead removal, cells were cultured in RPMI supplemented with IL-2 (30 IU/ml).

### HIV<sub>GKO</sub> constructs and viral production

The HIV<sub>GKO</sub> dual-reporter virus (Addgene, #112234) was used as the wild-type (WT) backbone. Tat-deficient derivatives were generated by site-directed mutagenesis. The TatLeak mutant was generated by substituting the Tat start site codon (ATG to ACG) to reduce Tat translation without disrupting overlapping Vpr's open reading frames. The TatNull construct was generated by introducing two premature stop codons downstream of the Tat coding sequence while preserving Vpr coding potential. All constructs verified by Sanger sequencing to confirm precise mutagenesis.

VSV-G-pseudotyped HIV<sub>GKO</sub> constructs were produced in HEK293T cells by co-transfection of HIV<sub>GKO</sub> plasmids and pCMV-VSV-G (Addgene, #8454) using TransIT®-2020 Transfection Reagent (Mirus Bio, #MIR 5404). Viral supernatants were collected at 48–72 h, filtered (0.45 µm), and virus titers determined by infection in Jurkat CD4<sup>+</sup> T cells for 3 days.

### Luciferase reporter assay to assess residual Tat activity

HeLa cells stably expressing an LTR-luciferase reporter (HeLa-LTR-Luc) were seeded in 12-well plates and transfected with HIV<sub>GKO</sub> constructs or control plasmid mKO2-N1 (Addgene #54625) using TransIT-LT1 (Mirus #MIR2306) according to the manufacturer's instructions. Transfection efficiency was monitored by mKO2 fluorescence and used for normalization of luciferase signals. Cells were harvested 48 h post-transfection and lysed in 1× passive lysis buffer (Promega, #E194A). Luciferase activity was quantified using the Luciferase Assay System (Promega, #E1501) on a Berthold luminometer and normalized to total protein content determined by Bradford assay (Bio-Rad #5000006). Data are presented as relative light units. All experiments were performed in triplicate.

### Viral infection and generation of integrated populations

Jurkat cells were infected with TatWT, TatLeak, and TatNull at multiplicity of infection (MOI) of ~0.2, a low-MOI condition used to enrich for single proviral integration and preserve integration-site heterogeneity across the population. Integrated populations were generated by sorting mKO2<sup>+</sup> cells using flow cytometry (BD FACSAria Fusion). For primary CD4<sup>+</sup> T cells, infection was performed on day 7 of differentiation by spinoculation (1740 × g, 2 h, 37°C) using 500 ng p24 per 10<sup>6</sup> cells. Infected cells were analyzed every 3 days by flow cytometry.

### Cellular stimulation and pharmacological treatments

Cells were stimulated with phorbol 12-myristate 13-acetate [PMA 10 nM (Fisher, #BP6851)]. Trichostatin A [TSA 2 µM (Sigma, #T-1952)], TNF-α (20 ng/ml, Cell Signaling Technology, #89025C), or JQ1 (10 µM, MedchemExpress, #HY-13030) for the indicated times. Cell counts and viability were determined using trypan blue exclusion and a Cellometer Auto T4 (Nexcelom).

**Table 1.** Primers used for quantitative RT-PCR

| Target         | Forward oligos           | Reverse oligos         |
|----------------|--------------------------|------------------------|
| Total HIV mRNA | GGTTAGACCAGATCTGAGCCTGG  | CAACAGACGGGCACACACTACT |
| Gag mRNA       | GCAATGAGCCAAGTAACAAATCCA | CCTTTTTCCTAGGGGCCCTGC  |
| Tat/Rev mRNA   | TCTATCAAAGCAACCCACCTC    | CGTCCCAGATAAGTGCTAAGG  |
| RPL13A         | GCCCTACGACAAGAAAAAGCG    | TACTTCCAGCCAACCTCGTGA  |

### Flow cytometry

Primary T cell activation was assessed by surface staining for CD25 (BD Biosciences, #560503) and CD69 (BD Biosciences, #561928). Cell cycle analysis was performed by fixation followed by staining with propidium iodide (PI) using FxCycle™ PI staining solution (Molecular probes, #F10797) and quantification of DNA content by flow cytometry. Intracellular HIV p24 staining was performed using an APC-conjugated anti-p24 antibody (MediMabs, MM-0289) following fixation with formaldehyde-containing buffer (BD Biosciences, 51-2090KZ) and permeabilization in Perm/Wash Buffer (BD Biosciences, 554723#). Data acquisition was performed on BD LSRFortessa or FACSCanto instruments. In Jurkat T cells, GFP expression was used as a readout of HIV transcription, while mKO2 expression was used to identify cells harboring integrated proviruses. In primary CD4<sup>+</sup> T cells, infected (integrated) cells were defined as GFP<sup>+</sup> or mKO2<sup>+</sup>; as previously described [31]. Mean fluorescence intensity (MFI) and frequencies of GFP<sup>+</sup> cells within the mKO2<sup>+</sup> population were quantified. Data were analyzed using FlowJo v10, and statistical analyses were performed using GraphPad Prism.

### Western blotting

Cells were lysed in RIPA buffer supplemented with protease inhibitors (Roche, #04293132001). Protein concentrations were determined by Bradford assay. Equal amounts of protein were resolved by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred to PVDF membranes. Membranes were probed with antibodies against Tat (Thermo Fisher, #MA1-71509), p55 Gag (BEI resources, #ARP-13417), and HDAC1 as loading control (Thermo Fisher, #PA1-860), followed by HRP-conjugated secondary antibodies. Signals were detected using ECL™ Prime Luminol Enhancer solutions (Cytiva) and quantified using Image Lab software (Bio-Rad Laboratories).

### RNA isolation and quantitative reverse transcriptase-polymerase chain reaction

Total RNA was extracted using RNeasy kit (Qiagen, #74106) and treated with DNase (Invitrogen, #AM1907). Complementary DNA was synthesized using random hexamer and SuperScript III First Strand Synthesis kit (Invitrogen, #18080051). Quantitative polymerase chain reaction (PCR) was performed using SensiFAST™ SYBR® No-ROX Kit (BioLine, #BIO-98020). HIV transcripts were normalized to RPL13A expression (Table 1) using the  $\Delta\Delta C_t$  method.

### Chromatin immunoprecipitation and library preparation

ChIP was performed as previously described [36]. Cells were crosslinked with 1% formaldehyde for 10 min and quenched with 0.125 M glycine. Chromatin was sonicated to 200–

400 bp fragments (Qsonica, #Q800R3 sonicator). A total of 20  $\mu$ g chromatin was used for 2  $\mu$ g antibody of Tat (Thermo Fisher, #MA1-71509), CDK9 (Protintech, #11705-1-AP), RNAPII (Active Motif #61801), S2P-RNAPII (Active Motif, #61803), S5P-RNAPII (Active Motif, #61805), or matched IgG controls, mouse IgG (Millipore # NI03), rabbit IgG (Fisher Scientific, #02-6102), and rat IgG (Thermo Fisher, #31933). The equivalent of 5% chromatin was saved as input. DNA was purified, treated with RNase A (Thermo Scientific, #FEREN0531), reverse crosslinked, and then digested with proteinase K (Fisher Scientific # BP1700100). The DNA was purified using a PCR clean kit (Qiagen, #28106).

ChIP enrichment was validated by qPCR prior to sequencing as previously described [35, 37]. ChIP DNA libraries were prepared using the NEB Next Ultra II DNA Library Prep Kit (NEB, #E7103) and indexed using unique dual indices. Libraries were sequenced on an Illumina platform with paired-end reads. Unless otherwise stated, ChIP-seq experiments were performed with two independent biological replicates per condition.

### ChIP-seq data processing, normalization, and peak calling

Sequencing quality was assessed using FastQC and MultiQC. Reads were trimmed using Trim Galore and aligned to a custom GRCh38 reference genome containing the HIV<sub>GKO</sub> provirus as an additional chromosome using Bowtie2 (very-sensitive preset and a maximum fragment length of 2000 bp). Alignments were sorted/indexed with samtools; duplicate reads were removed using Picard. Genome-wide signal tracks were generated using deepTools bamCoverage with RPGC normalization to enable quantitative comparison across samples, using hg38 effective genome size ( $2.91 \times 10^9$ ) with 10 bp bins and read extension while ignoring marked duplicates.

Peak calling was performed using MACS2 in paired-end-mode (-f BAMPE) with narrow peak detection and a false discovery rate (FDR) threshold of  $q < 0.05$  using matched input controls. Overlap analysis were performed using bedtools intersect; enrichment against randomized background was assessed using bedtools shuffle excluding blacklist regions.

TatNull-corrected Tat occupancy. To account for nonspecific background and transcription-coupled association, Tat ChIP-seq analyses at host genes used TatNull as an empirical background. For metagene and gene-level analyses, TatWT RPGC signal was corrected by subtracting the corresponding TatNull RPGC signal on a per-region basis prior to aggregation (referred to as “TatNull-corrected” Tat signal).

### Tat chromatin occupancy classification

Tat-bound genes were classified as Tat<sup>−</sup>, Tat<sup>+</sup>, or Tat<sup>++</sup> based on Tat ChIP-seq peak overlap and enrichment.

## Metagene profiling and gene-level ChIP-seq quantification

For metagene analysis, signal matrices were computed using deepTools computeMatrix in scale-region mode with 6 kb upstream, a scale gene body, and 6 kb downstream using 25 bp bins. PCGs and lncRNAs were analyzed separately and subdivided into Tat<sup>-</sup>, Tat<sup>+</sup>, and Tat<sup>++</sup> classes.

Gene-level ChIP-seq signal was quantified using deepTools multiBigwigSummary on preferred promoter and gene-body windows. Promoter regions were defined as TSS-50 to TSS+350 nt, and gene bodies were defined as TSS+350 nt to transcription end site (TES), thereby excluding promoter-proximal pausing regions from gene-body measurements. Violin plots and summary statistics were generated using ggplot2. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests with multiple-testing correction where indicated.

## Initiation and elongation indices at the HIV provirus

The initiation index was calculated as the ratio of S5P to total RNAPII signal at the HIV promoter (TSS-50 to TSS+350 nt) [38, 39]. The elongation index (EI) was calculated as S2P signal across the HIV gene body (TSS+350 nt to TES) normalized to the promoter RNAPII occupancy [40–42]. Indices were computed from RPGC-normalized signals and averaged across biological replicates.

## RNA-seq processing, quantification, and differential expression analysis

Total RNA libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit and sequenced on an Illumina platform. Raw paired-end RNA-seq reads were assessed using FastQC and summarized with MultiQC. Adapter trimming and quality filtering were performed using Cutadapt. Trimmed reads were quantified using Salmon in quasi-mapping mode against a combined transcriptome (GENCODE v44 human transcripts supplemented with HIV<sub>GKO</sub> transcripts). Transcript abundances were summarized to gene-level counts using tximport with lengthScaledTPM.

Gene-level counts were analyzed using DESeq2. Size factors were estimated using the median-of-ratios method; dispersions were fitted using the default model. Differential expression testing was performed using Wald tests with Benjamini–Hochberg FDR correction.

## Tat-dependent stimulation metrics and integrative analyses

$\Delta$ -signal (TatWT<sub>stim</sub> versus TatNull<sub>stim</sub>). For Fig. 4 and related analysis,  $\Delta$ -signal was defined as  $\log_2$  (TatWT<sub>stim</sub> / TatNull<sub>stim</sub>).

$\Delta\Delta$  (interaction-based Tat-dependent stimulation gain). To quantify Tat-dependent stimulation gain, a DESeq2 intersection framework was used with genotypes (TatWT versus TatNull), stimulation (Stim versus NS), and their interaction. For each gene, Tat-dependent stimulation gain ( $\Delta\Delta$ ) was defined as:  $\Delta\Delta_{\text{RNA}} = (\text{TatWT}_{\text{stim}} - \text{TatWT}_{\text{NS}}) - (\text{TatNull}_{\text{stim}} - \text{TatNull}_{\text{NS}})$ , using  $\log_2$  (normalized counts + 1). This interaction-based framework isolates the Tat-specific component of the stimulation response by controlling for basal expression differences and Tat-independent stimulation effects.  $\Delta\Delta_{\text{RNA}}$  values were used for gene-level and

population-level analyses. The HIV provirus was excluded from host-gene statistical testing and shown only as a scale reference where indicated. Genes with a significant interaction term (adjusted  $P < .05$  and  $|\log_2$  fold change|  $\geq 0.5$ ) were defined as Tat-dependent stimulation-response genes for downstream analyses.

Quartile-based intersection analysis (UpSet). For Fig. 6D, Tat-bound genes were ranked independently by  $\Delta\Delta_{\text{RNA}}$ ,  $\Delta\Delta_{\text{Tat}}$  gene-body occupancy, and  $\Delta\Delta_{\text{S2P}}$  gene-body occupancy, and the top quartile of each metric was selected. Intersections were computed to identify genes exhibiting coordinated Tat-dependent changes in Tat occupancy, elongation-associated chromatin features, and RNA output.

## Functional enrichment analysis

Gene Ontology enrichment analyses were performed using clusterProfiler (enrichGO). Background gene sets were defined as all annotated gene in org.Hs.eg.db. Adjusted  $P$ -values were calculated using the Benjamini–Hochberg method.

## Statistical analysis

Unless otherwise stated, data are presented as mean  $\pm$  s.e.m. Pairwise comparisons were performed using two-sided tests as appropriate [including Wilcoxon rank-sum tests for distribution comparisons and analysis of variance (ANOVA) where indicated]. Multiple-testing correction was applied using Benjamini–Hochberg methods where appropriate. All plots were generated using ggplot2 and custom R scripts. Sample sizes, replicates, and statistical tests are indicated in figure legends.

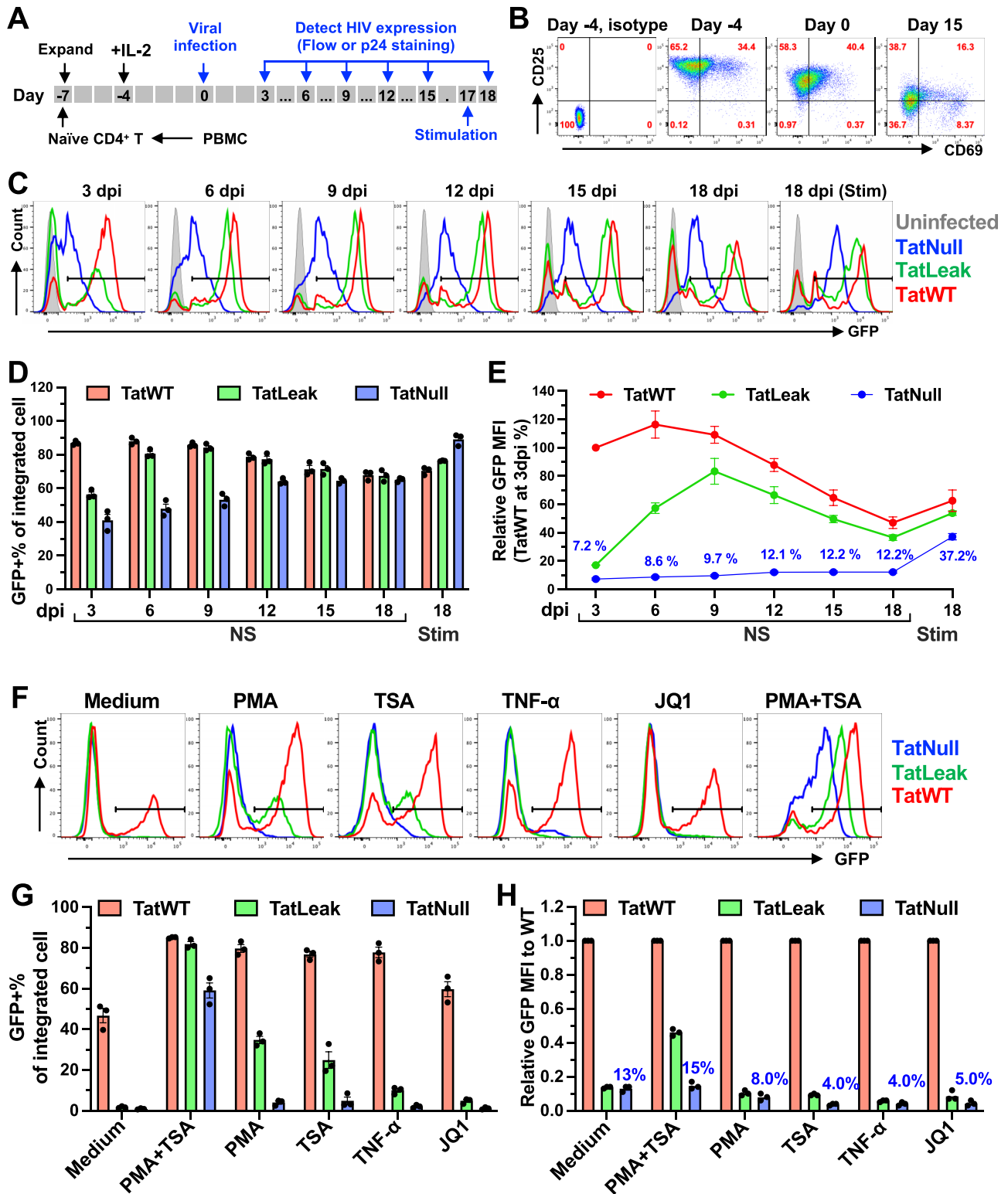
## Results

### Tat-independent transcription is quantitatively constrained across primary and CD4<sup>+</sup> T-transformed cells

To quantify host-driven HIV transcription in the absence of Tat, we engineered an isogenic Tat-deficient derivative of the HIV<sub>GKO</sub> dual-reporter virus [31] (Supplementary Fig. S1A). In this system, the HIV promoter (5'LTR) drives GFP expression as a readout of proviral transcription, while a constitutive EF1 $\alpha$  promoter drives mKO2 to mark cells harboring integrated proviruses independent of transcriptional activity (Supplementary Fig. S1A). To progressively disrupt Tat expression, we generated two Tat mutant viruses. Mutation of the Tat start codon (ATG to ACG; TatLeak) reduced Tat translation without perturbing overlapping viral reading frames (Vpr), whereas introduction of premature stop codons downstream of the Tat coding sequence generated a true Tat-null virus (TatNull), without altering Vpr coding potential (Supplementary Fig. S1A). In HeLa-LTR-luciferase reporter assays [36], TatLeak retained partial Tat transactivation activity, whereas TatNull exhibited no detectable Tat activity (Supplementary Fig. S1B), confirming a complete ablation.

To investigate Tat-dependent transcription in a physiologically relevant context, we employed a widely used primary CD4<sup>+</sup> T-cell infection model in which naïve cells are activated and expanded prior to infection [34, 43], followed by longitudinal monitoring of viral expression (Fig. 1A). Flow cytometric analysis of activation markers confirmed that cells remained in an activated or transitioning state at the time of infection (Fig. 1B), consistent with a transcriptionally





**Figure 1.** Tat-independent HIV transcription is quantitatively constrained in primary CD4<sup>+</sup> T and Jurkat cells. **(A)** Schematic of experimental workflow for activation, infection with TatWT, TatLeak, or TatNull HIV<sub>GKO</sub> viruses, and longitudinal monitoring of HIV expression. **(B)** Flow cytometry analysis of activation markers in uninfected cells. Representative CD25 versus CD69 staining at indicated time points (day -4 and day 0 relative to infection, and day 15 post-infection), with corresponding isotype controls. **(C)** Representative GFP fluorescence profiles of infected primary central memory CD4<sup>+</sup> T cells at indicated days post-infection (dpi). GFP expression was analyzed within the integrated (mKO2<sup>+</sup>) population. Quantification of proviral transcription in primary T cells measured as the fraction of GFP<sup>+</sup> cells **(D)** and GFP MFI normalized to TatWT at 3 dpi **(E)** across time points under non-stimulated (NS) and stimulated (10 nM PMA + 2  $\mu$ M TSA) conditions. **(F)** Representative GFP histograms of Jurkat cells infected with TatWT, TatLeak, or TatNull HIV<sub>GKO</sub> viruses following treatment with indicated latency-reversing agents. On 4 dpi, cells were treated overnight with 10 nM PMA, 2  $\mu$ M TSA, 20 ng/ml TNF, 10  $\mu$ M JQ1, or the combination 10 nM PMA + 2  $\mu$ M TSA. Quantification of GFP<sup>+</sup> frequency **(G)** and GFP MFI normalized to WT **(H)** in Jurkat cells under indicated stimulation conditions (16 h on 4 dpi). Data in panels (D–G) represent mean  $\pm$  s.e.m. with individual biological replicates shown as points.

permissive chromatin environment. Given prior reports that the constitutive EF1 $\alpha$  promoter can be partially silenced in primary T cells [31, 44], we directly compared GFP-based detection of HIV expression with intracellular p24 staining. At 3 dpi, cells infected with TatWT, TatLeak, and TatNull HIV<sub>GKO</sub> viruses were analyzed by flow cytometry using GFP, p24, and mKO2 signals (Supplementary Fig. S1C). Within the integrated cell population, defined as GFP<sup>+</sup> or mKO2<sup>+</sup> (and independently as p24<sup>+</sup> or mKO2<sup>+</sup>) [31], GFP and p24 measurements were highly concordant, both in the fraction of expressing cells (GFP<sup>+</sup> versus p24<sup>+</sup>) and in relative MFI (Supplementary Fig. S1C and D). Together, these results validate GFP as a robust and quantitative reporter for HIV gene expression in this system.

Using this framework, TatWT, TatLeak, and TatNull HIV<sub>GKO</sub> viruses exhibited comparable integration efficiencies in primary central memory CD4<sup>+</sup> T cells, as indicated by similar frequencies of integrated cells at early time points (Supplementary Fig. S1E and F), thereby excluding differential infection as a confounding factor. Proviral transcription was therefore quantified by GFP expression within the integrated population (Fig. 1C). Over time, however, the fraction of integrated cells declined more rapidly in TatWT infections compared to TatNull, consistent with a selective loss of cells expressing higher levels of viral gene products (Supplementary Fig. S1F). In line with this, TatWT proviruses rapidly reached high GFP expression, whereas the TatLeak and TatNull proviruses exhibited progressively reduced transcriptional output (Fig. 1C and D). TatLeak proviruses showed an early defect that partially recovered toward WT levels over time. In contrast, TatNull proviruses displayed a sustained reduction in transcriptional output despite a gradual increase in the proportion of GFP<sup>+</sup> cells (Fig. 1C and D).

To correct for the time-dependent loss of highly expressing TatWT cells, GFP MFI values were normalized to TatWT at 3 dpi (Fig. 1E). Using this approach, TatNull proviruses exhibited consistently low transcriptional output, reaching only ~7%–12% of TatWT levels across multiple time points (Fig. 1D). These results define a substantial, yet quantitatively constrained, Tat-independent transcription output driven by host factors alone, demonstrating that Tat is dispensable for proviral integration and basal transcriptional engagement but essential for robust transcriptional amplification. The distinct phenotypes of TatLeak and TatNull viruses further demonstrate that even minimal Tat expression is sufficient to restore near-WT transcriptional output, whereas complete Tat loss reveals a stable quantitative limit of host-driven transcription.

To determine whether strong activation can compensate for Tat loss, infected primary CD4<sup>+</sup> T cells were stimulated with PMA and TSA at 17 days post infection (dpi). Consistent with a pre-activated primary T-cell state, stimulation only modestly increased gene expression from TatWT proviruses, but markedly enhanced gene transcription from TatNull proviruses, with GFP MFI increasing ~3-fold. When normalized to TatWT at 3 dpi, this corresponded to ~37% of TatWT levels (Fig. 1C–E). Despite this increase, TatNull expression remained substantially below TatWT, even under strong activating conditions, revealing a quantitative ceiling for Tat-independent transcription.

We next examined Tat-independent transcription in Jurkat T cells using low-MOI infections (Supplementary Fig. S1G). Under basal conditions, TatNull proviruses exhibited minimal

GFP expression, whereas TatWT were transcriptionally active (Fig. 1E and F). Single latency-reversing agents induced modest activation of TatWT and TatLeak proviruses but had little effect on TatNull proviruses. Strikingly, combined PMA and TSA stimulation activated a large fraction of TatNull-infected cells, increasing the proportion of GFP<sup>+</sup> to >60%; however, GFP MFI remained limited to ~15% of TatWT levels, comparable to the ~13% observed under basal conditions (Fig. 1F and G). These findings demonstrate that strong upstream signaling can initiate transcription in the absence of Tat but cannot substitute for Tat-dependent transcriptional amplification.

Together, these results establish that host transcriptional machinery can engage the HIV promoter and initiate transcription, but that Tat-independent transcription is quantitatively constrained to ~4%–15% of WT output across cell types. The TatLeak intermediate phenotype further underscores the exquisite sensitivity of HIV transcriptional output to Tat dosage and validates TatNull as a stringent reference for defining the physiological baseline of host-driven proviral transcription.

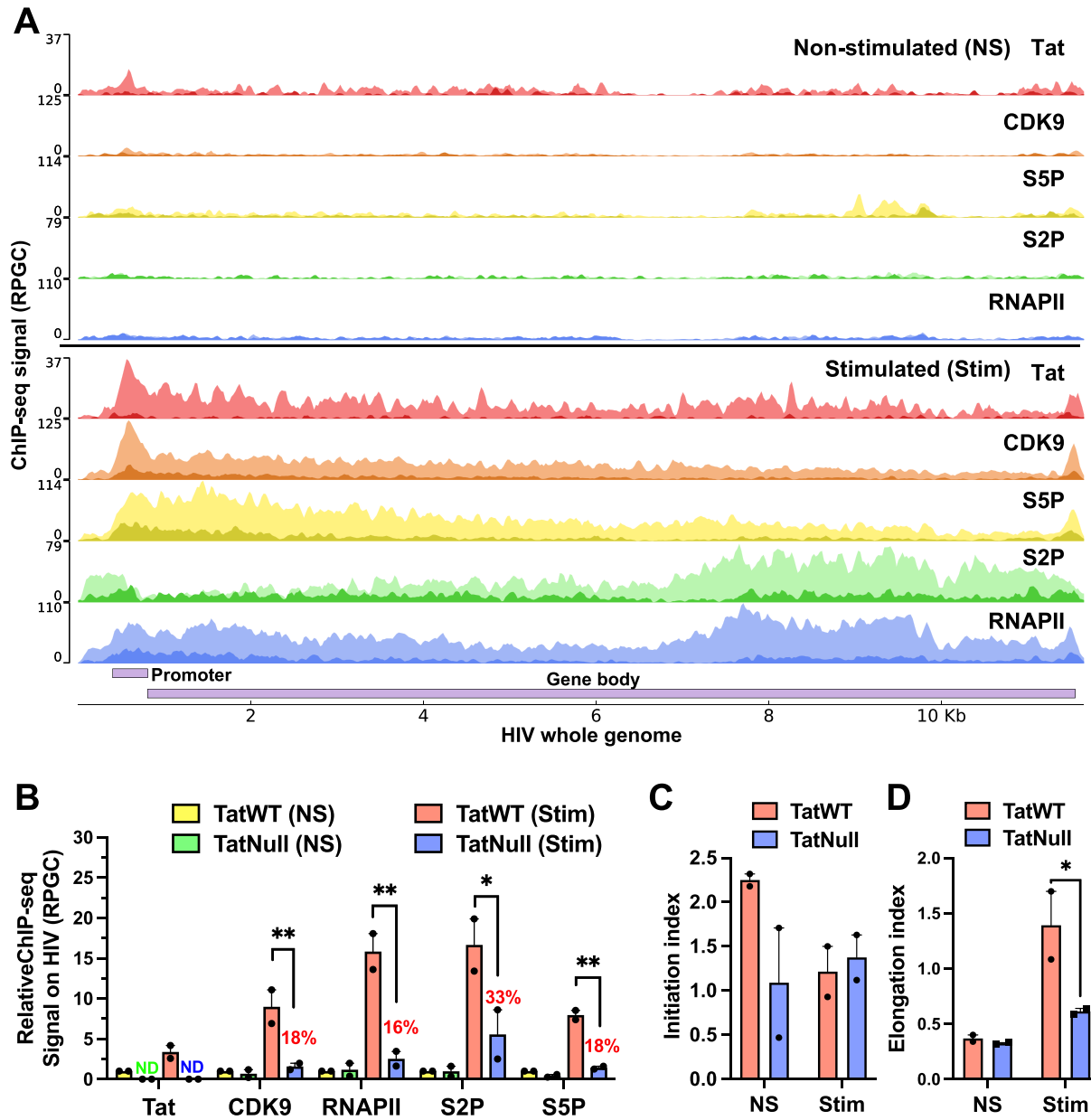
### Tat is required for efficient CDK9 and RNAPII recruitment and productive HIV elongation

To define the contribution of Tat to proviral transcription, we generated stable Jurkat cell populations harboring either TatWT or TatNull proviruses (Supplementary Fig. S2A). GFP analysis under NS and PMA + TSA-stimulated (Stim) conditions confirmed markedly reduced transcriptional activity in TatNull cells (Supplementary Fig. S2A). Consistently, total HIV and Gag transcripts were substantially reduced in TatNull cells (~6% of TatWT under NS and ~11% upon stimulation), while Tat/Rev transcripts showed a similar reduction at baseline (~6%) but increased to ~30% of TatWT upon stimulation (Supplementary Fig. S2B), suggesting partial recovery of spliced transcripts under strong activation.

Western blot analysis confirmed the absence of Tat and reduced Gag expression in TatNull cells (Supplementary Fig. S2C), and Sanger sequencing verified the introduced mutations (Supplementary Fig. S2D). Cell cycle analysis by propidium iodide (PI) staining showed no major differences between TatWT and TatNull populations (Supplementary Fig. S2E), indicating that the observed effects are not attributable to altered cell cycle distribution.

We next performed ChIP-seq profiling of Tat, CDK9, total RNAPII, RNAPII-Ser5P (S5P), and RNAPII-Ser2P (S2P) under basal and PMA + TSA-stimulated conditions. Genome browser visualization revealed minimal basal occupancy of Tat, CDK9, and RNAPII across the provirus in both genotypes (Fig. 2A). Upon stimulation, WT proviruses exhibited robust recruitment of Tat, CDK9, RNAPII, S2P, and S5P across both promoter and gene body, with Tat enrichment extending beyond the TAR region and closely paralleling elongation-associated factors. In contrast, TatNull proviruses displayed stimulus-induced recruitment of RNAPII and CDK9, but at substantially lower magnitude and with reduced gene-body occupancy (Fig. 2A). Consistent with this, both total RNAPII and RNAPII-Ser5P signals were globally reduced in TatNull proviruses, indicating diminished overall polymerase engagement.

Quantitative analysis of RPGC-normalized ChIP-seq signal across the provirus confirmed these patterns (Fig. 2B). Stimulation markedly increased CDK9, total RNAPII, S5P, and S2P



**Figure 2.** Tat is required for efficient CDK9 recruitment and productive HIV transcriptional elongation. **(A)** Genome browser tracks showing RPGC-normalized ChIP-seq signal for Tat, CDK9, total RNAPII, RNAPII-Ser5P (S5P), and RNAPII-Ser2P (S2P) across the integrated HIVGKO provirus in TatWT and TatNull Jurkat cells under NS and stimulated (PMA + TSA) conditions (TatWT: light color, TatNull: dark color). **(B)** Quantification of mean ChIP-seq signal across the HIVGKO provirus for each factor in TatWT and TatNull cells under NS and stimulated conditions, normalized to WT NS. Percent values indicate TatNull signal as a fraction of TatWT under stimulated conditions; ND, not detected. Points indicate biological replicates. **(C, D)** Initiation and elongation indices at the HIV promoter and gene body, respectively (defined in “Materials and methods” section), quantifying promoter engagement and productive elongation. Data in panels (B–D) are shown as mean  $\pm$  s.e.m. Statistical significance is indicated in the figure and assessed using two-way ANOVA; \* $P < .05$ , \*\* $P < .01$ .

occupancy on TatWT proviruses, whereas TatNull proviruses exhibited consistently attenuated recruitment of all marks. This reduction extended to promoter-proximal RNAPII and S5P signals, indicating reduced polymerase engagement at the promoter in the absence of Tat, although direct assessment of pre-initiation complex (PIC) assembly was not performed. Under stimulated conditions, total RNAPII occupancy on TatNull proviruses reached  $\sim 16\%$  of TatWT levels, closely mirroring the residual transcriptional output observed in Fig. 1.

To dissect distinct transcriptional phases, we calculated initiation and elongation indices (Fig. 2C and D). The initia-

tion index, defined as promoter-proximal S5P relative to total RNAPII [38, 39], was comparable between TatWT and TatNull proviruses and showed minimal change upon stimulation. In TatWT proviruses, the modest decrease in initiation index following stimulation likely reflects redistribution of RNAPII from the promoter into the gene body associated with enhanced elongation, rather than reduced initiation. These data indicate that Tat is largely dispensable for promoter-proximal RNAPII and paused polymerase distribution, despite reduced overall RNAPII levels. In contrast, the EI, defined as gene-body S2P relative to promoter RNAPII

[40–42], increased ~3.8-fold in TatWT proviruses but only ~1.8-fold in TatNull proviruses following stimulation. Thus, although TatNull proviruses retain the capacity for RNAPII recruitment and transcription initiation, they exhibit reduced RNAPII occupancy and a marked defect in P-TEFb-dependent pause release and productive elongation.

Together, these data demonstrate that Tat is not required for establishing promoter-proximal RNAPII engagement but enhances overall RNAPII occupancy and is essential for efficient CDK9-dependent productive elongation, providing a mechanistic basis for the quantitative ceiling of Tat-independent HIV transcription.

### Tat associates broadly with host chromatin but is quantitatively enriched at a restricted gene subset

Although Tat is classically viewed as a provirus-specific transcriptional regulator, we asked whether Tat also associates with host chromatin on a genome-wide scale. To identify bona fide Tat-binding sites, TatNull samples were used as a background control. Replicate correlation analysis revealed clear separation between TatWT and TatNull ChIP-seq profiles, with increased concordance after stimulation and tight clustering of TatNull samples at background levels (Supplementary Fig. S3A). Using this TatNull-controlled framework, Tat peaks were independently called in TatWT and TatWT + Stim samples, identifying 1468 high-confidence Tat binding sites genome-wide in the stimulated TatWT condition.

Tat binding co-localized extensively with transcriptional machinery. Between 40% and 65% of Tat peaks overlapped CDK9, total RNAPII, S5P, or S2P, representing >100-fold enrichment relative to shuffled genomic controls (Supplementary Fig. S3B). Tat peak intensity correlated most strongly with CDK9 and total RNAPII occupancy, and more modestly with RNAPII phosphorylation states (Supplementary Fig. S3C), consistent with Tat engagement at transcriptionally active loci and its established interaction with P-TEFb.

Consistent with Tat's preferential association with active transcription complexes [45, 46], increased Tat abundance upon stimulation dramatically expanded the repertoire of Tat-bound genes, increasing from 318 loci at baseline to 1267 after stimulation, with 112 shared between conditions (Fig. 3A). Tat peaks were distributed across promoters, introns, and downstream regions, and stimulation increased Tat occupancy proportionally across genomic features rather than redistributing binding to specific regions (Fig. 3B). Among stimulated Tat-bound genes, PCGs ( $n = 894$ ) predominated, followed by long noncoding RNAs (lncRNAs;  $n = 267$ ) (Fig. 3C).

Metagene analysis (TatNull-corrected, RPGC-normalized signal) revealed a characteristic bimodal Tat distribution at PCGs, with enrichment at both transcription start sites (TSSs) and TESs, which was enhanced under increased Tat abundance after stimulation and accompanied by increased genebody signal (Fig. 3D). Tat-bound lncRNAs exhibited broader promoter-proximal enrichment with similar Tat abundance-dependent increases (Fig. 3E). Thus, increased Tat availability primarily increased the magnitude of Tat association rather than altering its positional pattern along genes.

Functional annotation of Tat-bound PCGs revealed strong enrichment for chromatin-associated processes, including chromatin organization, nucleosome assembly, and epigenetic regulation (Supplementary Fig. S4A). Tat-bound lncR-

NAs displayed more limited functional enrichment, largely restricted to RNA metabolic and ribonucleoprotein-associated processes (Supplementary Fig. S4B), consistent with heterogeneous lncRNA function and incomplete annotation. Stratification by Tat signal intensity uncovered a distinct subset of genes with disproportionately high Tat occupancy (Tat<sup>++</sup>;  $n = 81$ ; Supplementary Table S1), exhibiting  $\geq 2$ -fold higher Tat binding in TatWT + Stim compared to TatNull + Stim (Fig. 3F). These Tat<sup>++</sup> genes were strongly enriched for chromatin regulatory pathways (Fig. 3G). Notably, HIV served as an internal reference point, highlighting the exceptional magnitude of Tat recruitment at the provirus relative to host genes. A corresponding Tat<sup>++</sup> lncRNA subset ( $n = 54$ ; Supplementary Table S2) was also identified (Supplementary Fig. S4C), although functional interpretation was limited by sparse annotation.

Together, these results reveal that Tat associates broadly with host chromatin yet exhibits pronounced quantitative selectivity, with maximal recruitment at the HIV provirus and a restricted subset of host genes.

### Tat-bound host genes exhibit elevated RNAPII and CDK9 occupancy independently of Tat, with Tat-dependent amplification upon stimulation

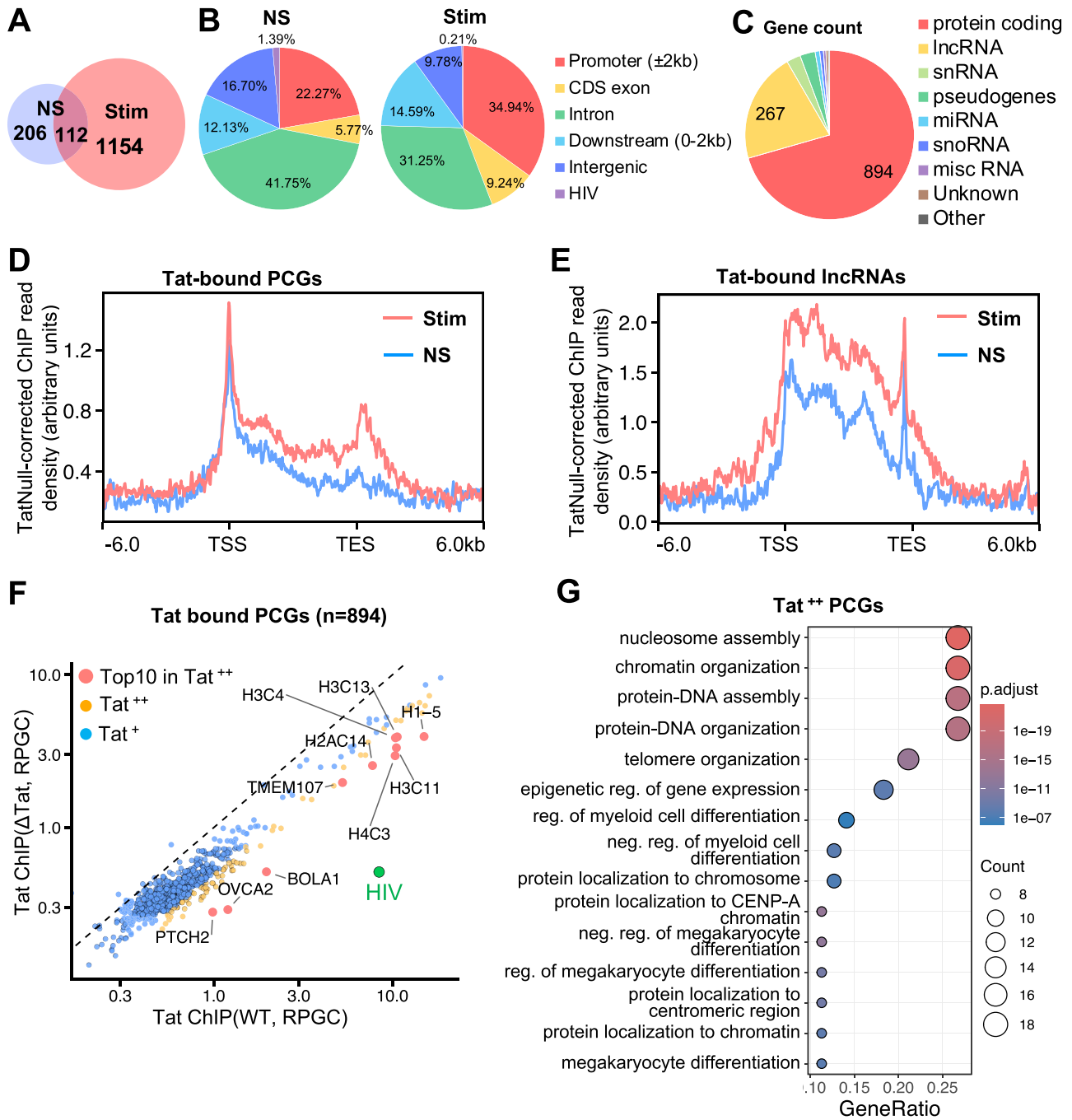
To determine whether Tat preferentially associates with transcriptionally active chromatin, we generated length- and GC-content-matched control gene sets lacking detectable Tat (Tat<sup>-</sup>) (Supplementary Fig. S5A–C). Genes were then stratified into three classes: Tat<sup>-</sup> (no Tat peaks), Tat<sup>+</sup> (Tat-bound genes), and Tat<sup>++</sup> (Tat-bound genes with  $\geq 2$ -fold higher Tat ChIP-seq signal in WT + Stim relative to TatNull + Stim).

Metagene profiling revealed a stepwise increase in RNAPII occupancy from Tat<sup>-</sup> to Tat<sup>+</sup> to Tat<sup>++</sup> PCGs under unstimulated conditions in both TatWT and TatNull cells (Fig. 4A). Because TatNull cells lack Tat, this hierarchy reflects intrinsic transcriptional engagement at Tat-associated loci rather than a direct Tat effect. CDK9, S2P, and S5P exhibited a similar Tat<sup>-</sup> < Tat<sup>+</sup> < Tat<sup>++</sup> ordering across genotypes (Supplementary Fig. S6A–C), indicating that Tat preferentially associates with chromatin environments already characterized by active transcriptional machinery. Gene-level quantification corroborated these trends: median RNAPII signal increased progressively across Tat classes in all conditions and in both genotypes (Fig. 4B). Stimulation with PMA + TSA globally redistributed RNAPII and reduced the relative separation between classes, yet Tat<sup>+</sup> and Tat<sup>++</sup> genes remained consistently more RNAPII-rich than Tat<sup>-</sup> controls, and this stratification persisted in TatNull samples, reinforcing that they are not established by Tat. These patterns were mirrored in metagene profiles of CDK9, S2P, and S5P (Supplementary Fig. S7).

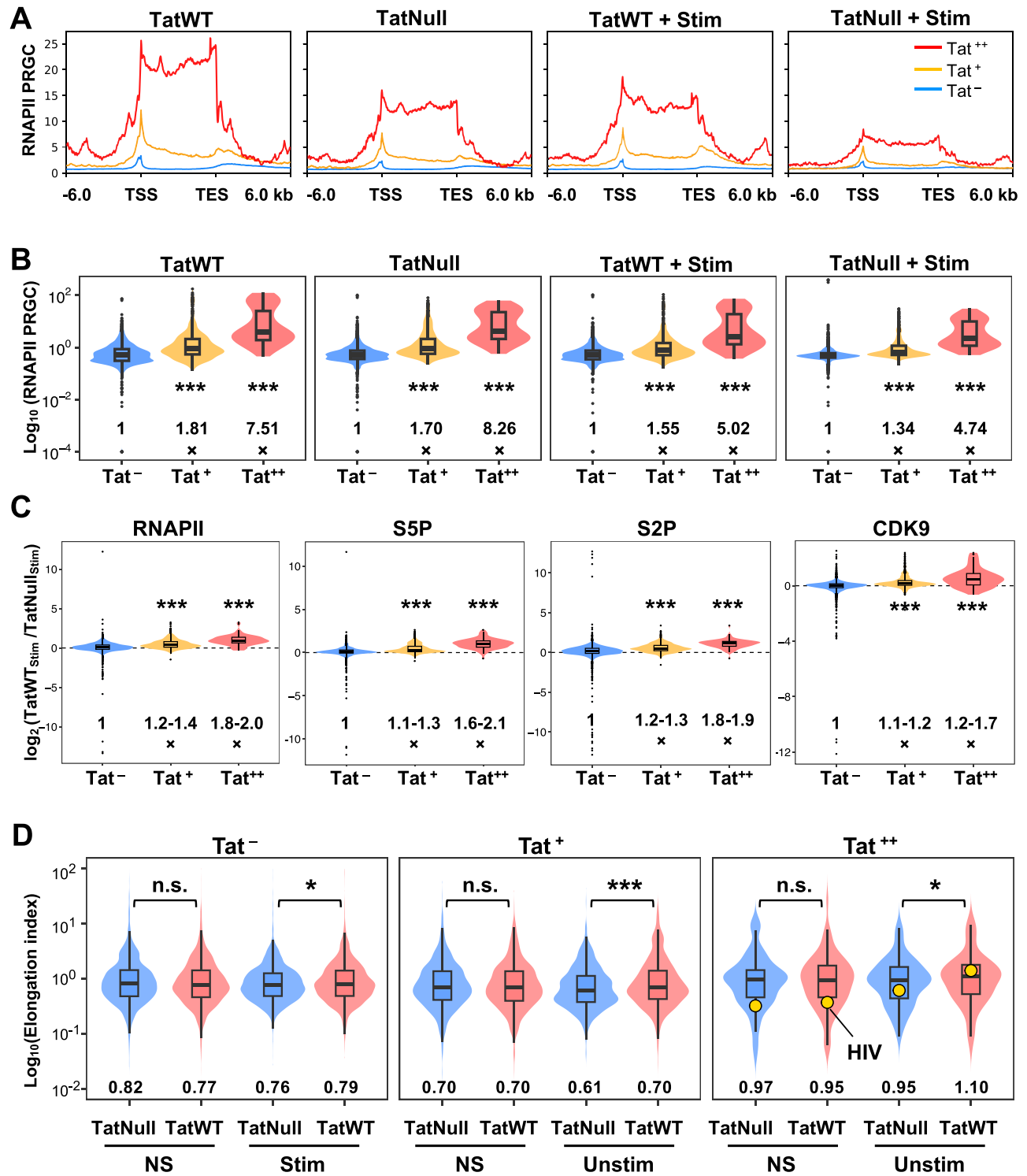
To isolate Tat-dependent effects while controlling for global stimulation-induced transcription redistribution, we computed per-gene  $\Delta$ -signals as  $\log_2$  (TatWT + Stim / TatNull + Stim), using TatNull + Stim as matched Tat-deficient reference. Tat<sup>+</sup> and Tat<sup>++</sup> PCGs exhibited significantly higher RNAPII, S2P, S5P, and CDK9  $\Delta$ -signals than Tat<sup>-</sup> genes (Fig. 4C), with RNAPII increases of ~1.3-fold and ~1.85-fold, respectively. Thus, Tat does not create transcriptional competence *de novo* but instead amplifies recruitment and activity of elongation-associated transcriptional machinery at pre-engaged loci.

We next asked whether these increases translated into greater elongation efficiency by calculating an EI





**Figure 3.** Genome-wide mapping identifies Tat-bound host genes and a highly Tat-enriched subset. **(A)** Overlap of Tat-bound genes detected in unstimulated (NS) and PMA + TSA-stimulated (Stim) TatWT cells. **(B)** Genomic distribution of Tat ChIP-seq peaks across promoters, exons, introns, downstream regions, intergenic regions, and the HIV provirus under NS and Stim conditions. **(C)** Biotype composition of Tat-bound host genes ( $n = 1267$ ), including PCGs ( $n = 894$ ) and long noncoding RNAs (lncRNAs,  $n = 267$ ). **(D, E)** Metagene profiles of TatNull-corrected, RPGC-normalized Tat ChIP-seq signal across Tat-bound PCGs ( $n = 894$ ) and lncRNAs ( $n = 267$ ) under NS and Stim conditions. Signal is averaged across scaled gene bodies with  $\pm 6$  kb flanking regions. **(F)** Gene-level comparison of Tat ChIP-seq signal between TatWT and TatNull cells across Tat-bound PCGs. Tat<sup>+</sup> ( $n = 894$ ), Tat<sup>++</sup> ( $n = 81$ ), and representative Tat<sup>++</sup> genes are indicated; the HIV provirus is shown as a reference point (green). **(G)** Gene Ontology enrichment of Tat<sup>++</sup> PCGs ( $n = 81$ ). Dot size indicates gene counts and color indicates Benjamini-Hochberg-adjusted  $P$  values.



**Figure 4.** Tat-associated host genes exhibit elevated RNAPII and CDK9 loading independent of RNA output. **(A)** Metagene profiles of RNAPII ChIP-seq signal across PCGs stratified by Tat occupancy (Tat<sup>-</sup>, Tat<sup>+</sup>, Tat<sup>++</sup>), shown for TatWT and TatNull cells under NS and stimulated (PMA + TSA) conditions. Profiles are averaged across scaled gene bodies with  $\pm 6$  kb flanking regions. **(B)** Gene-level RNAPII occupancy across Tat classes in TatWT and TatNull cells. Violin plots show log<sub>10</sub> RPK-normalized, RPK-normalized RNAPII signal over gene bodies. Boxes denote the interquartile range with the median indicated. Numbers below indicate median fold enrichment relative to Tat<sup>-</sup> within the same condition. **(C)** Stimulation-induced changes in chromatin occupancy by Tat class. Per-gene  $\Delta$ -signal values are shown as violin plots for RNAPII, S5P, S2P, and CDK9 across Tat classes. Values below each group indicate the interquartile range (25th–75th percentile) fold change of TatWT/TatNull, normalized to Tat<sup>-</sup> (linear scale). **(D)** EI across PCGs stratified by Tat occupancy (Tat<sup>-</sup>, Tat<sup>+</sup>, Tat<sup>++</sup>) in TatWT and TatNull cells under NS and Stim conditions. EI was calculated as the ratio of RNAPII S2P signal over the gene body to promoter signal (S2P<sub>body</sub> / S2P<sub>promoter</sub>) using RPK-normalized ChIP-seq data and is shown on a log<sub>10</sub> scale. Violin plots show gene-level EI distributions with median and interquartile range indicated, and the bottom values indicate the median EI of the group. The HIV provirus EI is overlaid as a single point within the Tat<sup>++</sup> group, calculated using the same metric. Statistical significance was assessed using Wilcoxon rank-sum tests (n.s. no significance, \* $P < 1 \times 10^{-3}$ , \*\*\* $P < 1 \times 10^{-10}$ ).

(gene-body/promoter S2P) (Fig. 4D). Under unstimulated conditions, EI values were indistinguishable between TatWT and TatNull cells across all Tat binding classes, indicating that basal elongation efficiency at host genes is largely Tat independent. Following stimulation and Tat accumulation, TatWT cells showed modest EI increases at Tat-associated genes relative to TatNull, most pronounced at Tat<sup>+</sup> loci. Tat<sup>++</sup> genes exhibited a similar trend with reduced statistical significance, likely due to already elevated basal EI limiting dynamic range. In contrast, Tat<sup>-</sup> genes displayed only minor genotype-dependent differences, reflecting with global stimulation effects rather than direct Tat regulation. This modest EI response stands in sharp contrast to the HIV provirus, where EI rose dramatically from comparable WT and TatNull values under basal conditions (0.37 versus 0.32) to a markedly higher EI in TatWT upon stimulation (1.4) (Fig. 2D). Thus, although Tat-dependent increases in S2P occupancy are evident at Tat-bound loci (Fig. 4C), the absence of a focused Tat-TAR recruitment mechanism at host genes constrains how efficiently these increases in RNAPII and P-TEFb occupancy are converted to productive elongation.

Tat-bound lncRNAs displayed analogous behavior. Tat<sup>+</sup> and Tat<sup>++</sup> lncRNAs showed elevated RNAPII occupancy in both TatWT and TatNull cells under basal conditions (Supplementary Fig. S8A and B), consistent with a permissive chromatin state. Upon Tat accumulation by stimulation, Tat<sup>+</sup> and Tat<sup>++</sup> lncRNAs exhibited significantly increased RNAPII, S2P, S5P, and CDK9  $\Delta$ -signals relative to Tat<sup>-</sup> controls (Supplementary Fig. S8C), yet EI changes were minimal (Supplementary Fig. S8D), suggesting distinct elongation constraints at lncRNAs loci.

Together, these results support a model in which Tat preferentially associates with transcriptionally active chromatin landscapes and selectively amplifies RNAPII and P-TEFb recruitment. Baseline host-driven transcription establishes a permissive state onto which Tat imposes a secondary amplification layer, mirroring the architecture at the HIV provirus, but with far greater elongation control at the virus locus.

### Tat selectively enhances expression at Tat-associated host genes, but with modest magnitude compared to HIV

To determine whether Tat chromatin occupancy translates into functional regulation of host transcription, we analyzed RNA-seq profiles from TatWT and TatNull-infected cells under NS and stimulated conditions (PMA + TSA). Genes were stratified by Tat ChIP-seq binding (Tat<sup>-</sup>, Tat<sup>+</sup>, Tat<sup>++</sup>; Fig. 4). Restricting the analysis to genes with detectable expression yielded 2442 Tat<sup>-</sup>, 823 Tat<sup>+</sup>, and 78 Tat<sup>++</sup> PCGs, representing the subset of Tat-bound loci that are transcriptionally active.

Across all conditions, Tat-associated PCGs (Tat<sup>+</sup> and Tat<sup>++</sup>) displayed higher absolute expression levels than Tat<sup>-</sup> genes, including in NS and TatNull infected cells (Fig. 5A). This indicates that Tat binding preferentially occurs at loci already embedded in transcriptionally permissive chromatin rather than establishing basal transcription. Consistent with this interpretation, differences in gene expression between TatWT and TatNull under NS conditions were minimal across all gene classes (median TatWT/TatNull fold change: Tat<sup>-</sup> = 1.03 $\times$ , Tat<sup>+</sup> = 1.08 $\times$ , Tat<sup>++</sup> = 1.13 $\times$ ; Fig. 5B), demon-

strating that Tat is largely dispensable for basal host gene expression.

In contrast, basal HIV transcription showed a marked Tat dependence. On the same scale used for host genes, the TatWT-TatNull separation at the provirus was 1.92, corresponding to an  $\sim$ 3.8-fold difference (Fig. 5B), highlighting a fundamental distinction between viral and host transcriptional control (Fig. 5B).

Following Tat accumulation upon stimulation, Tat-associated genes exhibited a modest but progressive shift toward higher expression in TatWT relative to TatNull cells, proportional to Tat occupancy (median TatWT/TatNull fold change: Tat<sup>-</sup> = 0.99 $\times$ , Tat<sup>+</sup> = 1.10 $\times$ , Tat<sup>++</sup> = 1.23 $\times$ ; Fig. 5B). In contrast, HIV transcription showed a dramatically larger TatWT-TatNull separation of 3.45 ( $\sim$ 11-fold difference), underscoring the disproportionate dependence of viral transcription on Tat.

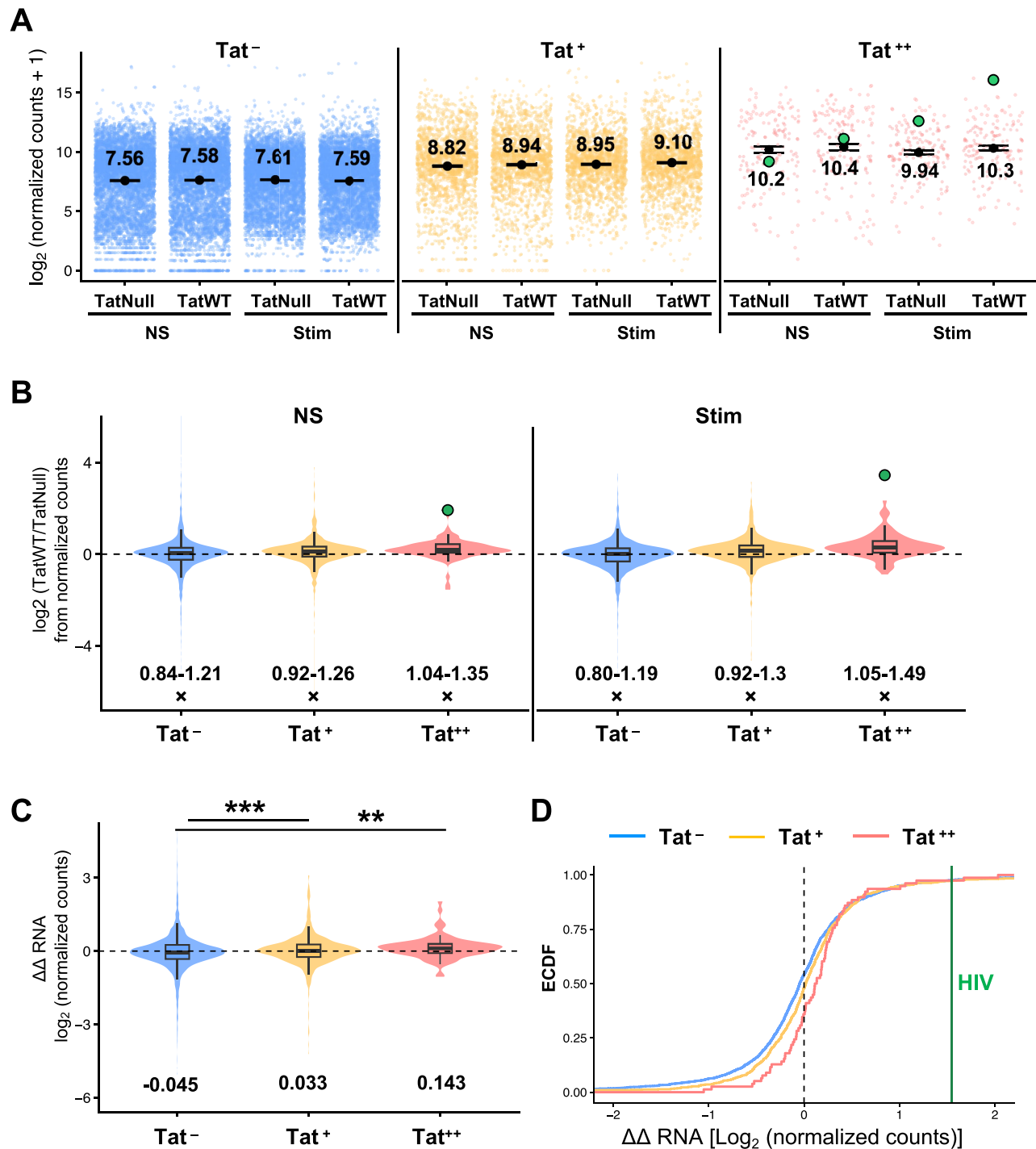
To quantify Tat-dependent amplification during stimulation, we defined a Tat-dependent stimulation gain metric:  $\Delta\Delta\text{RNA} = (\text{TatWT}_{\text{stim}} - \text{TatWT}_{\text{NS}}) - (\text{TatNull}_{\text{stim}} - \text{TatNull}_{\text{NS}})$ . Tat<sup>-</sup> genes exhibited no Tat-dependent gain (median = 0.045), whereas Tat<sup>+</sup> and Tat<sup>++</sup> genes displayed small but significant increases (medians 0.033 and 0.143; Fig. 5C), with cumulative distributions shifting progressively rightward from Tat<sup>-</sup> to Tat<sup>+</sup> to Tat<sup>++</sup> (Fig. 5D). By comparison, HIV exhibited a  $\Delta\Delta\text{RNA}$  of 1.54, over an order of magnitude larger than host genes. Because HIV already displays substantial Tat dependence at baseline, this metric underestimates Tat's full effect at the provirus; HIV is shown as a scale reference rather than a direct comparator (Fig. 5D).

lncRNAs showed a qualitatively similar pattern. Tat-associated lncRNAs had higher basal expression than Tat<sup>-</sup> loci (Supplementary Fig. S9A), with minimal TatWT-TatNull differences under NS conditions (Supplementary Fig. S9B). Following Tat accumulation by stimulation, Tat<sup>+</sup> and Tat<sup>++</sup> lncRNAs showed modest WT-biased shifts and limited Tat-dependent gains (Supplementary Fig. S9C and D), with greater heterogeneity than PCGs but similarly small effect sizes.

Collectively, these results demonstrate that Tat preferentially associates with host genes embedded in transcriptionally permissive chromatin and confers only modest amplification of their expression. In contrast, Tat exerts a quantitatively dominant effect at the HIV promoter, revealing a fundamental asymmetry between viral and host transcriptional regulation: Tat acts primarily as a specialized viral amplifier superimposed on a broadly permissive host transcriptional landscape.

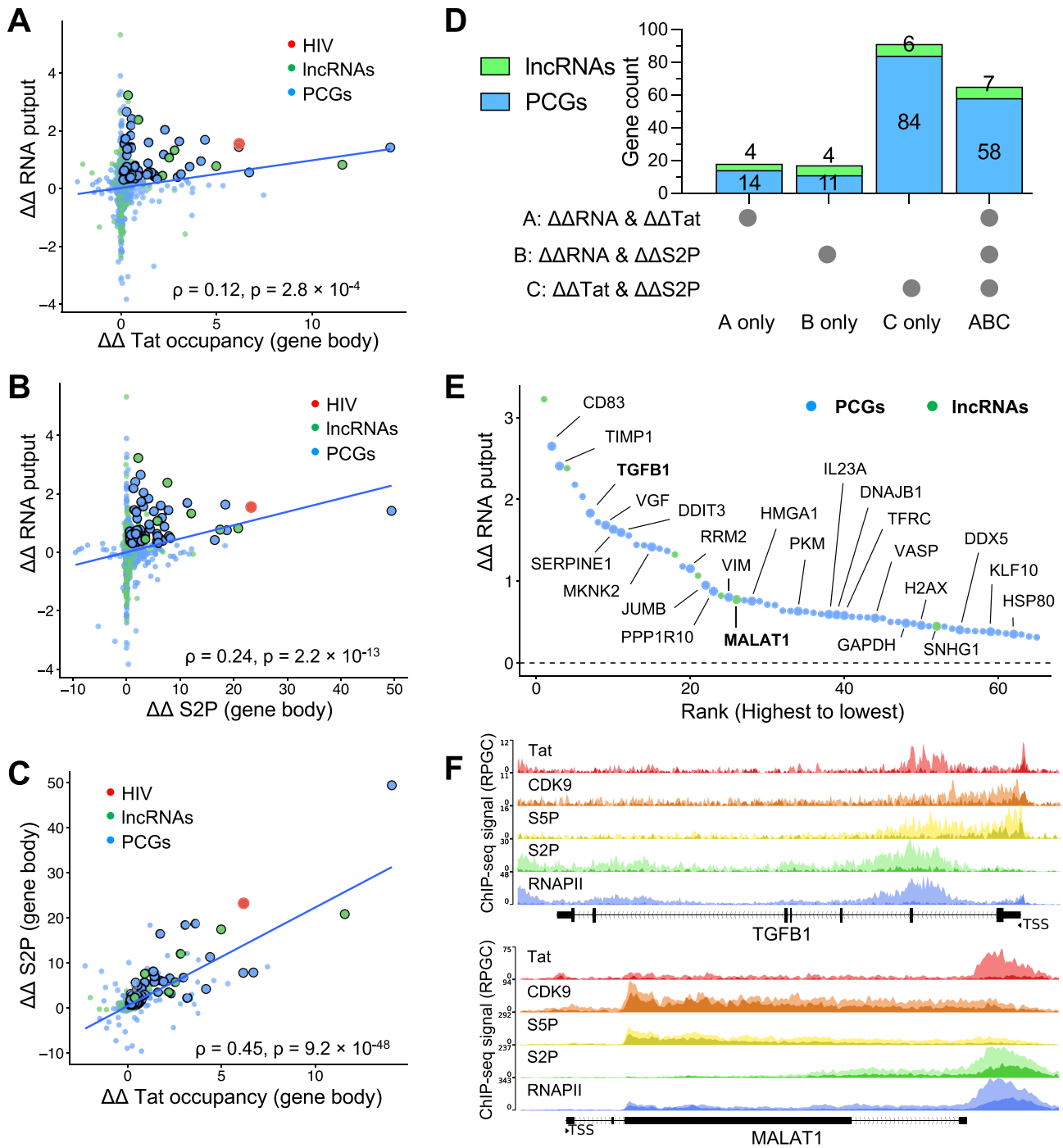
### Integrated ChIP-seq and RNA-seq reveal coordinated Tat-associated elongation with limited RNA amplification at host genes

To link Tat-associated chromatin features with transcriptional output, we integrated ChIP-seq and RNA-seq using a  $\Delta\Delta$  framework that isolates the Tat-specific component of the transcriptional response under matched stimulation conditions. RNA-seq analysis was restricted to expressed Tat-bound loci, yielding 978 Tat-bound loci (834 PCGs and 100 lncRNAs). For each gene,  $\Delta\Delta\text{RNA}$  quantified the additional RNA gain in TatWT relative to TatNull cells after controlling for basal transcription and Tat-independent activation. HIV was included as a reference point for scale.



**Figure 5.** Tat selectively enhances stimulation-dependent transcription at Tat-associated host genes. **(A)** Gene-level expression of host PCGs stratified by Tat chromatin occupancy. Genes were classified as Tat-negative (Tat<sup>-</sup>,  $n = 3461$ ), Tat-bound (Tat<sup>+</sup>,  $n = 894$ ), or high-confidence Tat-associated (Tat<sup>++</sup>,  $n = 65$ ) based on Tat ChIP-seq signal. Expression is shown as log<sub>2</sub>(normalized counts + 1) in TatWT and TatNull cells under NS and stimulated (PMA + TSA; Stim) conditions. Each dot represents an individual gene-sample measurement, whereas  $n$  denotes the number of unique genes in each Tat group. Black points and horizontal bars indicate the mean  $\pm$  SEM within each group, with mean values shown above. The HIV provirus is shown as a single green dot for reference and was excluded from host gene quantification and statistical analyses. **(B)** Gene-level Tat dependence of basal and stimulated conditions. Violin plots depict per-gene log<sub>2</sub> (TatWT/TatNull) values for PCGs grouped by Tat occupancy (Tat<sup>-</sup>, Tat<sup>+</sup>, Tat<sup>++</sup>) under NS and Stim conditions. Overlaid boxplots mark the median and interquartile range; the dashed line indicates no Tat dependence (log<sub>2</sub> = 0). Values below each group indicate the interquartile range (25th–75th percentile) fold change of TatWT/TatNull, normalized to Tat<sup>-</sup> (linear scale). **(C)** Tat-dependent stimulation gain ( $\Delta\Delta$ RNA) at the gene level. Violin plots show the distribution of  $\Delta\Delta$ RNA values stratified by Tat occupancy. Boxplots indicate median and interquartile range; the dashed line denotes  $\Delta\Delta = 0$  (no Tat-dependent stimulation gain). Statistical comparisons between Tat<sup>-</sup> and Tat<sup>+</sup> or Tat<sup>++</sup> genes were performed using two-sided Wilcoxon rank-sum tests with Benjamini–Hochberg correction (\*\* $P < .01$ , \*\*\* $P < .001$ ). **(D)** Population-level distribution of Tat-dependent stimulation gain. Empirical cumulative distribution functions of  $\Delta\Delta$  values are shown for PCGs grouped by Tat occupancy. The vertical dashed line indicates  $\Delta\Delta = 0$ . The solid green line marks the  $\Delta\Delta$  value of HIV for scale reference only and was not included in host-gene distributions or statistical analyses. Analysis in panels (B–D) were performed on the same Tat-stratified PCGs set defined in panel (A).





**Figure 6.** Integrated chromatin and transcriptional analyses reveal coordinated Tat-associated elongation and RNA output at a subset of host genes. **(A–C)** Relationships between stimulation-dependent Tat occupancy, elongation-associated RNAPII modification, and RNA output across Tat-bound host genes ( $n = 932$  total genes; 832 PCGs and 100 lncRNAs). Scatter plots show  $\Delta\Delta$  Tat gene-body occupancy versus  $\Delta\Delta$  RNA (A),  $\Delta\Delta$  S2P gene-body occupancy versus  $\Delta\Delta$  RNA (B), and  $\Delta\Delta$  Tat versus  $\Delta\Delta$  S2P (C). Each point represents a Tat-bound PCG or lncRNA; the HIV provirus is shown in as a reference point. Linear regression lines are shown for visualization. Spearman correlations were computed using host genes only (HIV excluded). Genes belonging to the intersection set defined in panel (D) are highlighted as enlarged point in panels (A–C). **(D)** Intersection of Tat-regulated host genes exhibiting coordinated increases in Tat occupancy, elongation-associated chromatin features, and RNA output. Genes were ranked by  $\Delta\Delta$  RNA,  $\Delta\Delta$  Tat, and  $\Delta\Delta$  S2P, and the top quartiles were intersected. The UpSet plot shows set membership (dots) and gene counts per category (bars), stratified by gene biotype (PCGs and lncRNAs). Genes shared across all three features (A + B + C) define a subset with coordinated Tat-associated elongation and transcriptional gain. **(E)** Ranked distribution of host genes within the A + B + C intersection set. Genes are ordered by  $\Delta\Delta$  RNA (highest to lowest), with PCGs and lncRNAs indicated as in panels (A–C). Selected genes previously implicated in HIV-related pathways are highlighted. The dashed horizontal line indicates  $\Delta\Delta$  RNA = 0. **(F)** Representative genome browser views of TGFB1 and MALAT1 showing RPKC-normalized ChIP-seq signal for Tat, CDK9, RNAPII, S5P, and S2P, illustrating Tat-associated recruitment of elongation machinery at host genes.

Across host genes, increases in Tat occupancy within gene bodies ( $\Delta\Delta\text{Tat\_body}$ ) showed only a weak association with  $\Delta\Delta\text{RNA}$  (Spearman  $\rho = 0.12$ ; Fig. 6A), and promoter-proximal Tat occupancy was not predictive of RNA gain (Supplementary Fig. S10A). Thus, Tat binding alone is insufficient to explain host transcriptional output. By contrast, elongation-associated chromatin features were more informative: Tat-dependent increases in RNAPII-S2P phosphorylation across gene bodies ( $\Delta\Delta\text{S2P\_body}$ ) correlated with  $\Delta\Delta\text{RNA}$  ( $\rho = 0.24$ ; Fig. 6B) and were strongly coupled to  $\Delta\Delta\text{Tat\_body}$  ( $\rho = 0.45$ ; Fig. 6C). Similar relationships were observed for CDK9 and total RNAPII occupancy (Supplementary Fig. S10B–D). These results place Tat upstream of elongation-associated RNAPII modification, rather than directly upstream of RNA output.

To identify loci with coordinated Tat-dependent regulation, we ranked Tat-bound genes by  $\Delta\Delta\text{RNA}$  and intersected the top quartiles of  $\Delta\Delta\text{RNA}$ ,  $\Delta\Delta\text{Tat\_body}$ , and  $\Delta\Delta\text{S2P\_body}$ . This revealed a distinct subset of 65 genes (A + B + C), predominantly protein-coding, enriched for loci previously reported to change during HIV infection or to influence the viral life cycle (Fig. 6D and E). *TGFB1* and *MALAT1* illustrate this pattern [47, 48]: both show stimulation-dependent Tat recruitment together with increased CDK9 and RNAPII occupancy across gene bodies, consistent with Tat-associated engagement of elongation machinery at host genes (Fig. 6F).

Together, these findings demonstrate that Tat recruitment at host genes is tightly coupled to Tat-dependent increases in S2P occupancy that only partially predict RNA output, consistent with limited elongation gains observed at these genes. In contrast, HIV transcription operates in a distinct regime where Tat-dependent elongation is directly translated into disproportionately large RNA amplification. Tat therefore functions not as a global regulator of host transcription, but as a specialized viral amplifier operating within an already permissive host transcriptional landscape.

## Discussion

This study establishes a quantitative, chromatin-preserving framework to disentangle basal host-driven HIV transcription from Tat-dependent amplification. Using an isogenic Tat-Null HIV<sub>GKO</sub> system integrated with matched ChIP-seq and RNA-seq analyses, we define the physiological magnitude of Tat-independent transcription and its mechanistic relationship to transcriptional elongation. Across primary CD4<sup>+</sup> T cells and Jurkat T cells, Tat deficiency restricts proviral output to ~4%–15% of WT levels, accompanied by proportional reductions in RNAPII occupancy and elongation-associated chromatin marks. At the genome scale, Tat preferentially associates with transcriptionally engaged host loci enriched for RNAPII, CDK9, and elongation signatures, yet its functional impact on host transcription is modest compared with its disproportionate effect at the HIV LTR. These findings support a model in which Tat operates primarily as a selective viral amplifier of transcriptional elongation rather than a global regulator of host gene expression.

Tat is dispensable for promoter engagement but is required for efficient transition into productive elongation. Importantly, Tat loss does not result in accumulation of promoter-proximal paused RNAPII; instead, TatNull proviruses exhibit reduced overall RNAPII occupancy together with a disproportionate defect in gene-body Ser2-

phosphorylated RNAPII. In TatWT proviruses, initiation and promoter-proximal RNAPII engagement are efficient, whereas stimulation-induced CDK9 recruitment and accumulation of Ser2-phosphorylated RNAPII across the gene body are markedly attenuated in TatNull cells. Consistent with this, integrative analyses of host genes indicate that elongation-associated chromatin features, rather than Tat occupancy per se, are the strongest predictors of stimulation-dependent RNA output. This placement of Tat upstream of elongation capacity aligns with its canonical role in recruiting P-TEFb and the super elongation complex to promote pause release [2, 3, 49–52]. Efficient elongation likely supports production of full-length genomic RNA and coordinated accumulation of multiply spliced transcripts required for viral gene expression.

The requirement for combined PMA + TSA stimulation further highlights the importance of chromatin accessibility in enabling Tat-independent transcription. Activation of a single pathway (e.g., NF- $\kappa$ B via TNF- $\alpha$ ) was insufficient, whereas dual stimulation was required, indicating that Tat-independent transcription arises from convergent regulatory inputs. In TatNull proviruses, limited RNAPII occupancy despite evidence of initiation suggests suboptimal promoter accessibility and/or inefficient PIC assembly. Concurrently, reduced CDK9 recruitment and Ser2 phosphorylation reflect constrained P-TEFb activity, potentially due to incomplete release from the 7SK complex [53, 54]. Together, these findings support a model in which cellular signaling and chromatin accessibility set the level of promoter engagement, while Tat is specifically required to convert these initiation events into productive elongation.

Previous Tat-null models based on single proviral integrations demonstrated the indispensability of Tat for sustained HIV transcription, but could not capture the integration-site diversity and chromatin heterogeneity that shape proviral behavior [55–60]. By quantifying transcription across thousands of independently integrated proviruses, including in acutely infected primary CD4<sup>+</sup> T cells, the TatNull HIV<sub>GKO</sub> platform defines the magnitude of Tat-independent transcription within a fully chromatinized genomic context. This quantitative resolution is particularly relevant for Tat-targeted therapeutic strategies, which suppress but do not abolish viral transcription and therefore require an accurate understanding of residual Tat-independent activity.

Importantly, interpretation of these measurements is influenced by cellular context. The primary CD4<sup>+</sup> T-cell system used here reflects an activated or transitioning state rather than a fully quiescent latency model, as bona fide latency in primary cells requires entry into a non-permissive, quiescent memory state with globally reduced transcriptional activity [61]. Accordingly, the basal transcription observed likely reflects ongoing chromatin engagement in an activated environment rather than true latency.

Although Tat is classically described as binding TAR RNA at the HIV promoter, our ChIP-seq analyses reveal Tat occupancy across the provirus and at a defined subset of host loci. At the provirus, Tat-signal is strongly TSS-centric, consistent with focused TAR-dependent recruitment. In contrast, Tat binding at host genes lacks a dominant peak and broadly overlaps regions enriched for RNAPII and elongation marks, indicating association with transcriptionally active chromatin rather than sequence-specific targeting. Given Tat's lack of intrinsic DNA-binding activity and its extensive interactions with RNAPII and P-TEFb [36, 49, 62], Tat is likely recruited to

host chromatin through protein–protein interactions. Within this framework, Tat associates with pre-engaged transcriptional complexes at host loci but exerts only modest effects on transcription output. In contrast, Tat recruitment at the HIV provirus is spatially focused and functionally dominant, revealing a fundamental asymmetry between viral and host transcriptional regulation. While Tat has been reported to interact with chromatin remodeling complexes such as PBAF/SWI-SNF [7], our genome-wide data indicate that Tat predominantly co-localizes with RNAPII-P-TEFb elongation complexes at both viral and host loci, with its functional impact dictated by the mode of recruitment: TAR-dependent engagement at the HIV LTR drives efficient and sustained transcriptional elongation, whereas association with RNAPII-P-TEFb at host genes results in more modest and context-dependent transcriptional amplification.

Reassessing Tat's influence on host transcription under physiological expression levels further clarifies this asymmetry. Endogenous Tat was detectable at a defined subset of host loci, predominantly PCGs, with fewer lncRNAs, consistent with earlier reports of Tat occupancy at active chromatin [45, 46]. However, Tat-associated transcriptional modulation was substantially smaller than previously reported in Tat overexpression systems [45], which likely distorted regulatory stoichiometry. Under physiological conditions, Tat confers only modest stimulation-dependent RNA gains at host loci and limited effects at lncRNAs, indicating that Tat does not broadly reprogram host transcription.

Although downstream phenotypes were not directly assessed, the selective upregulation of Tat-associated genes, particularly those involved in chromatin organization, stress responses, and cellular regulation, suggests potential functional consequences. For example, strongly Tat-responsive genes (e.g. TGFB1, SERPINE1) are linked to signaling, survival, and immune modulation pathways, which may influence cellular fitness and viral output between TatWT and TatNull conditions. However, given the modest and restricted nature of these effects, any phenotypic consequences are likely indirect and arise from cumulative changes in specific genes rather than broad reprogramming of cell identity. Further functional studies will be required to define these effects.

Several considerations inform these conclusions. Cellular stimulation induces global transcriptional remodeling that may compress Tat-dependent differences at host loci when assessed at the population level, and indirect effects of other viral proteins cannot be fully excluded. Nonetheless, by integrating chromatin occupancy with direct measurements of RNA output, our analyses provide a conservative and functionally grounded assessment of Tat-dependent regulation. This approach is particularly important for lncRNAs and other regulatory elements whose transcriptional consequences are difficult to infer from chromatin features alone.

A major strength of this work is the use of the isogenic Tat-Null HIV<sub>GKO</sub> platform, which preserves native proviral chromatin architecture while enabling quantitative dissection of Tat-dependent and Tat-independent transcription across heterogeneous integration sites. This system provides a generalizable framework for interrogating early viral transcription and host regulatory pathways that operate independently of Tat. For example, BRD4, a key modulator of HIV transcription, influences proviral output in both TatWT and TatNull contexts, thereby separating Tat-dependent from Tat-independent mechanisms of regulation [35, 63–65]. Such applications il-

lustrate how the TatNull platform can be leveraged to resolve host-virus regulatory interactions and identify host factors that sustain transcription in the absence of Tat.

Together, our findings refine current models of HIV transcription by defining the physiological magnitude of Tat-independent expression and by revealing a fundamental asymmetry between viral and host transcriptional regulation. Tat acts not as a broad host transcriptional regulator, but as a specialized viral amplifier that exploits pre-existing transcriptionally permissive chromatin to drive disproportionately large increases in proviral output. Identifying the host factors and chromatin regulators that sustain this residual transcription will be critical for understanding how HIV persists under conditions of partial transcriptional repression. The Tat-Null HIV<sub>GKO</sub> framework offers a versatile platform for probing early transcriptional regulation and informing strategies aimed at achieving durable transcriptional silencing of HIV.

## Acknowledgements

We thank the Flow Cytometry Core and the Genomics Core at The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology for expert technical support and assistance with data generation and analysis. We are also grateful to members of the Valente laboratory for helpful discussions and critical feedback throughout the study.

*Author contributions:* Chuan Li (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Software [lead], Supervision [lead], Validation [lead], Visualization [lead], Writing—original draft [equal], Writing—review & editing [equal]), Yuan Ma (Investigation [supporting], Methodology [supporting]), Yi Wang (Investigation [supporting], Methodology [supporting]), Huiming Yang (Investigation [supporting], Methodology [supporting]), and Susana Valente (Funding acquisition [lead], Project administration [equal], Resources [equal], Supervision [equal], Writing—review & editing [lead]).

## Supplementary data

Supplementary data is available at NAR online.

## Conflict of interest

None declared.

## Funding

This work was supported by the NIH/NIAID grants R33 AI140439, R37 AI165137, and the HOPE Martin Delaney Collaboratory UM1 AI164559. This work is also supported by the Miami Center for AIDS Research FY26 Pilot Awards (P30AI073961). Funding to pay the Open Access publication charges for this article was provided by UM1 AI164559.

## Data availability

Sequencing data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE318861. Data will be released upon publication.

## References

- Bowman EA, Kelly WG. RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: a tail of two kinases. *Nucleus* 2014;5:224–36. <https://doi.org/10.4161/nucl.29347>
- Sobhian B, Laguette N, Yatim A *et al.* HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell* 2010;38:439–51. <https://doi.org/10.1016/j.molcel.2010.04.012>
- He N, Liu M, Hsu J *et al.* HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol Cell* 2010;38:428–38. <https://doi.org/10.1016/j.molcel.2010.04.013>
- Wei P, Garber ME, Fang SM *et al.* A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* 1998;92:451–62. [https://doi.org/10.1016/S0092-8674\(00\)80939-3](https://doi.org/10.1016/S0092-8674(00)80939-3)
- Dayton AL, Sodroski JG, Rosen CA *et al.* The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* 1986;44:941–7. [https://doi.org/10.1016/0092-8674\(86\)90017-6](https://doi.org/10.1016/0092-8674(86)90017-6)
- Deymier MJ, Claiborne DT, Ende Z *et al.* Particle infectivity of HIV-1 full-length genome infectious molecular clones in a subtype C heterosexual transmission pair following high fidelity amplification and unbiased cloning. *Virology* 2014;468–470:454–61. <https://doi.org/10.1016/j.virol.2014.08.018>
- Rafati H, Parra M, Hakre S *et al.* Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. *PLoS Biol* 2011;9:e1001206. <https://doi.org/10.1371/journal.pbio.1001206>
- Treand C, du Chene I, Bres V *et al.* Requirement for SWI/SNF chromatin-remodeling complex in Tat-mediated activation of the HIV-1 promoter. *EMBO J* 2006;25:1690–9. <https://doi.org/10.1038/sj.emboj.7601074>
- Mahmoudi T, Parra M, Vries RG *et al.* The SWI/SNF chromatin-remodeling complex is a cofactor for Tat transactivation of the HIV promoter. *J Biol Chem* 2006;281:19960–8. <https://doi.org/10.1074/jbc.M603336200>
- Agbottah E, Deng L, Dannenberg LO *et al.* Effect of SWI/SNF chromatin remodeling complex on HIV-1 Tat activated transcription. *Retrovirology* 2006;3:48. <https://doi.org/10.1186/1742-4690-3-48>
- Lusic M, Marcello A, Cereseto A *et al.* Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. *EMBO J* 2003;22:6550–61. <https://doi.org/10.1093/emboj/cdg631>
- Marzio G, Tyagi M, Gutierrez MI *et al.* HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. *Proc Natl Acad Sci USA* 1998;95:13519–24. <https://doi.org/10.1073/pnas.95.23.13519>
- Benkirane M, Chun RF, Xiao H *et al.* Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat. *J Biol Chem* 1998;273:24898–905. <https://doi.org/10.1074/jbc.273.38.24898>
- Mbonye U, Karn J. Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. *Virology* 2014;454–455:328–39. <https://doi.org/10.1016/j.virol.2014.02.008>
- Ott M, Geyer M, Zhou Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe* 2011;10:426–35. <https://doi.org/10.1016/j.chom.2011.11.002>
- Yu K, Liu H, Pan T. HIV-1 Tat: molecular switch in viral persistence and emerging technologies for functional cure. *Int J Mol Sci* 2025;26:6311. <https://doi.org/10.3390/ijms26136311>
- Duverger A, Wolschendorf F, Zhang M *et al.* 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. *J Virol* 2013;87:2264–77. <https://doi.org/10.1128/JVI.01594-12>
- Wu BY, Woffendin C, Duckett CS *et al.* Regulation of human retroviral latency by the NF-kappa B/I kappa B family: inhibition of human immunodeficiency virus replication by I kappa B through a Rev-dependent mechanism. *Proc Natl Acad Sci USA* 1995;92:1480–4. <https://doi.org/10.1073/pnas.92.5.1480>
- Shukla A, Ramirez NP, D'Orso I. HIV-1 proviral transcription and latency in the new era. *Viruses* 2020;12:555. <https://doi.org/10.3390/v12050555>
- D'Orso I. The HIV-1 transcriptional program: from initiation to elongation control. *J Mol Biol* 2025;437:168690. <https://doi.org/10.1016/j.jmb.2024.168690>
- Li C, Mousseau G, Valente ST. Tat inhibition by didehydro-Cortistatin A promotes heterochromatin formation at the HIV-1 long terminal repeat. *Epigenetics Chromatin* 2019;12:23. <https://doi.org/10.1186/s13072-019-0267-8>
- Mousseau G, Kessing CF, Fromentin R *et al.* The Tat inhibitor didehydro-Cortistatin A prevents HIV-1 reactivation from latency. *mBio* 2015;6:e00465. <https://doi.org/10.1128/mBio.00465-15>
- Jin H, Sun Y, Li D *et al.* Strong *in vivo* inhibition of HIV-1 replication by Nullbasic, a Tat mutant. *mBio* 2019;10:e01769–19. <https://doi.org/10.1128/mBio.01769-19>
- Jin H, Li D, Sivakumaran H *et al.* Shutdown of HIV-1 transcription in T cells by Nullbasic, a mutant Tat protein. *mBio* 2016;7:e00518–16. <https://doi.org/10.1128/mBio.00518-16>
- Mousseau G, Aneja R, Clementz MA *et al.* Resistance to the Tat inhibitor didehydro-Cortistatin A is mediated by heightened basal HIV-1 transcription. *mBio* 2019;10:e01750–18. <https://doi.org/10.1128/mBio.01750-18>
- Aguiar-Cordova E, Chinen J, Donehower L *et al.* A sensitive reporter cell line for HIV-1 tat activity, HIV-1 inhibitors, and T cell activation effects. *AIDS Res Hum Retroviruses* 1994;10:295–301. <https://doi.org/10.1089/aid.1994.10.295>
- Li C, Wang HB, Kuang WD *et al.* Naf1 regulates HIV-1 latency by suppressing viral promoter-driven gene expression in primary CD4+ T cells. *J Virol* 2017;91:e01830–16. <https://doi.org/10.1128/JVI.01830-16>
- Jordan A, Bisgrove D, Verdin E. HIV reproducibly establishes a latent infection after acute infection of T cells *in vitro*. *EMBO J* 2003;22:1868–77. <https://doi.org/10.1093/emboj/cdg188>
- Chang LJ, Zhang C. Infection and replication of Tat- human immunodeficiency viruses: genetic analyses of LTR and tat mutations in primary and long-term human lymphoid cells. *Virology* 1995;211:157–69. <https://doi.org/10.1006/viro.1995.1388>
- Battivelli E, Dahabieh MS, Abdel-Mohsen M *et al.* Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4(+) T cells. *eLife* 2018;7:e34655. <https://doi.org/10.7554/eLife.34655>
- Battivelli E, Verdin E. HIV(GKO): a tool to assess HIV-1 latency reversal agents in human primary CD4(+) T cells. *Bio Protoc* 2018;8:e3050. <https://doi.org/10.21769/BioProtoc.3050>
- Naranjo O, Torices S, Clifford PR *et al.* AKT signaling modulates latent viral reservoir viability in HIV-1-infected blood-brain barrier pericytes. *J Biol Chem* 2024;300:105526. <https://doi.org/10.1016/j.jbc.2023.105526>
- de Armas LR, Gavegnano C, Pallikkuth S *et al.* The effect of JAK1/2 inhibitors on HIV reservoir using primary lymphoid cell model of HIV latency. *Front Immunol* 2021;12:720697. <https://doi.org/10.1128/JVI.01297-20>
- Sarabia I, Huang SH, Ward AR *et al.* The intact non-inducible latent HIV-1 reservoir is established in an *in vitro* primary T(CM) cell model of latency. *J Virol* 2021;95:e01297–20. <https://doi.org/10.1128/JVI.01297-20>
- Li C, Ma Y, Mori LP *et al.* The human chromatin remodeling complex p400 restricts HIV-1 transcription in a Tat-dependent manner. *Nucleic Acids Res* 2025;53:gakf1. <https://doi.org/10.1093/nar/gkaf1323>



36. Li C, Mori LP, Lyu S *et al.* The chaperone protein p32 stabilizes HIV-1 Tat and strengthens the p-TEFb/RNAPII/TAR complex promoting HIV transcription elongation. *Proc Natl Acad Sci USA* 2023;120:e2217476120. <https://doi.org/10.1073/pnas.2217476120>
37. Mori LP, Li C, Venables T *et al.* The p400 complex promotes HIV-1 latency by suppressing viral transcription and altering the host cell state. *Nucleic Acids Res* 2025;53:gkaf1323. <https://doi.org/10.1093/nar/gkaf764>
38. Core L, Adelman K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* 2019;33:960–82. <https://doi.org/10.1101/gad.325142.119>
39. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* 2012;13:720–31. <https://doi.org/10.1038/nrg3293>
40. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* 2015;16:167–77. <https://doi.org/10.1038/nrm3953>
41. Dahlberg O, Shilkova O, Tang M *et al.* P-TEFb, the super elongation complex and mediator regulate a subset of non-paused genes during early *Drosophila* embryo development. *PLoS Genet* 2015;11:e1004971. <https://doi.org/10.1371/journal.pgen.1004971>
42. Mahat DB, Kwak H, Booth GT *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* 2016;11:1455–76. <https://doi.org/10.1038/nprot.2016.086>
43. Bosque A, Planelles V. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood* 2009;113:58–65. <https://doi.org/10.1182/blood-2008-07-168393>
44. Kim Y, Cameron PU, Lewin SR *et al.* Limitations of dual-fluorescent HIV reporter viruses in a model of pre-activation latency. *J Intern AIDS Soc* 2019;22:e25425. <https://doi.org/10.1002/jia2.25425>
45. Reeder JE, Kwak YT, McNamara RP *et al.* HIV Tat controls RNA polymerase II and the epigenetic landscape to transcriptionally reprogram target immune cells. *eLife* 2015;4:e08955. <https://doi.org/10.7554/eLife.08955>
46. Marban C, Su T, Ferrari R *et al.* Genome-wide binding map of the HIV-1 Tat protein to the human genome. *PLoS One* 2011;6:e26894. <https://doi.org/10.1371/journal.pone.0026894>
47. Qu D, Sun WW, Li L *et al.* Long noncoding RNA MALAT1 releases epigenetic silencing of HIV-1 replication by displacing the polycomb repressive complex 2 from binding to the LTR promoter. *Nucleic Acids Res* 2019;47:3013–27. <https://doi.org/10.1093/nar/gkz117>
48. Feltl D, Zavadova E, Pala M *et al.* The dynamics of plasma transforming growth factor beta 1 (TGF-beta1) level during radiotherapy with or without simultaneous chemotherapy in advanced head and neck cancer. *Oral Oncol* 2005;41:208–13. <https://doi.org/10.1016/j.oraloncology.2004.09.005>
49. Gu J, Babayeva ND, Suwa Y *et al.* Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4. *Cell Cycle* 2014;13:1788–97. <https://doi.org/10.4161/cc.28756>
50. Lu H, Li Z, Xue Y *et al.* AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. *Proc Natl Acad Sci USA* 2014;111:E15–24.
51. Schulze-Gahmen U, Echeverria I, Stjepanovic G *et al.* Insights into HIV-1 proviral transcription from integrative structure and dynamics of the Tat:AFF4:p-TEFb:TAR complex. *eLife* 2016;5:e15910. <https://doi.org/10.7554/eLife.15910>
52. Asamitsu K, Okamoto T. The Tat/P-TEFb protein–protein interaction determining transcriptional activation of HIV. *CPD* 2017;23:4091–7. <https://doi.org/10.2174/1381612823666170710164148>
53. Fujinaga K, Huang F, Peterlin BM. P-TEFb: the master regulator of transcription elongation. *Mol Cell* 2023;83:393–403. <https://doi.org/10.1016/j.molcel.2022.12.006>
54. Liu P, Xiang Y, Fujinaga K *et al.* Release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. *J Biol Chem* 2014;289:9918–25. <https://doi.org/10.1074/jbc.M113.539015>
55. Hyder U, Shukla A, Challa A *et al.* HIV-1 proviral genome engineering with CRISPR-Cas9 for mechanistic studies. *Viruses* 2024;16:287. <https://doi.org/10.3390/v16020287>
56. Lucic B, Chen HC, Kuzman M *et al.* Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat Commun* 2019;10:4059. <https://doi.org/10.1038/s41467-019-12046-3>
57. Marini B, Kertesz-Farkas A, Ali H *et al.* Nuclear architecture dictates HIV-1 integration site selection. *Nature* 2015;521:227–31. <https://doi.org/10.1038/nature14226>
58. Singh PK, Bedwell GJ, Engelman AN. Spatial and genomic correlates of HIV-1 integration site targeting. *Cells* 2022;11:655. <https://doi.org/10.3390/cells11040655>
59. Einkauf KB, Osborn MR, Gao C *et al.* Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell* 2022;185:266–82.e15. <https://doi.org/10.1016/j.cell.2021.12.011>
60. Vansant G, Chen HC, Zorita E *et al.* The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res* 2020;48:7801–17. <https://doi.org/10.1093/nar/gkaa536>
61. Mbonye U, Agaponova A, Yang M *et al.* Protocol for intracellular immunofluorescence measurements of latent HIV reactivation in a primary CD4(+) T cell model. *STAR Protocols* 2024;5:103398. <https://doi.org/10.1016/j.xpro.2024.103398>
62. Cujec TP, Cho H, Maldonado E *et al.* The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol Cell Biol* 1997;17:1817–23. <https://doi.org/10.1128/MCB.17.4.1817>
63. Conrad RJ, Fozouni P, Thomas S *et al.* The short isoform of BRD4 promotes HIV-1 latency by engaging repressive SWI/SNF chromatin-remodeling complexes. *Mol Cell* 2017;67:1001–12.e6. <https://doi.org/10.1016/j.molcel.2017.07.025>
64. Bigrove DA, Mahmoudi T, Henklein P *et al.* Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proc Natl Acad Sci USA* 2007;104:13690–5. <https://doi.org/10.1073/pnas.0705053104>
65. Yang Z, Yik JH, Chen R *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* 2005;19:535–45. <https://doi.org/10.1016/j.molcel.2005.06.029>