Single-cell transcriptional landscapes reveal HIV-1–driven aberrant host gene transcription as a potential therapeutic target

Runxia Liu[1,*], **Yang-Hui Jimmy Yeh**[1,*], **Ales Varabyou**[2,*], **Jack A. Collora**[1], **Scott Sherrill-Mix**[3], **C. Conover Talbot Jr.**[4], **Sameet Mehta**[5], **Kristen Albrecht**[1], **Haiping Hao**[4], **Hao Zhang**[6], **Ross A. Pollack**[7], **Subul A. Beg**[7], **Rachela M. Calvi**[8], **Jianfei Hu**[9], **Christine M. Durand**[7], **Richard F. Ambinder**[7], **Rebecca Hoh**[10], **Steven G. Deeks**[10], **Jennifer Chiarella**[8], **Serena Spudich**[8], **Daniel C. Douek**[9], **Frederic D. Bushman**[3], **Mihaela Pertea**[2,11], **Ya-Chi Ho**[1,†]

[1]Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06519, USA.

[2]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

[3]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

[4]Institute for Basic Biomedical Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA.

[5]Yale Center for Genome Analysis, Yale University, New Haven, CT 06519, USA.

[6]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.

[7]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

[8]Department of Neurology, Yale University School of Medicine, New Haven, CT 06519, USA.

[9]Vaccine Research Center, National Institute of Health, Bethesda, MD 20892, USA.

[10]Department of Medicine, University of California, San Francisco, CA 94110, USA.

[†]Corresponding author. ya-chi.ho@yale.edu.

[*]These authors contributed equally to this work.

SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/12/543/eaaz0802/DC1

[11]Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

Abstract

Understanding HIV-1–host interactions can identify the cellular environment supporting HIV-1 reactivation and mechanisms of clonal expansion. We developed HIV-1 SortSeq to isolate rare HIV-1–infected cells from virally suppressed, HIV-1–infected individuals upon early latency reversal. Single-cell transcriptome analysis of HIV-1 SortSeq[+] cells revealed enrichment of nonsense-mediated RNA decay and viral transcription pathways. HIV-1 SortSeq[+] cells up-regulated cellular factors that can support HIV-1 transcription (*IMPDH1* and *JAK1*) or promote cellular survival (*IL2* and *IKBKB*). HIV-1–host RNA landscape analysis at the integration site revealed that HIV-1 drives high aberrant host gene transcription downstream, but not upstream, of the integration site through HIV-1–to–host aberrant splicing, in which HIV-1 RNA splices into the host RNA and aberrantly drives host RNA transcription. HIV-1–induced aberrant transcription was driven by the HIV-1 promoter as shown by CRISPR-dCas9–mediated HIV-1–specific activation and could be suppressed by CRISPR-dCas9–mediated inhibition of HIV-1 5' long terminal repeat. Overall, we identified cellular factors supporting HIV-1 reactivation and HIV-1–driven aberrant host gene transcription as potential therapeutic targets to disrupt HIV-1 persistence.

INTRODUCTION

Despite effective antiretroviral therapy (ART), HIV-1 persists in latently infected CD4[+] T cells as a major barrier to cure (1–3). During treatment interruptions, CD4[+] T cells from the latent reservoir contribute to viral rebound (4). To cure HIV-1 infection, all cells harboring infectious HIV-1 proviruses need to be recognized and eliminated by the immune system. However, latent HIV-1 proviruses are transcriptionally inactive and do not present antigens to immune effectors. Transcriptional interference by the host gene may induce transcriptional read-through or viral promoter occlusion and prevent HIV-1 transcription (5, 6). Transcriptional blocks (7), such as the lack of active forms of cellular transcription factors (8–13) and the HIV-1 Tat feedback loop (14–17), prevent HIV-1 transcriptional initiation and elongation. Therefore, despite maximum T cell activation, HIV-1 reactivation is stochastic, as some HIV-1 proviruses remain transcriptionally inactive (18). It remains unclear which cellular factors contribute to maximum HIV-1 expression upon latency reversal, particularly in CD4[+] T cells from HIV-1–infected individuals. Understanding the cellular environment supporting HIV-1 expression upon latency reversal is required to effectively target HIV-1–infected cells.

The rarity of HIV-1–infected cells in vivo and the lack of cellular surface markers that can distinguish HIV-1–infected cells harboring intact proviruses make it challenging to study HIV-1–host interactions in vivo. Only 1 to 100 CD4[+] T cells harbor intact HIV-1 proviruses in ART-treated, virally suppressed, HIV-1–infected individuals (18–20). Using HIV-1 Env expression as a surrogate, broadly neutralizing antibodies can capture HIV-1–infected cells for transcriptome analysis (21). However, induction of readily detectable amount of HIV-1 Env expression requires >30 hours of mitogen stimulation (21). In contrast, using HIV-1 RNA as a surrogate, fluorescence in situ hybridization (FISH)–based methods can identify

HIV-1–infected cells within 24 hours of stimulation (22). However, the branched DNA amplification methods used to amplify the low amount of HIV-1 RNA in FISH-based methods degrade RNA and prevent downstream transcriptome analysis. Therefore, the cellular environment supporting HIV-1 transcription during early latency reversal remains unknown.

HIV-1–infected cells undergo clonal expansion, and these clonally expanded cells increase over time (23–25). More than 50% of HIV-1–infected cells harboring replication-competent HIV-1 proviruses are maintained through clonal expansion (26–28), making this proliferating latent reservoir a major concern. Clonal expansion is presumably driven by antigen stimulation (29, 30) and homeostatic proliferation (31–33). Recent evidence suggests that proliferation of some HIV-1–infected cells in vivo is associated with HIV-1 integration sites, particularly in individuals treated with years of suppressive ART (23–25): First, although HIV-1 integration sites during in vitro infection span the transcription unit of the host genome, HIV-1 integration sites found in vivo are enriched in a small region of certain cancer-related genes (such as within 90 of the 370-kb *BACH2* transcription unit) as expanded clones (23, 24). Second, in vivo enrichment of HIV-1 integration into these cancer-related genes is exclusively in the same orientation as the host transcription unit, whereas HIV-1 integration into these cancer-related genes in vitro can be either the same or opposite orientation (23, 24). This indicates that both the location and orientation of HIV-1 at the integration site may be associated with preferential proliferation. Third, HIV-1 integration in the same orientation as the host transcription unit, such as *BACH2* and *STAT5B*, induces HIV-1 RNA splicing into the host RNA, creating HIV-1–host chimeric RNA that can contribute to HIV-1 persistence (34). These lines of evidence suggest that an unknown mechanism, which cannot be explained by antigen-driven proliferation and homeostatic proliferation, favors the proliferation of HIV-1–infected cells in a manner dependent on integration location and orientation.

We propose that HIV-1 reactivation changes the host cellular transcriptional landscape both globally in infected CD4[+] T cells and locally at the integration site. Here, we developed HIV-1 SortSeq to identify rare HIV-1–infected cells from ART-treated, virally suppressed, HIV-1–infected individuals and examined the transcriptional landscape of HIV-1–infected cells within 24 hours of latency reversal. We further developed a CRISPR-dCas9–based HIV-1–specific activation and inhibition system to examine the impact of HIV-1 promoter activity on the host gene at the integration site. Our approach examines the cellular environment supporting full HIV-1 reactivation and mechanisms of HIV-1 integration site-related proliferation.

RESULTS

HIV-1 SortSeq identifies HIV-1–infected cells from individuals

To examine HIV-1–host genetic interactions upon early latency reversal, we developed HIV-1 SortSeq to identify rare (1 to 100/10[6]) (18, 20, 35, 36) CD4[+] T cells containing inducible HIV-1 from ART-treated, virally suppressed, HIV-1–infected individuals (Fig. 1; fig. S1, A and B; and table S1). Briefly, we treated CD4[+] T cells with phorbol 12-myristate 13-acetate (PMA) and ionomycin ex vivo for 16 hours in the presence of enfuvirtide to

induce HIV-1 latency reversal without causing cellular proliferation or new rounds of infection (Fig. 1A). Although this approach cannot answer why latent HIV-1 remains latent, we proposed to examine why inducible HIV-1 can be induced through examination of the cellular environment supporting HIV-1 transcription during early (<24 hours) latency reversal. Cells were fixed, permeabilized, and hybridized with HIV-1 RNA-specific probes targeting 5' and 3' HIV-1 (data files S2 and S3) under RNA-preserving conditions (37) and sorted by flow cytometry. The use of one set of 96 fluorescently labeled probes targeting 5' HIV-1 RNA and another set of 96 fluorescently labeled probes targeting 3' HIV-1 overcomes the HIV-1 sequence diversity in clinical samples and maximizes HIV-1 RNA signal without RNA degradation caused by branched DNA amplification (22). HIV-1 SortSeq detected primary CD4[+] T cells infected with NL4–3 reference strain or clinical isolates (Fig. 1B) and CD4[+] T cells from HIV-1–infected individuals (Fig. 1C) with high specificity (Fig. 1D). HIV-1 SortSeq captured one HIV-1–infected cell per million uninfected cells (Fig. 1E). The frequency of HIV-1 SortSeq[+] cells was correlated with the size of the latent reservoir measured by the quantitative viral outgrowth assay (Fig. 1F). The RNA in HIV-1 SortSeq cells was sufficient for single-cell complementary DNA library construction (fig. S1C).

We captured 86 HIV-1 SortSeq[+] cells in bulk from seven participants and 48 HIV-1 SortSeq[+] single cells from 14 ART-treated, virally suppressed, HIV-1–infected individuals (data file S4). Instead of choosing participants with a known, comparatively large latent reservoir (21), we did not preselect participants to have a better representation of the diversity of the latent reservoir. The low number of cells we identified reflects the rarity of cells containing inducible HIV-1. We identified spliced HIV-1 RNA, no large internal deletions, and no APOBEC3G-mediated hypermutations in HIV-1 SortSeq[+] cells, indicating a likely intact HIV-1 genome (fig. S2). Of note, HIV-1 harboring small internal deletions or missense mutations can still produce readily detectable amounts of HIV-1 RNA (38). Using identical HIV-1 *env* V3-V4 sequences as an indicator of clonally expanded HIV-1–infected cells (39), we identified the same expanded clone by HIV-1 SortSeq and from viral outgrowth culture positive wells, indicating detection of clonally expanded replication-competent HIV-1 (fig. S3). We then identified HIV-1 SortSeq[+] and SortSeq[−] cells from the same HIV-1–infected individuals through two-way flow cytometric single-cell sorting for single-cell RNA sequencing (RNA-seq). HIV-1 SortSeq[+] and SortSeq[−] cells from HIV-1–infected individuals were sorted directly into tubes containing RNA-preserving buffer to maximize RNA capture. Although this method captures HIV-1 SortSeq[+] and SortSeq[−] single cells, it does not allow flow cytometry confirmation of sorting purity. Therefore, we used the presence of HIV-1 RNA reads, as shown on Integrative Genomic Browser and HIV BLAST (Los Alamos National Laboratory), to ensure that HIV-1 SortSeq[+] cells were authentic HIV-1–infected cells.

HIV-1 SortSeq[+] cells are polarized in T$_H$1 cells

We first examined the T cell activation status in HIV-1 SortSeq[+] and SortSeq[−] cells (Fig. 2). To avoid batch effects in transcriptome analysis, all HIV-1 SortSeq cells processed at Yale University, but not at Johns Hopkins University, were included for transcriptome analysis. From 28 HIV-1 SortSeq[+] cells and 43 HIV-1 SortSeq[−] cells from 10 ART-treated, virally suppressed, HIV-1–infected individuals, we found that both HIV-1 SortSeq[+] and SortSeq[−]

cells expressed RNA encoding early activation markers (CD69 and CD25), but not late activation markers [CD38 and human lymphocyte antigen DR (HLA-DR)], suggesting that HIV-1 SortSeq captured early activation events. We found that the extent of T cell activation (as measured by RNA expression levels of T cell activation markers CD69, CD25, CD38, and HLA-DR) was comparable between HIV-1 SortSeq$^+$ and SortSeq$^-$ cells (Fig. 2, A to D; P = not significant).

We next examined the T cell polarization phenotypes of HIV-1 SortSeq cells (Fig. 2). Using signature cytokine profiles, we found that only HIV-1 SortSeq$^+$ cells were enriched in T helper 1 (T$_H$1) effector cytokines *IFNG* (P = 0.018; Fig. 2F) and *IL2* (P = 0.029; Fig. 2G), suggesting an enrichment of HIV-1–infected cells in T$_H$1 (40), but not in T$_H$2, T$_H$17, or regulatory T cells (T$_{reg}$). This is consistent with the finding that HIV-1–infected cells are mainly memory CD4$^+$ T cells and T$_H$1 cells (31, 40), whereas HIV-1–uninfected cells can be either naïve or memory CD4$^+$ T cells.

Single-cell transcriptional landscape identifies up-regulation of cellular factors involving HIV-1 transcription, cellular survival, and immune response

We compared the single-cell transcriptional landscape between HIV-1 SortSeq$^+$ and SortSeq$^-$ cells. We found that both groups of cells demonstrated a heterogeneous transcriptional profile (Fig. 3, A and B), reflecting the unique and diverse cellular environment in each cell during early latency reversal. This is consistent with the diverse and heterogenous nature of CD4$^+$ T cells (41, 42). In comparison, the distribution of transcriptional profiles of housekeeping genes such as *B2M* and *UBC* was relatively homogeneous across samples (fig. S4, A and B), suggesting that the heterogeneity in transcription of these genes was not because of different amounts of cellular transcription in individual cells. We identified 447 differentially expressed genes in HIV-1 SortSeq$^+$ compared to SortSeq$^-$ cells (data file S5). Gene ontology analysis from the 395 genes up-regulated in HIV-1 SortSeq$^+$ cells showed molecular function enrichment for RNA binding proteins (fig. S5A) and biological function enrichment in nonsense-mediated RNA decay (NMD), RNA processing, and viral transcription (fig. S5B), suggesting the importance of cellular and viral RNA transcription and processing during early latency reversal.

We identified up-regulation of cellular factors *IMPDH1* (Fig. 3C) and *JAK1* (Fig. 3D) that may support HIV-1 expression, up-regulation of NMD pathways involving *UPF2* (Fig. 3E and fig. S5B), and up-regulation of *IKBKB* and *IL2* in HIV-1 SortSeq$^+$ cells (Fig. 2G and Fig. 3F). We found up-regulation of immune regulatory cytokine *LTA* (lymphotoxin-α) and chemokines *CCL3* [macrophage inflammatory protein–1 (MIP-1α)], *CCL4* (MIP-1β), and *XCL1* (lymphoactin; Fig. 3, G to J), all of which may inhibit HIV-1 replication (43). We note that Tat can induce *LTA*, *CCL3*, *CCL4*, and *XCL1* expression (44, 45) as a host immune response against HIV-1 infection (46). The expression of these genes enriched in HIV-1 SortSeq$^+$ cells {~5 to 10 log$_2$[transcripts per million (TPM) + 1]; Fig. 3, G to J} was comparable to the expression of housekeeping genes *B2M* and *UBC* [~4 to 8 log$_2$(TPM + 1); fig. S4, C and D]. Using constitutively expressed gene *EF1A* as a reference (fig. S6A), we confirmed up-regulation of *IMPDH1*, *JAK1*, *UPF2*, and *IKBKB* by quantitative

polymerase chain reaction (qPCR; fig. S6, B to E) normalized to *EF1A* expression (fig. S6, F to I).

HIV-1 proviruses that are integrated into cancer-related genes can be inducible

We next examined HIV-1–host interactions locally at the integration site. To identify HIV-1 integration sites in HIV-1 SortSeq[+] cells, we examined the HIV-1–host chimeric RNA that indicates the integration sites of HIV-1 proviruses. To exclude sequencing artifacts (47), only HIV-1–host chimeric RNA reads that captured the definite HIV-1–host RNA junction of a known HIV-1 splice site and a known host splice site (in HIV-1–host RNA splicing) or the exact end of the 5' and 3' HIV-1 long terminal repeat (LTR; in read-through transcription) were considered authentic. The sequencing reads used for analysis are shown in fig. S7 (see also tables S2 and S3). After stringent filtering for sequencing artifacts, manual examination using the University of California, Santa Cruz (UCSC) Genome Browser BLAT and identification of canonical splice sites, we found 19 HIV-1–host chimeric RNA species, three of which share the same integration site (table S3). The low number of HIV-1–host chimeric RNA reflects not only the low frequency of HIV-1–host chimeric RNA but also the stringency of our RNA sequence examination.

Whether HIV-1 proviruses that are integrated into cancer-related genes are intact or defective remains unclear. Because most of the HIV-1 proviruses are defective (18, 20), methods that examine HIV-1 integration site and HIV-1 genome integrity at the same time capture mainly defective proviruses (19, 48). On the other hand, methods that can capture intact HIV-1 from viral outgrowth positive cultures require multiple rounds of in vitro infection, making integration site analysis unfeasible (26–28). Using HIV-1 SortSeq, we examined whether HIV-1 copies that have integrated into cancer-related genes are inducible. Using a previously reported list of 2983 cancer-related genes (49) and 3804 housekeeping genes (50), we compared the proportion of cancer-related genes in the human genome and in integration sites identified in HIV-1 DNA and HIV-1–host chimeric RNA analysis during in vitro infection (47) and in cells from HIV-1–infected individuals (23) (Fig. 4). HIV-1 integration sites captured in HIV-1 SortSeq were enriched in both cancer-related genes (29.4%) and housekeeping genes (29.4%; Fig. 4), reflecting HIV-1 preferential integration into active transcription units (18, 51, 52). Overall, we found that HIV-1 proviruses that are integrated into cancer-related genes can be inducible and are therefore putatively intact.

HIV-1 integration in the same orientation in cancer-related genes is associated with HIV-1–host RNA aberrant splicing

We next examined whether HIV-1 integration orientation affected HIV-1–host interactions (fig. S8). To examine whether the host promoter or the HIV-1 promoter drives host gene transcription, we performed strand-specific RNA-seq and examined the sense-strand RNA sequences. We found that HIV-1 3' LTR dominated over the host promoter during read-through transcription upon latency reversal (fig. S7C), compatible with previous findings of 3' LTR dominance (5, 47). Although HIV-1 proviral integration orientation was roughly equal for the same and convergent orientations (fig. S8), we found that HIV-1 integration into cancer-related genes more frequently lead to HIV-1–host splicing: Five (62.5%) HIV-1–host splicing resulted in eight integrations into a cancer-related gene, whereas none of the

nine (0%) HIV-1 integrations into noncancer-related genes lead to HIV-1–host splicing (fig. S8). HIV-1 integration into cancer-related genes induced chimeric RNA spliced between canonical splice donors and canonical splice acceptors in HIV-1 and cancer-related genes such as *SMARCC1*, *PYHIN1*, *MIR155HG*, *BACH2*, and *NFATC3* (Fig. 5A and fig. S7D), some of which (*SMARCC1*, *BACH2,* and *NFATC3*) were previously reported integration sites in clonally expanded HIV-1–infected cells from HIV-1–infected individuals (23, 24). Our finding suggests that although HIV-1 can integrate into both cancer-related genes and noncancer-related genes, HIV-1 integration in the same orientation as the cancer-related genes is associated with HIV-1–host RNA aberrant splicing.

The HIV-1 promoter dominates over the host promoter and drives aberrant host gene transcription at the integration site

We identified HIV-1 splicing from the HIV-1 major splice donor into the canonical acceptor of exon 8 of *NFATC3* (Fig. 5A). *NFATC3* is important for T cell activation (53) and HIV-1 reactivation (54). HIV-1 integration into *NFATC3* upstream of this splice junction was previously reported in clonally expanded cells in HIV-1–infected individuals (23). To understand whether HIV-1 integration affects *NFATC3* transcription, we compared *NFATC3* transcripts in different single cells (Fig. 5B). In cells which did not have HIV-1 integrated into *NFATC3*, all exons of *NFATC3* remained transcribed (Fig. 5B). However, in the single cell that contained HIV-1 integrated into *NFATC3* (154_21), all exons downstream of HIV-1 integration site were transcribed; however, we did not detect transcription upstream of the HIV-1 integration site (Fig. 5B). These results indicate that HIV-1 LTR dominates over the host at the integration site and drives aberrant *NFATC3* transcription.

The HIV-1 promoter drives aberrant host protein expression at the integration site

Although the host gene may suppress HIV-1 transcription through transcriptional interference, as demonstrated in cell line clones in which an HIV-1 reporter is integrated into noncancer-related genes *PP5*, *UBA2* (5), and *HPRT* (6), the presence of HIV-1 RNA splicing into *BACH2* and *STAT5B* RNA (34) suggests that HIV-1 can escape host gene transcriptional interference. We examined the impact of HIV-1 on host gene expression at the integration site to understand whether HIV-1 insertional activation of the host gene is driven by HIV-1 LTR. Although HIV-1–infected cell line models may not reflect the quiescent state of resting memory CD4$^+$ T cells, the use of cell line clones allows examination of HIV-1–host interactions at both transcription and translation levels. Using an HIV-1 reporter provirus [NL4–3-d6-dEnv-drGFP; (55)] that contained all splice elements to allow examination of HIV-1–host splicing, we infected Jurkat T cells at a low multiplicity of infection, which we then sorted into single cells and grew into individual clones. Instead of using targeted integration of HIV-1 reporter into specific genes such as *BACH2* and *STAT5B*, our approach recapitulated HIV-1 integration into the host intron in vivo without additional transcriptional interference caused by the transcriptional terminator cassettes in lentiviral vectors used in targeted integration methods. We established three Jurkat T cell clones harboring HIV-1 proviral reporter proviruses integrated into the introns of three cancer-related genes, such as *RAP1B*, *VAV1*, and *SPECC1*, in the same orientation as the host gene transcription unit, which resembled the integration sites observed in vivo (Fig. 6).

Of note, HIV-1 integration into *VAV1* was reported in clonally expanded HIV-1–infected cells from HIV-1–infected individuals (23).

We examined the host gene transcriptional landscape at the integration site in these three HIV-1–Jurkat cell clones using RNA-seq of polyadenylated RNA (Fig. 6, A to C). First, we found that HIV-1 integration did not affect host gene transcription upstream of the integration site (Fig. 6), suggesting an orientation-dependent effect of HIV-1 on the host gene expression. Second, we found that host gene expression was highly increased (>5-fold) downstream of the HIV-1 integration site in all three HIV-1–Jurkat clones (Fig. 6). Third, zooming into the intron in which HIV-1 is integrated, we found that HIV-1 induced intron retention at the integration site (Fig. 6). Fourth, we captured HIV-1–host chimeric RNA splicing events from canonical HIV-1 splice donor to canonical host splice acceptor. HIV-1 aberrant splicing activated cryptic host exons (Fig. 6). The aberrant splicing events followed the GT|AG mRNA processing rule and are therefore unlikely to be sequencing artifacts. Last, HIV-1 transcription and intra–HIV-1 splicing remained intact (Fig. 6E), suggesting that the HIV-1 RNA can splice into human RNA without affecting HIV-1 RNA splicing. Together, these findings suggest that HIV-1–induced aberrant host transcription is not simply an HIV-1–host splicing event: HIV-1 drives downstream host gene expression, induces host intron retention, and activates cryptic exons.

To understand whether HIV-1–induced aberrant host gene transcription leads to aberrant host protein translation, we performed a Western blot on HIV-1–Jurkat clones 1G2 (in which HIV-1 is integrated upstream of the translation start site of *RAP1B*) and 8B10 (in which HIV-1 is integrated downstream of the translation start site of *VAV1*). We found that HIV-1 integration upstream of the *RAP1B* translation start site induced increased Ras-related protein Rap-1b (RAP1B) expression (Fig. 6D), whereas HIV-1 integration downstream of the *VAV1* translation start site induced truncated Proto-oncogene vav (VAV1) protein expression (Fig. 6D). In the 8B10-VAV1 clone, HIV-1 integrated downstream of the translation start codon and cut into the middle of the protein-coding region, thereby causing truncation of the VAV1 protein (Fig. 6A). In the 1G2-RAP1B clone, HIV-1 integrated upstream of the translation start codon (Fig. 6B), and thus, the protein-coding region remained intact. Therefore, there was no RAP1B truncation in the 1G2-RAP1B clone. Of note, an N-terminal truncation is known to increase the oncogenic potency of VAV1 (56). Overall, we showed that HIV-1–induced aberrant host gene transcription at the integration site leads to aberrant host protein expression.

HIV-1–driven aberrant transcription can be suppressed by CRISPR-mediated inhibition of the HIV-1 LTR

We next examined whether it is the HIV-1 promoter that drives the aberrant host protein expression. We hypothesized that HIV-1–specific activation would increase host gene expression at the integration site, whereas HIV-1–specific inhibition would suppress host gene expression at the integration site. We constructed dCas9-VP64–mediated HIV-1–specific activation (CRISPRa) and dCas9-Krab–mediated HIV-1–specific inhibition (CRISPRi) systems (57) using guide RNAs (gRNAs) targeting HIV-1 LTR (58) (Fig. 7A). The CRISPRa and CRISPRi system has been previously established to examine how

targeted activation and suppression of cellular genes can change host-pathogen interactions (57). We found that activation of HIV-1 LTR drives aberrant VAV1 protein expression. Thus, HIV-1–induced aberrant host protein expression is driven by HIV-1 LTR. Inhibition of HIV-1–LTR reduces aberrant VAV1 protein expression (Fig. 7, B and C). Using RNA landscape mapping at the HIV-1 integration site in all three clones, we found that HIV-1–specific CRISPRi inhibition suppressed HIV-1–driven aberrant host gene transcription at the integration site in all three cell line clones (Fig. 8), restoring the expression of the host gene to that of the uninfected Jurkat cells (Fig. 8). Overall, our results suggest that HIV-1–induced aberrant host gene transcription can be targeted by CRISPR-based or potentially small molecule–based disruption of HIV-1 LTR function.

DISCUSSION

Our comprehensive study—encompassing global single-cell transcriptional landscape to HIV-1–host interactions at the integration site, clinical samples to cell line validation, and moving from mechanisms to potential therapeutic targets—provides insights on HIV-1 persistence and HIV-1 eradication strategies. The stochastic nature of HIV-1 reactivation makes it challenging to expose all HIV-1–infected cells for immune clearance (18). Nonetheless, the identification of cellular factors required for HIV-1 transcription may help to understand transcriptional blocks on HIV-1 reactivation (7).

We found enrichment of HIV-1 SortSeq$^+$ cells in cells with a T_H1 phenotype. CD4$^+$ T cells in peripheral blood are polarized toward T_H1 about 10-fold more often than other polarizations (59). HIV-1 infects T_H1 cells at a higher frequency, followed by T_H0, then by T_H2 cells (60). This is further evidenced by the fact that HIV-1–infected cells are enriched in the T_H1 population (40), memory T cells (31), and HIV-1–specific CD4$^+$ T cells (29). On the basis of the integration sites that we captured, there is no evidence suggesting that the enrichment in T_H1 phenotype is driven by the integration site.

Inosine-5'-monophosphate dehydrogenase (IMPDH) is the rate-limiting enzyme required for guanine nucleotide de novo synthesis in CD4$^+$ T cells. Inhibition of IMPDH by a U.S. Food and Drug Administration (FDA)–approved drug mycophenolic acid [the active form of mycophenolate mofetil (MMF)] suppresses HIV-1 replication (61). *JAK1* is required for Tat-dependent HIV-1 gene expression (62) and is involved in T cell activation. Janus kinase 1 (JAK1) inhibition by FDA-approved drug ruxolitinib suppresses HIV-1 replication (63). Although the role of *IMPDH1* and *JAK1* in HIV-1 latency remains unknown, targeting IMPDH by MMF (NCT03262441) and JAK1 by ruxolitinib (NCT02475655) is currently being examined in ongoing clinical trials. Although numerous cellular factors and small-molecule compounds had been proposed to alter HIV-1 expression in cell line models, HIV-1 SortSeq identifies cellular factors that are enriched in HIV-1–infected cells from virally suppressed individuals as high-priority therapeutic targets. Our result suggests that transcriptome analysis by HIV-1 SortSeq identified not only a transcriptome signature of HIV-1–infected cells upon latency reversal but also cellular factors that can serve as therapeutic targets in HIV-1 eradication strategies.

NMD is a cellular surveillance mechanism that identifies and degrades aberrant mRNA containing premature stop codons and long aberrant introns. Regulator of nonsense transcripts 2 (UPF2), a key player in NMD pathways, can block HIV-1 RNA nuclear export (64, 65). Although UPF2 (along with other NMD proteins such as UPF1) restricts human T cell leukemia virus–1 (HTLV-1) expression and that the UPF2 restriction can be counteracted by HTLV-1 Tax (66), whether UPF2 can function as a restriction factor inhibiting HIV-1 expression upon latency reversal remains to be examined.

Whether HIV-1 drives clonal expansion of HIV-1–infected cells remains unclear. HTLV-1 Tax associates with inhibitor of nuclear factor κB (IκB) kinase and degrades nuclear factor κB inhibitor IκB. Together with Tax-induced interleukin-2 (IL-2) and IL-2Rα (CD25) transcription (67) and other oncogenic mechanisms, HTLV-1–infected cells undergo clonal expansion (68). HIV-1 Tat is the major driver of stochastic activation of the infected cells (17). Tat is also known to induce IL-2 expression (69) and *IKBKB* activity (44). Up-regulation of *IL2* and *IKBKB*, potentially in part by Tat, could be a mechanism protecting activated and proliferating CD4[+] T cells from apoptosis (70).

HIV-1 insertional activation of cancer-related genes is a potential mechanism for HIV-1 integration site-related proliferation (34). We found that HIV-1–driven aberrant host gene transcription affected processes beyond splicing between HIV-1 and the host gene. First, we found that HIV-1 proviruses integrated into cancer-related genes can be inducible and are putatively intact. Second, location- and integration orientation-dependent HIV-1–driven aberrant host gene transcription drove high (>5-fold increase) downstream host gene expression at the integration site. This suggests that HIV-1–driven aberrant host gene transcription may potentially contribute to integration site-dependent HIV-1 proliferation. Third, HIV-1–driven aberrant host gene transcription induced aberrant splicing, intron retention, and cryptic exon activation at the integration site, which can potentially induce NMD (71). Last, we found that the HIV-1 LTR drives aberrant host gene transcription, and this HIV-1 LTR-driven aberrant transcription can be targeted by CRISPR-based inhibition of the HIV-1 LTR.

The HIV-1 promoter LTR is relatively conserved in all infected cells, regardless of the integration site. We propose that inhibiting HIV-1 LTR-driven transcription disrupts HIV-1–driven aberrant host gene transcription. HIV-1 LTR-driven transcription can be inhibited ex vivo by IMPDH1 inhibitor MMF (61, 72) or JAK inhibitor ruxolitinib (63). Furthermore, gene therapy approaches may also disrupt HIV-1–driven aberrant host gene transcription, which targets HIV-1 LTR, and *gag* may remove elements required for HIV-1–driven host gene transcription (73). Therefore, despite the heterogeneity of HIV-1 integration sites, targeting the relatively conserved HIV-1 LTR was sufficient to disrupt HIV-1–driven aberrant transcription.

Our study is limited to examination of HIV-1–infected CD4[+] T cells harboring induced HIV-1 RNA upon latency reversal, not in the quiescent latent state. Because latent HIV-1 does not express HIV-1 RNA or HIV-1 proteins for detection, we had to induce maximal HIV-1 expression by stimulating CD4[+] T cells with PMA and ionomycin ex vivo. Analysis of HIV-1–infected CD4[+] T cells activated by other reactivation reagents, such as anti-CD3/

CD28 costimulation and latency-reversing agents, may reveal different transcriptional signatures and HIV-1–host transcriptional interaction. Another limitation is that HIV-1 SortSeq only captures a subset of HIV-1–infected cells in which HIV-1 proviruses are reactivated by a single round of PMA/ionomycin stimulation. The replication-competent, noninduced HIV-1 may not be reactivated despite multiple rounds of T cells activation (18). Thus, HIV-1 SortSeq, using HIV-1 RNA expression as surrogates, would not capture the latent reservoir that is not reactivated after one round of activation. Last, HIV-1 SortSeq involves sample fixation and permeabilization to deliver HIV-1 RNA probes into cells and multiple washing steps to remove nonspecific binding. The fixation and permeabilization cause partial RNA degradation, and cells may be lost during the washing steps. To overcome this barrier, we optimized an RNA-preserving method to preserve RNA integrity and to reduce cell loss during washing.

Overall, our single-cell transcriptome analysis of HIV-1–infected cells from HIV-1–infected individuals examined the transcriptional landscape both globally in cells and locally at the integration site. Although many cellular factors and small-molecule compounds have been implicated in HIV-1 persistence in cell line models, our study identifies key cellular factors and HIV-1 persistence mechanism in cells from HIV-1–infected individuals and suggests priority targets in HIV-1 cure strategies.

MATERIALS AND METHODS

Study design

We conducted a prospective, cross-sectional study recruiting adult HIV-1–infected individuals. The primary study objectives were to map the single-cell atlas of HIV-1–infected cells upon ex vivo latency reversal. The inclusion criteria were adult HIV-1–infected individuals under suppressive ART for >1 year with undetectable plasma viral load for >6 months. Peripheral blood and leukapheresis samples were obtained from a total of 25 HIV-1 study participants (table S1). This study was approved by the Johns Hopkins University and Yale University Institutional Review Boards. All participants were provided written informed consent. Additional methods details are available in the Supplementary Materials.

Statistical analysis

Analyses were performed using MedCalc and Prism software (GraphPad). Statistical tests are indicated in the figure legends. P values of <0.05 in two-tailed testing were considered statistically significant. In the regression analysis of the frequency of HIV-1–infected cells detected by HIV-1 SortSeq versus the frequency of input HIV-1–infected cells, R^2 and P values were calculated on log-transformed data using MedCalc. In the correlation analysis between the frequency of HIV-1 SortSeq$^+$ cells and the size of the latent reservoir as measured by viral outgrowth assays, log-transformed data were tested for a normal distribution using D'Agostino-Pearson test. The Pearson correlation coefficient and P value were calculated on log-transformed data using MedCalc. In single-cell analysis, DEsingle was used to compare the transcriptome of SortSeq$^+$ and SortSeq$^-$ cells using de-fault parameters (74), and P values of <0.05 after Benjamini-Hochberg correction were

considered significant (75). Red lines denote median, and dashed lines denote 75th and 25th percentiles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

REFERENCES AND NOTES

1. Chun T-W, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA, Fauci AS, Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc. Natl. Acad. Sci. U.S.A 94, 13193–13197 (1997). [PubMed: 9371822]

2. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF, Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278, 1295–1300 (1997). [PubMed: 9360927]

3. Wong JK, Hezareh M, Günthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD, Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278, 1291–1295 (1997). [PubMed: 9360926]

4. De Scheerder M-A, Vrancken B, Dellicour S, Schlub T, Lee E, Shao W, Rutsaert S, Verhofstede C, Kerre T, Malfait T, Hemelsoet D, Coppens M, Dhondt A, De Looze D, Vermassen F, Lemey P, Palmer S, Vandekerckhove L, HIV rebound is predominantly fueled by genetically identical viral expansions from diverse reservoirs. Cell Host Microbe 26, 347–358.e7 (2019). [PubMed: 31471273]

5. Lenasi T, Contreras X, Peterlin BM, Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe 4, 123–133 (2008). [PubMed: 18692772]

6. Han Y, Lin YB, An W, Xu J, Yang H-C, O'Connell K, Dordai D, Boeke JD, Siliciano JD, Siliciano RF, Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. Cell Host Microbe 4, 134–146 (2008). [PubMed: 18692773]

7. Yukl SA, Kaiser P, Kim P, Telwatte S, Joshi SK, Vu M, Lampiris H, Wong JK, HIV latency in isolated patient CD4+ T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. Sci. Transl. Med 10, eaap9927 (2018).

8. Duh EJ, Maury WJ, Folks TM, Fauci AS, Rabson AB, Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. Proc. Natl. Acad. Sci. U.S.A 86, 5974–5978 (1989). [PubMed: 2762307]

9. Böhnlein E, Lowenthal JW, Siekevitz M, Ballard DW, Franza BR, Greene WC, The same inducible nuclear proteins regulates mitogen activation of both the interleukin-2 receptor-alpha gene and type 1 HIV. Cell 53, 827–836 (1988). [PubMed: 2836068]

10. Dandekar DH, Ganesh KN, Mitra D, HIV-1 Tat directly binds to NFκB enhancer sequence: Role in viral and cellular gene expression. Nucleic Acids Res. 32, 1270–1278 (2004). [PubMed: 14981150]

11. Kinoshita S, Su L, Amano M, Timmerman LA, Kaneshima H, Nolan GP, The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. Immunity 6, 235–244 (1997). [PubMed: 9075924]

12. Nabel G, Baltimore D, An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature 326, 711–713 (1987). [PubMed: 3031512]

13. West MJ, Lowe AD, Karn J, Activation of human immunodeficiency virus transcription in T cells revisited: NF-κB p65 stimulates transcriptional elongation. J. Virol 75, 8524–8537 (2001). [PubMed: 11507198]

14. Cujec TP, Cho H, Maldonado E, Meyer J, Reinberg D, Peterlin BM, The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. Mol. Cell. Biol 17, 1817–1823 (1997). [PubMed: 9121429]

15. Herrmann CH, Rice AP, Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: Candidate for a Tat cofactor. J. Virol 69, 1612–1620 (1995). [PubMed: 7853496]

16. Tyagi M, Pearson RJ, Karn J, Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. J. Virol 84, 6425–6437 (2010). [PubMed: 20410271]

17. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV, Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122, 169–182 (2005). [PubMed: 16051143]

18. Ho Y-C, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD, Siliciano RF, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell 155, 540–551 (2013). [PubMed: 24243014]

19. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, Fray EJ, Beg SA, Antar AAR, Jenike KM, Bertagnolli LN, Capoferri AA, Kufera JT, Timmons A, Nobles C, Gregg J, Wada N, Ho Y-C, Zhang H, Margolick JB, Blankson JN, Deeks SG, Bushman FD, Siliciano JD, Laird GM, Siliciano RF, A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature 566, 120–125 (2019). [PubMed: 30700913]

20. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho Y-C, Richman DD, Deeks SG, Siliciano JD, Siliciano RF, Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat. Med 22, 1043–1049 (2016). [PubMed: 27500724]

21. Cohn LB, da Silva IT, Valieris R, Huang AS, Lorenzi JCC, Cohen YZ, Pai JA, Butler AL, Caskey M, Jankovic M, Nussenzweig MC, Clonal CD4+ T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation. Nat. Med 24, 604–609 (2018). [PubMed: 29686423]

22. Baxter AE, Niessl J, Fromentin R, Richard J, Porichis F, Charlebois R, Massanella M, Brassard N, Alsahafi N, Delgado G-G, Routy J-P, Walker BD, Finzi A, Chomont N, Kaufmann DE, Single-cell characterization of viral translation-competent reservoirs in HIV-infected individuals. Cell Host Microbe 20, 368–380 (2016). [PubMed: 27545045]

23. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH, Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science 345, 179–183 (2014). [PubMed: 24968937]

24. Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM, Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science 345, 570–573 (2014). [PubMed: 25011556]

25. Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic

M, Nussenzweig MC, HIV-1 integration landscape during latent and active infection. Cell 160, 420–432 (2015). [PubMed: 25635456]

26. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DI, Keele BF, Ho Y-C, Siliciano JD, Siliciano RF, Proliferation of latently infected CD4+ T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J. Exp. Med 214, 959–972 (2017). [PubMed: 28341641]

27. Lorenzi JC, Cohen YZ, Cohn LB, Kreider EF, Barton JP, Learn GH, Oliveira T, Lavine CL, Horwitz JA, Settler A, Jankovic M, Seaman MS, Chakraborty AK, Hahn BH, Caskey M, Nussenzweig MC, Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. Proc. Natl. Acad. Sci. U.S.A 113, E7908–E7916 (2016).

28. Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, Luke BT, Shao W, Hughes SH, Coffin JM, Kearney MF, Mellors JW, Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLOS Pathog. 13, e1006283 (2017).

29. Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, Okamoto Y, Casazza JP, Kuruppu J, Kunstman K, Wolinsky S, Grossman Z, Dybul M, Oxenius A, Price DA, Connors M, Koup RA, HIV preferentially infects HIV-specific CD4+ T cells. Nature 417, 95–98 (2002). [PubMed: 11986671]

30. Simonetti FR, Sobolewski MD, Fyne E, Shao W, Spindler J, Hattori J, Anderson EM, Watters SA, Hill S, Wu X, Wells D, Su L, Luke BT, Halvas EK, Besson G, Penrose KJ, Yang Z, Kwan RW, Van Waes C, Uldrick T, Citrin DE, Kovacs J, Polis MA, Rehm CA, Gorelick R, Piatak M, Keele BF, Kearney MF, Coffin JM, Hughes SH, Mellors JW, Maldarelli F, Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. Proc. Natl. Acad. Sci. U.S.A 113, 1883–1888 (2016). [PubMed: 26858442]

31. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, Boucher G, Boulassel M-R, Ghattas G, Brenchley JM, Schacker TW, Hill BJ, Douek DC, Routy J-P, Haddad EK, Sékaly R-P, HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat. Med 15, 893–900 (2009). [PubMed: 19543283]

32. Vandergeeten C, Fromentin R, DaFonseca S, Lawani MB, Sereti I, Lederman MM, Ramgopal M, Routy JP, Sékaly R-P, Chomont N, Interleukin-7 promotes HIV persistence during antiretroviral therapy. Blood 121, 4321–4329 (2013). [PubMed: 23589672]

33. Bosque A, Famiglietti M, Weyrich AS, Goulston C, Planelles V, Homeostatic proliferation fails to efficiently reactivate HIV-1 latently infected central memory CD4+ T cells. PLOS Pathog. 7, e1002288 (2011).

34. Cesana D, Santoni de Sio FR, Rudilosso L, Gallina P, Calabria A, Beretta S, Merelli I, Bruzzesi E, Passerini L, Nozza S, Vicenzi E, Poli G, Gregori S, Tambussi G, Montini E, HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. Nat. Commun 8, 498 (2017). [PubMed: 28887441]

35. Eriksson S, Graf EH, Dahl V, Strain MC, Yukl SA, Lysenko ES, Bosch RJ, Lai J, Chioma S, Emad F, Abdel-Mohsen M, Hoh R, Hecht F, Hunt P, Somsouk M, Wong J, Johnston R, Siliciano RF, Richman DD, O'Doherty U, Palmer S, Deeks SG, Siliciano JD, Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. PLOS Pathog. 9, e1003174 (2013).

36. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, Siliciano RF, Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat. Med 9, 727–728 (2003). [PubMed: 12754504]

37. Klemm S, Semrau S, Wiebrands K, Mooijman D, Faddah DA, Jaenisch R, van A. Oudenaarden, Transcriptional profiling of cells sorted by RNA abundance. Nat. Methods 11, 549–551 (2014). [PubMed: 24681693]

38. Pollack RA, Jones RB, Pertea M, Bruner KM, Martin AR, Thomas AS, Capoferri AA, Beg SA, Huang S-H, Karandish S, Hao H, Halper-Stromberg E, Yong PC, Kovacs C, Benko E, Siliciano RF, Ho Y-C, Defective HIV-1 proviruses are expressed and can be recognized by cytotoxic T lymphocytes, which shape the proviral landscape. Cell Host Microbe 21, 494–506.e4 (2017). [PubMed: 28407485]

39. Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, Kumar MR, Beg SA, Capoferri AA, Ray SC, Ho Y-C, Hill AL, Siliciano JD, Siliciano RF, Expanded cellular clones

carrying replication-competent HIV-1 persist, wax, and wane. Proc. Natl. Acad. Sci. U.S.A 115, E2575–E2584 (2018).

40. Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, Kuo H-H, Hua S, Chen HR, Ouyang Z, Reddy K, Dong K, Ndung'u T, Walker BD, Rosenberg ES, Yu XG, Lichterfeld M, Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J. Clin. Invest 127, 2689–2696 (2017). [PubMed: 28628034]

41. Abbas AK, Murphy KM, Sher A, Functional diversity of helper T lymphocytes. Nature 383, 787–793 (1996). [PubMed: 8893001]

42. Sallusto F, Lanzavecchia A, Heterogeneity of CD4+ memory T cells: Functional modules for tailored immunity. Eur. J. Immunol 39, 2076–2082 (2009). [PubMed: 19672903]

43. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P, Identification of RANTES, MIP-1α, and MIP-1β as the major HIV-suppressive factors produced by CD8+ T cells. Science 270, 1811–1815 (1995). [PubMed: 8525373]

44. Fiume G, Vecchio E, De Laurentiis A, Trimboli F, Palmieri C, Pisano A, Falcone C, Pontoriero M, Rossi A, Scialdone A, Fasanella Masci F, Scala G, Quinto I, Human immunodeficiency virus-1 Tat activates NF-κB via physical interaction with IκB-α and p65. Nucleic Acids Res. 40, 3548–3562 (2012). [PubMed: 22187158]

45. Kim BO, Liu Y, Zhou BY, He JJ, Induction of C chemokine XCL1 (lymphotactin/single C motif-1α/activation-induced, T cell-derived and chemokine-related cytokine) expression by HIV-1 Tat protein. J. Immunol 172, 1888–1895 (2004). [PubMed: 14734774]

46. Abdelwahab SF, Cocchi F, Bagley KC, Kamin-Lewis R, Gallo RC, DeVico A, Lewis GK, HIV-1-suppressive factors are secreted by CD4+ T cells during primary immune responses. Proc. Natl. Acad. Sci. U.S.A 100, 15006–15010 (2003).

47. Sherrill-Mix S, Ocwieja KE, Bushman FD, Gene activity in primary T cells infected with HIV89.6: Intron retention and induction of genomic repeats. Retrovirology 12, 79 (2015). [PubMed: 26377088]

48. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ, Rosenberg ES, Chun T-W, Li JZ, Yu XG, Lichterfeld M, Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J. Clin. Invest 129, 988–998 (2019). [PubMed: 30688658]

49. Sadelain M, Papapetrou EP, Bushman FD, Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer 12, 51–58 (2012).

50. Eisenberg E, Levanon EY, Human housekeeping genes, revisited. Trends Genet. 29, 569–574 (2013). [PubMed: 23810203]

51. Schröder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F, HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110, 521–529 (2002). [PubMed: 12202041]

52. Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, Pierson TC, Margolick JB, Siliciano RF, Siliciano JD, Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J. Virol 78, 6122–6133 (2004). [PubMed: 15163705]

53. Macian F, NFAT proteins: Key regulators of T-cell development and function. Nat. Rev. Immunol 5, 472–484 (2005). [PubMed: 15928679]

54. Chan JK, Bhattacharyya D, Lassen KG, Ruelas D, Greene WC, Calcium/calcineurin synergizes with prostratin to promote NF-κB dependent activation of latent HIV. PLOS ONE 8, e77749 (2013).

55. Yang HC, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, Zhou Y, Shrum CK, Han Y, Liu JO, Zhang H, Margolick JB, Siliciano RF, Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. J. Clin. Invest 119, 3473–3486 (2009). [PubMed: 19805909]

56. Katzav S, Cleveland JL, Heslop HE, Pulido D, Loss of the amino-terminal helix-loop-helix domain of the vav proto-oncogene activates its transforming potential. Mol. Cell. Biol 11, 1912–1920 (1991). [PubMed: 2005887]

57. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS, Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661 (2014). [PubMed: 25307932]

58. Qu D, Li C, Sang F, Li Q, Jiang Z-Q, Xu L-R, Guo H-J, Zhang C, Wang J-H, The variances of Sp1 and NF-κB elements correlate with the greater capacity of Chinese HIV-1 B'-LTR for driving gene expression. Sci. Rep 6, 34532 (2016).

59. Saito S, Sakai M, Sasaki Y, Tanebe K, Tsuda H, Michimata T, Quantitative analysis of peripheral blood Th0, Th1, Th2 and the Th1:Th2 cell ratio during normal human pregnancy and preeclampsia. Clin. Exp. Immunol 117, 550–555 (1999). [PubMed: 10469061]

60. Bahbouhi B, Al-Harthi L, Enriching for HIV-infected cells using anti-gp41 antibodies indirectly conjugated to magnetic microbeads. Biotechniques 36, 139–147 (2004). [PubMed: 14740496]

61. Chapuis AG, Rizzardi G. Paolo, D'Agostino C, Attinger A, Knabenhans C, Fleury S, Acha-Orbea H, Pantaleo G, Effects of mycophenolic acid on human immunodeficiency virus infection in vitro and in vivo. Nat. Med 6, 762–768 (2000). [PubMed: 10888924]

62. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS, Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe 4, 495–504 (2008). [PubMed: 18976975]

63. Gavegnano C, Detorio M, Montero C, Bosque A, Planelles V, Schinazi RF, Ruxolitinib and tofacitinib are potent and selective inhibitors of HIV-1 replication and virus reactivation in vitro. Antimicrob. Agents Chemother 58, 1977–1986 (2014). [PubMed: 24419350]

64. Ajamian L, Abel K, Rao S, Vyboh K, Garcia-de-Gracia F, Soto-Rifo R, Kulozik AE, Gehring NH, Mouland AJ, HIV-1 recruits UPF1 but Excludes UPF2 to promote nucleocytoplasmic export of the genomic RNA. Biomolecules 5, 2808–2839 (2015). [PubMed: 26492277]

65. Rao S, Amorim R, Niu M, Breton Y, Tremblay MJ, Mouland AJ, Host mRNA decay proteins influence HIV-1 replication and viral gene expression in primary monocyte-derived macrophages. Retrovirology 16, 3 (2019). [PubMed: 30732620]

66. Mocquet V, Neusiedler J, Rende F, Cluet D, Robin J-P, Terme J-M, Dodon M. Duc, Wittmann J, Morris C, Le Hir H, Ciminale V, Jalinot P, The human T-lymphotropic virus type 1 tax protein inhibits nonsense-mediated mRNA decay by interacting with INT6/EIF3E and UPF1. J. Virol 86, 7530–7543 (2012). [PubMed: 22553336]

67. Sun S-C, Ballard DW, Persistent activation of NF-κB by the tax transforming protein of HTLV-1: Hijacking cellular IκB kinases. Oncogene 18, 6948–6958 (1999). [PubMed: 10602469]

68. Gillet NA, Malani N, Melamed A, Gormley N, Carter R, Bentley D, Berry C, Bushman FD, Taylor GP, Bangham CR, The host genomic environment of the provirus determines the abundance of HTLV-1–infected T-cell clones. Blood 117, 3113–3122 (2011). [PubMed: 21228324]

69. Ott M, Emiliani S, Van Lint C, Herbein G, Lovett J, Chirmule N, McCloskey T, Pahwa S, Verdin E, Immune hyperactivation of HIV-1-infected T cells mediated by Tat and the CD28 pathway. Science 275, 1481–1485 (1997). [PubMed: 9045614]

70. Hayden MS, Ghosh S, NF-κB in immunobiology. Cell Res. 21, 223–244 (2011). [PubMed: 21243012]

71. Jacob AG, Smith CWJ, Intron retention as a component of regulated gene expression programs. Hum. Genet 136, 1043–1057 (2017). [PubMed: 28391524]

72. Reeves DB, Duke ER, Hughes SM, Prlic M, Hladik F, Schiffer JT, Anti-proliferative therapy for HIV cure: A compound interest approach. Sci. Rep 7, 4011 (2017). [PubMed: 28638104]

73. Dash PK, Kaminski R, Bella R, Su H, Mathews S, Ahooyi TM, Chen C, Mancuso P, Sariyer R, Ferrante P, Donadoni M, Robinson JA, Sillman B, Lin Z, Hilaire JR, Banoub M, Elango M, Gautam N, Mosley RL, Poluektova LY, McMillan J, Bade AN, Gorantla S, Sariyer IK, Burdo TH, Young W-B, Amini S, Gordon J, Jacobson JM, Edagwa B, Khalili K, Gendelman HE, Sequential LASER ART and CRISPR treatments eliminate HIV-1 in a subset of infected humanized mice. Nat. Commun 10, 2753 (2019). [PubMed: 31266936]

74. Miao Z, Deng K, Wang X, Zhang X, DEsingle for detecting three types of differential expression in single-cell RNA-seq data. Bioinformatics 34, 3223–3224 (2018). [PubMed: 29688277]

75. Benjamini Y, Hochberg Y, Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B. Methodol 57, 289–300 (1995).

76. Bullen CK, Laird GM, Durand CM, Siliciano JD, Siliciano RF, New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. Nat. Med 20, 425–429 (2014). [PubMed: 24658076]

77. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT, Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat. Biotechnol 33, 243–246 (2015). [PubMed: 25748911]

78. Kim D, Langmead B, Salzberg SL, HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]

79. Jenkins DF, Faits T, Briars E, Pro S. Carrasco, Cunningham S, Campbell JD, Yajima M, Johnson W. Evan, Interactive single cell RNA-Seq analysis with the Single Cell Toolkit (SCTK). bioRxiv 2018, 329755 (2018).

80. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97 (2016). [PubMed: 27141961]

81. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol 34, 184–191 (2016). [PubMed: 26780180]

82. Bolger AM, Lohse M, Usadel B, Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]

83. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

84. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T, deepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191 (2014). [PubMed: 24799436]

85. Koslicki D, Topological entropy of DNA sequences. Bioinformatics 27, 1061–1067 (2011). [PubMed: 21317142]

86. Quinlan AR, Hall IM, BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

87. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD, Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res. 40, 10345–10355 (2012).
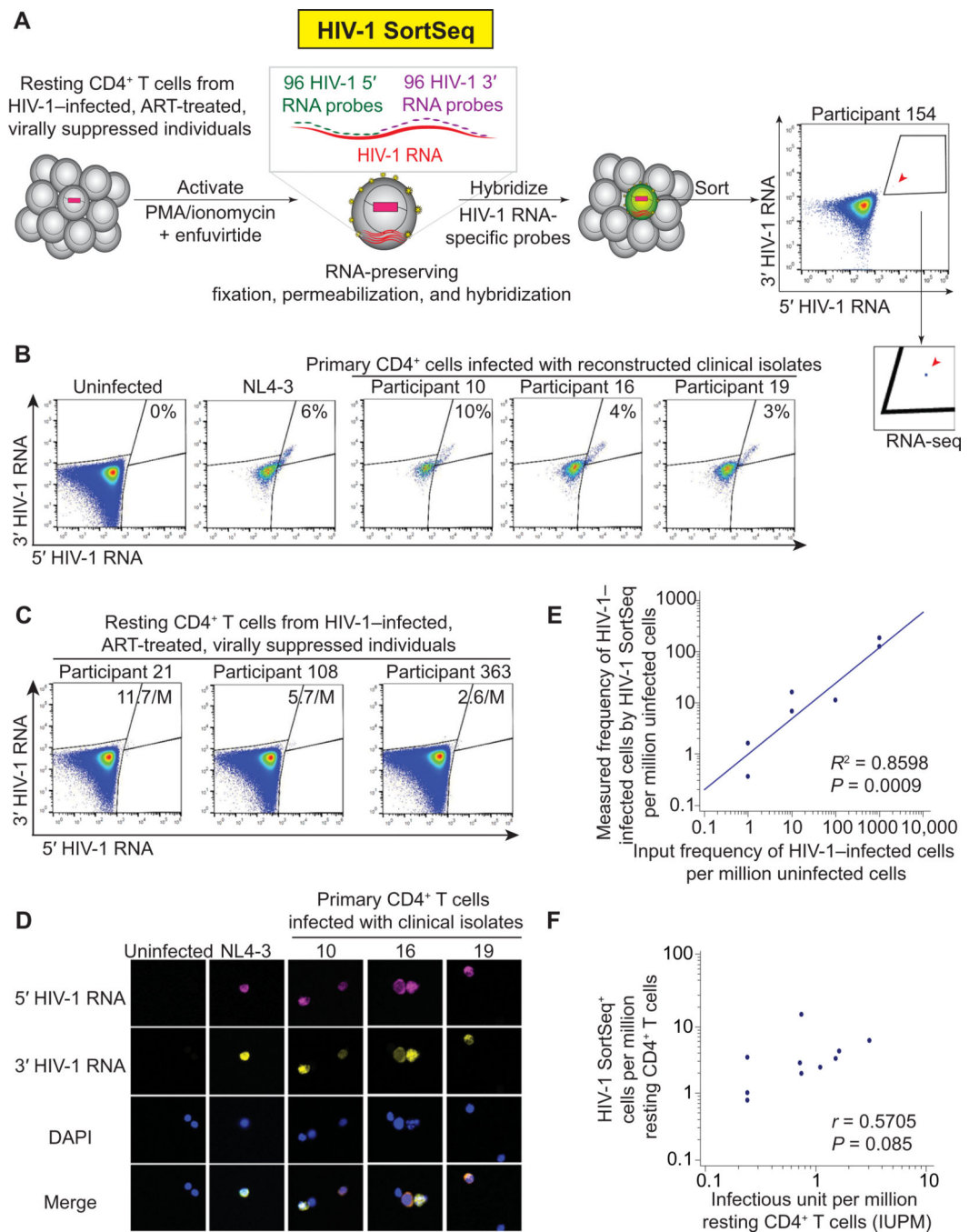
Fig. 1. HIV-1 SortSeq identifies CD4+ T cells harboring inducible HIV-1 from ART-treated, virally suppressed, HIV-1–infected individuals.

(**A**) Experimental scheme of HIV-1 SortSeq. CD4+ T cells from ART-treated, virally suppressed, HIV-1–infected individuals were sorted two-way into HIV-1 SortSeq+ and SortSeq− single cells. A stringent gating strategy away from the HIV-1 SortSeq− cells was used. (**B**) HIV-1 SortSeq gating strategy for primary cells infected with clinical isolates. The gating strategy was used to demonstrate the four quadrants of HIV-1 5' and 3' positive staining, not for sorting. (**C**) HIV-1 SortSeq gating strategy for CD4+ T cells isolated from

HIV-1–infected, ART-treated, and virally suppressed individuals upon latency reversal. The gating strategy was used to demonstrate the four quadrants of HIV-1 5' and 3' staining, not for sorting. (**D**) Fluorescent microscopic imaging of primary CD4[+] T cells infected with the NL4–3 reference strain and reconstructed clinical isolates at ~10% of infectivity. DAPI, 4,6-diamidino2-phenylindole nuclear staining. (**E**) Regression analysis of the frequency of HIV-1–infected cells detected by HIV-1 SortSeq versus the frequency of input HIV-1–infected cells. (**F**) Correlation between the frequency of HIV-1 SortSeq[+] cells and the size of the latent reservoir as measured by viral outgrowth assays.
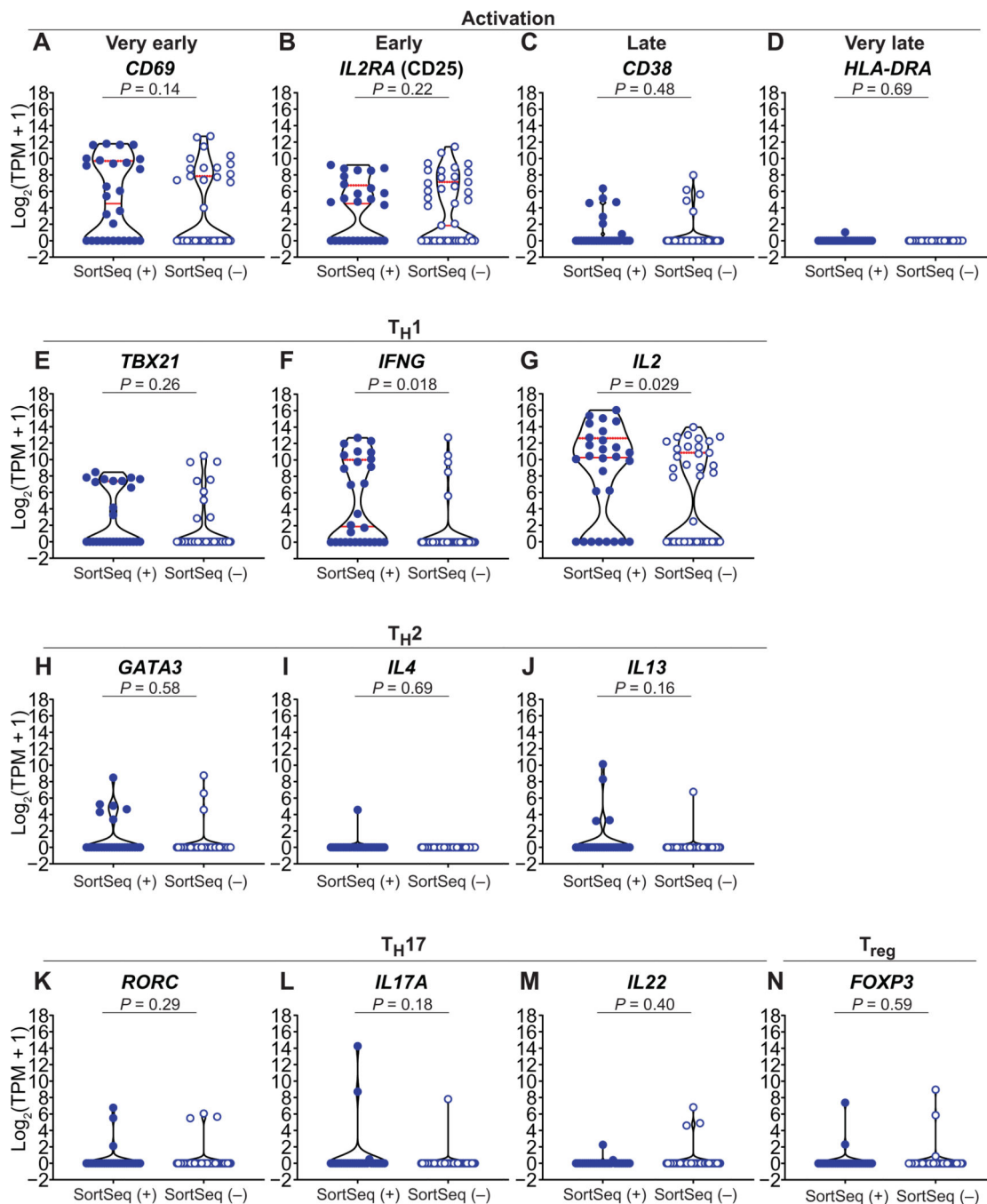
Fig. 2. HIV-1 SortSeq+ cells are polarized to T_H1 phenotype.
(**A** to **D**) RNA expression of T cell activation markers *CD69* (A), *IL2RA* (CD25) (B), *CD38* (C), and *HLA-DRA* (D) from HIV-1 SortSeq+ and SortSeq− cells from ART-treated, virally suppressed, HIV-1–infected individuals. (**E** to **N**) RNA expression of representative T cell polarization signatures of T_H1 (E to G), T_H2 (H to J), T_H17 (K to M), and T_reg (N) from HIV-1 SortSeq+ and SortSeq− cells. Each dot represents a single cell from 28 SortSeq+ and 43 SortSeq− cells. Red lines denote median expression. Dashed red lines denote 75th expression percentile.
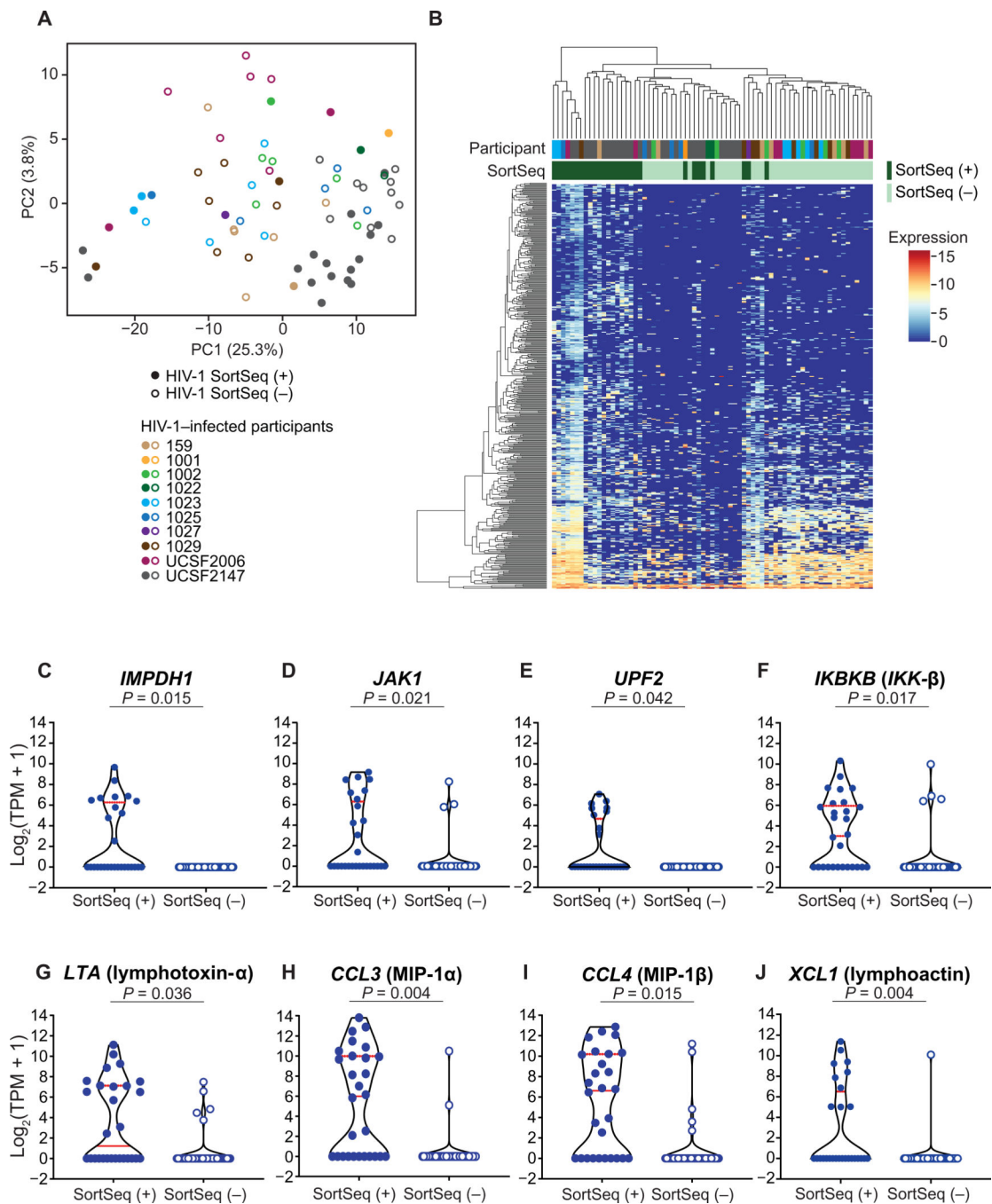
Fig. 3. The single-cell transcriptional landscape of HIV-1–infected cells upon ex vivo activation.
(**A**) The principal component analysis (PCA) demonstrates the distribution of HIV-1 SortSeq
$^{+}$ (closed circles) and SortSeq$^{−}$ cells (open circles) from ART-treated, virally suppressed,
HIV-1–infected individuals. (**B**) Heterogeneous transcriptional profile of HIV-1–infected
cells within 24 hours of latency reversal. The heat map shows significantly differentially
expressed genes between HIV-1 SortSeq^{+} and SortSeq$^{−}$ cells. Values indicate expression
measured as $\log_2(\text{TPM} + 1)$. The participant ID is color-coded as shown in (A). (**C** to **J**)
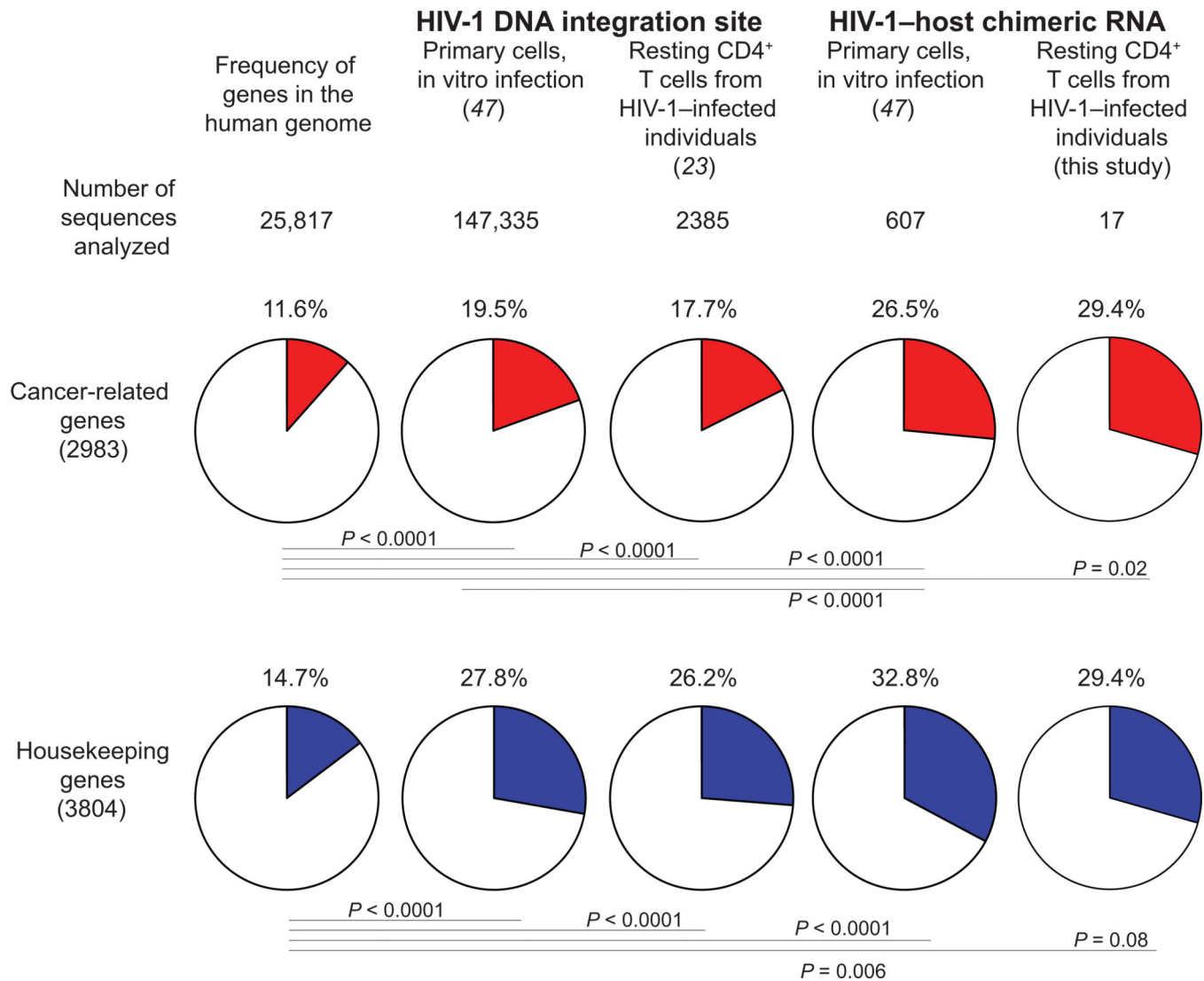Expression of selected significantly differentially expressed genes in HIV-1 SortSeq single

cells. Each dot represents a single cell from 28 SortSeq$^+$ and 43 SortSeq$^-$ cells. Red lines denote median expression. PC, principal component.

Number of sequences analyzed	Frequency of genes in the human genome	**HIV-1 DNA integration site**		**HIV-1–host chimeric RNA**	
		Primary cells, in vitro infection (*47*)	Resting CD4⁺ T cells from HIV-1–infected individuals (*23*)	Primary cells, in vitro infection (*47*)	Resting CD4⁺ T cells from HIV-1–infected individuals (this study)
	25,817	147,335	2385	607	17

Cancer-related genes (2983): 11.6%, 19.5%, 17.7%, 26.5%, 29.4%

$P < 0.0001$ $P < 0.0001$ $P < 0.0001$ $P = 0.02$
$P < 0.0001$

Housekeeping genes (3804): 14.7%, 27.8%, 26.2%, 32.8%, 29.4%

$P < 0.0001$ $P < 0.0001$ $P < 0.0001$ $P = 0.08$
$P = 0.006$

Fig. 4. HIV-1 proviruses integrated into cancer-related genes can be inducible.
Using a previously reported list of 2983 cancer-related genes (49) and 3804 housekeeping genes (50), we compared the proportion of cancer-related genes in the human genome, integration sites identified in HIV-1 DNA during in vitro infection (47), integration sites identified in HIV-1 DNA from HIV-1–infected, virally suppressed individuals (23), integration sites in HIV-1–host chimeric RNA during in vitro infection (47), and integration sites identified in HIV-1–host chimeric RNA in this study. *P* values, Fisher's exact test.
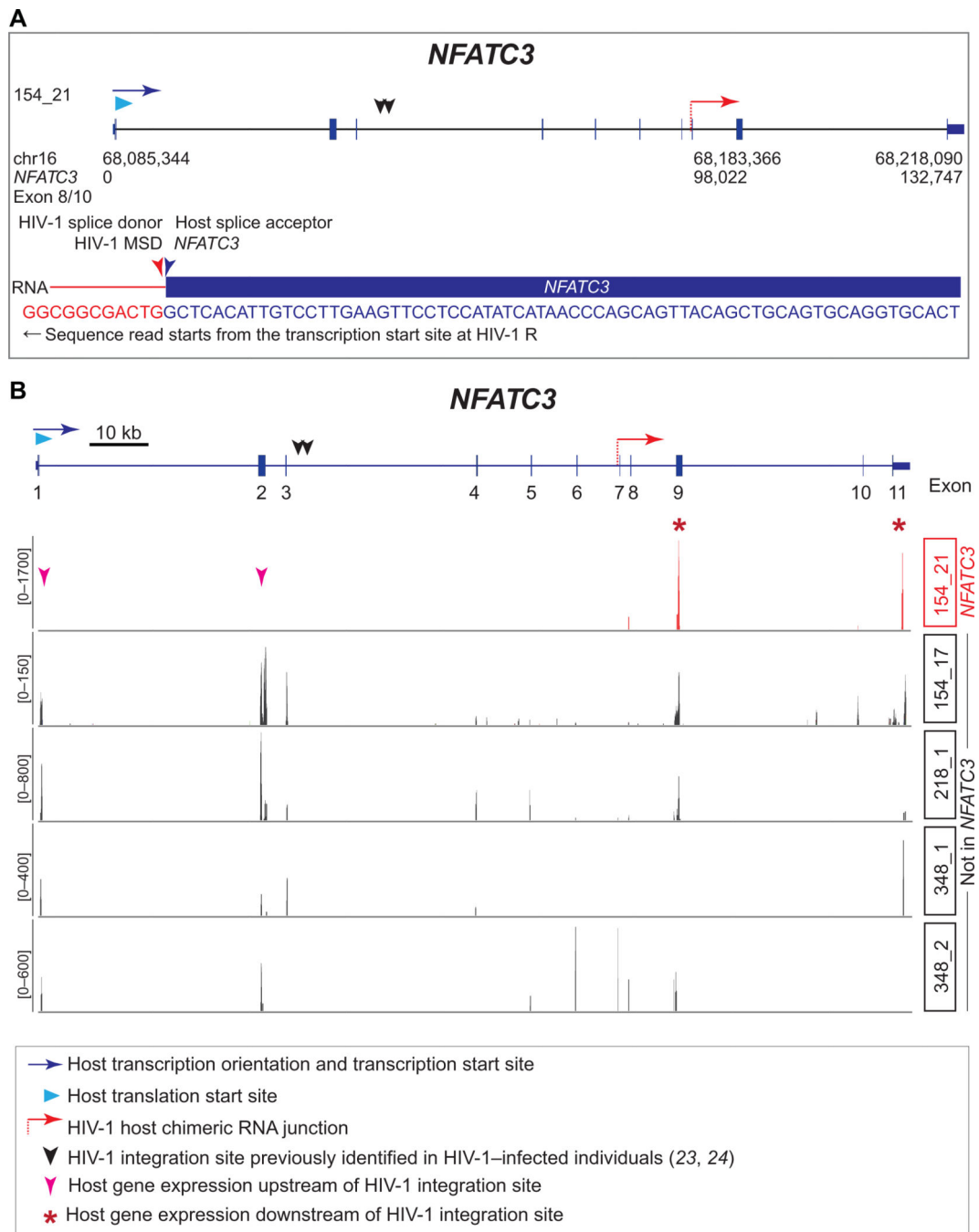
Fig. 5. HIV-1 dominates over the host promoter and drives *NFATC3* expression downstream of HIV-1 integration site through aberrant splicing.
(**A**) HIV-1 LTR drives the transcription of *NFATC3* through aberrant splicing from HIV-1 major splice donor (MSD) into host canonical splice acceptor site identified in HIV-1 SortSeq cells from an ART-treated, virally suppressed, HIV-1–infected individual. The HIV-1–host chimeric RNA junction is indicated as a dashed red arrow. (**B**) Transcription in the single cell (154_21) in which HIV-1 integrated into *NFATC3*. The expected transcription
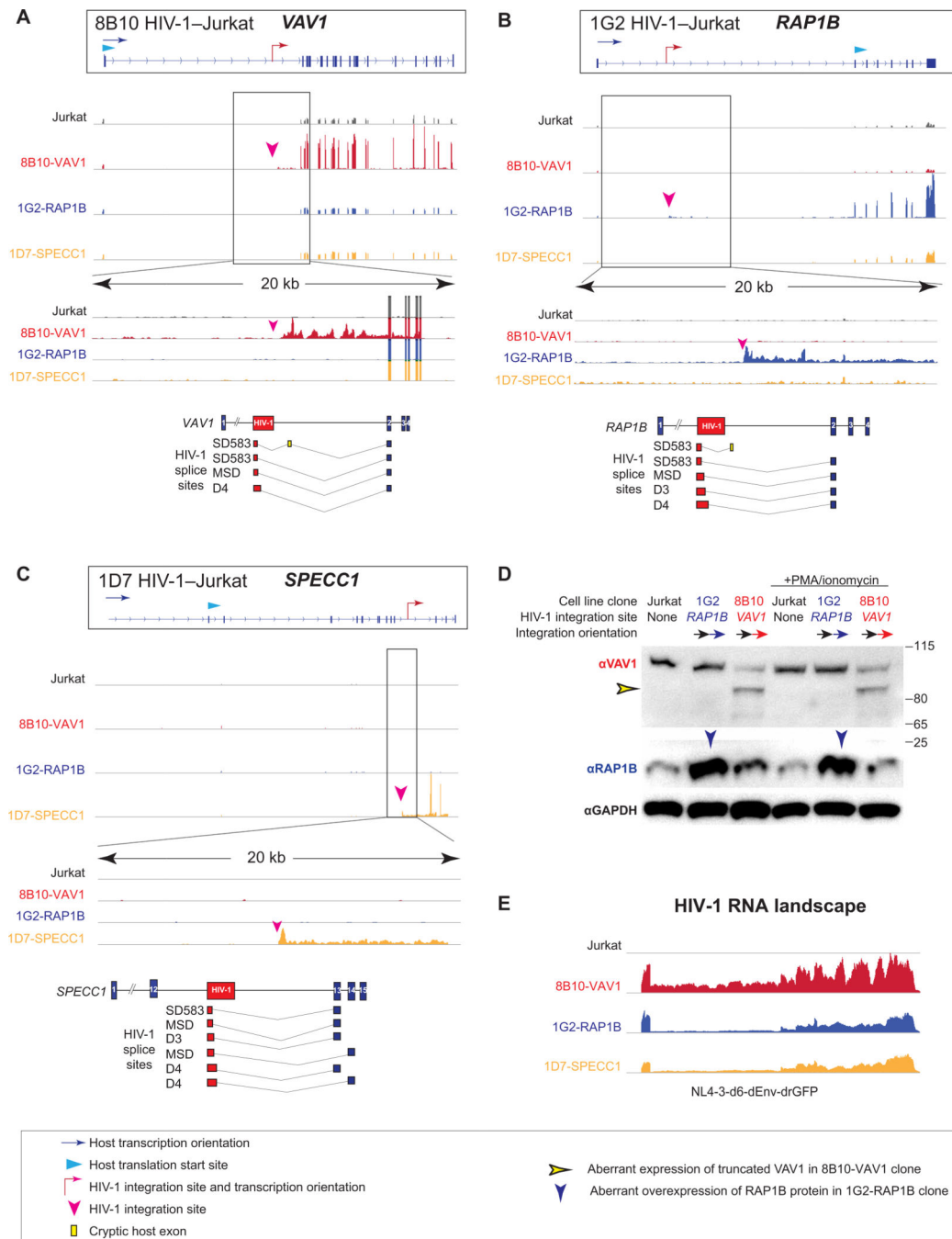
of upstream exons is indicated by magenta arrowheads, and the transcription of the downstream exons is indicated by red asterisks.

Fig. 6. HIV-1 drives aberrant cancer-related gene transcription and induces aberrant protein expression at the integration site.

(**A** to **C**) HIV-1 integration effect on the transcription of cancer-related genes *VAV1* in HIV-1–Jurkat clone 8B10 (A), *RAP1B* in HIV-1–Jurkat clone 1G2 (B), and *SPECC1* in HIV-1–Jurkat clone 1D7 (C) at the integration sites. Normalized RNA transcription landscapes with enlarged 20-kb windows across integration sites. HIV-1 proviruses are integrated downstream of the translation start site of *VAV1* and *SPECC1* in HIV-1–Jurkat clones 8B10 and 1D7, respectively. HIV-1 integration is upstream of the translation start site

of *RAP1B* in HIV-1–Jurkat clone 1G2. HIV-1–host chimeric RNA captured in RNA-seq is depicted as HIV-1 RNA (red boxes), splice junction (gray lines), canonically spliced host exons (blue boxes), and cryptic host exons (yellow bars). (**D**) Western blot of VAV1 and RAP1B expression in HIV-1–Jurkat clones. The blue arrowhead indicates HIV-1–driven RAP1B protein expression, and the yellow arrowhead indicates HIV-1–driven truncated VAV1 protein expression. αGAPDH, anti–glyceraldehyde-3-phosphate dehydrogenase. (**E**) HIV-1 genomic RNA landscape in HIV-1–Jurkat T cell clones.
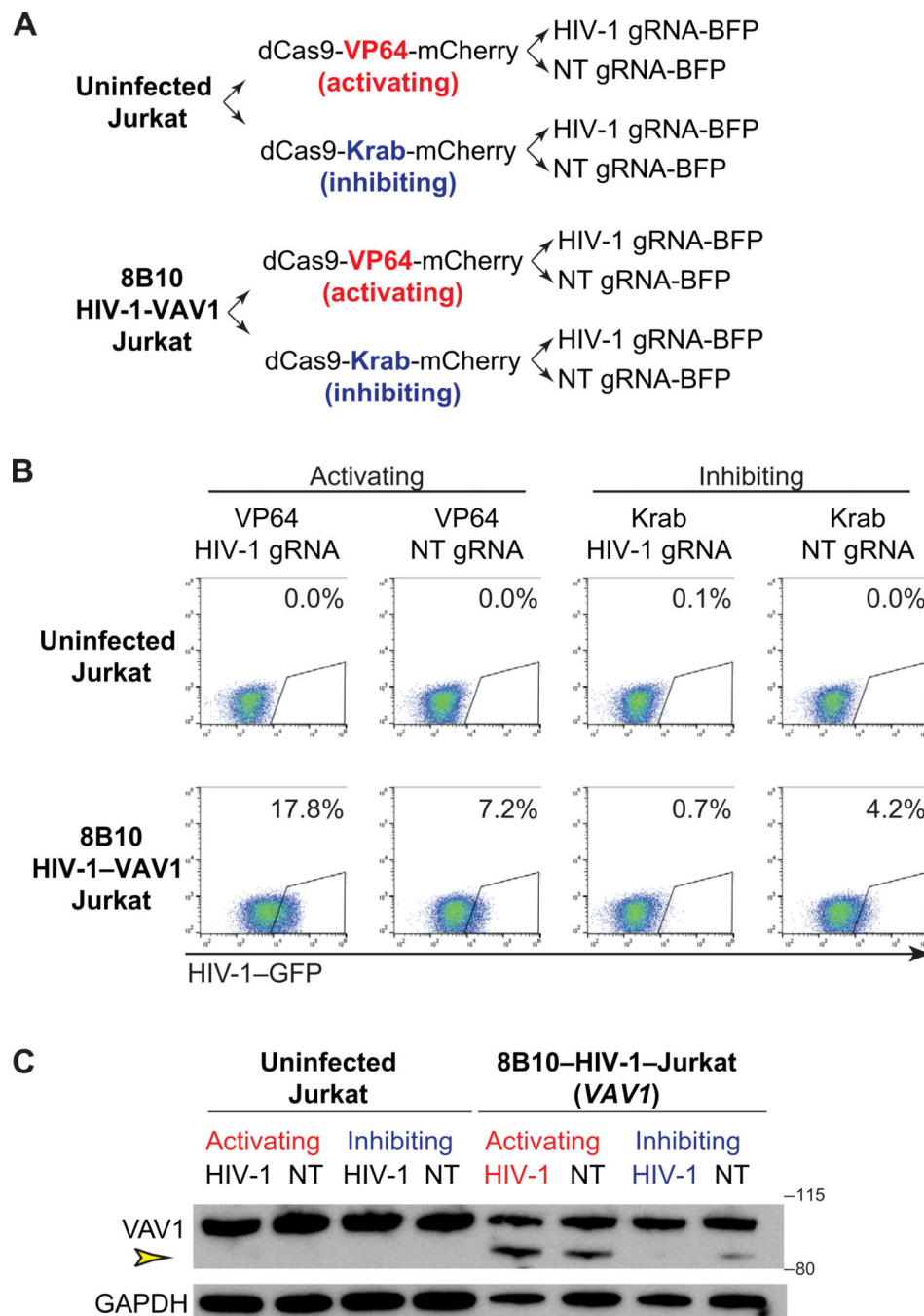
Fig. 7. HIV-1–driven aberrant host gene transcription at the integration site can be suppressed by CRISPR-dCas9–mediated HIV-1 LTR inhibition.

(**A**) CRISPR-dCas9–based HIV-1 LTR-specific activation and inhibition system. Uninfected Jurkat T cells and HIV-1–Jurkat T cell clone 8B10 (in which HIV-1 is integrated into *VAV1*) were transduced with dCas9-VP64-mCherry or dCas9-Krab-mCherry and isolated by flow cytometric sorting. These CRISPR-ready cells were then transduced with lentiviruses carrying HIV-1–specific guide RNA (gRNA) targeting HIV-1 LTR or non-targeting (NT) gRNA as the negative control. (**B**) dCas9-VP64–mediated HIV-1 activation and dCas9-

Krab–mediated HIV-1 inhibition as measured by flow cytometry. (**C**) The effect of dCas9-mediated HIV-1 activation and repression on aberrant VAV1 protein expression. Yellow arrowhead denotes HIV-1–driven aberrant and truncated VAV1 protein.

Fig. 8. CRISPRi-mediated HIV-1–specific inhibition restores HIV-1–driven aberrant host gene transcription to that of uninfected cells.

From CRISPR-ready, gRNA transduced uninfected and HIV-1–infected Jurkat T cell clones 8B10 (in which HIV-1 is integrated into *VAV1*), 1G2 (in which HIV-1 is integrated into *RAP1B*), and 1D7 (in which HIV-1 is integrated into *SPECC1*); HIV-1–green fluorescent protein–positive (GFP[+]) cells in CRISPRa/HIV-1 gRNA and CRISPRa/nontargeting gRNA systems were sorted for RNA landscape analysis at the integration site. HIV-1–GFP[−] cells in CRISPRi/HIV-1 gRNA and CRISPRi/nontargeting gRNA systems were sorted similarly.

Peaks show normalized RNA transcription in the corresponding genes *VAV1* (**A**), *RAP1B* (**B**), and *SPECC1* (**C**).