### Title

Genome-wide CRISPR/Cas9 transcriptional activation screen identifies a histone acetyltransferase inhibitor complex as a regulator of HIV-1 integration




### Journal

Nucleic Acids Research, 50(12)

### ISSN

0305-1048



### Authors

Zhang, Qiong

Wang, Shaobo

Li, Wanyu

et al.



### Publication Date

2022-07-08

### DOI

10.1093/nar/gkac464

*Published online 17 June 2022* *Nucleic Acids Research, 2022, Vol. 50, No. 12* ***6687–6701***
*https://doi.org/10.1093/nar/gkac464*

# Genome-wide CRISPR/Cas9 transcriptional activation screen identifies a histone acetyltransferase inhibitor complex as a regulator of HIV-1 integration

**Qiong Zhang[1,†], Shaobo Wang[1,†], Wanyu Li[1], Edwin Yau[1], Hui Hui[1], Parmit Kumar Singh[2,3], Vasudevan Achuthan[2,3], Maile Ann Young Karris[4], Alan N. Engelman[2,3] and Tariq M. Rana[1,*]**

[1]Division of Genetics, Department of Pediatrics, Program in Immunology, Institute for Genomic Medicine, UCSD Center for AIDS Research, University of California San Diego, 9500 Gilman Drive MC 0762, La Jolla, CA 92093, USA, [2]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, [3]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA and [4]Division of Infectious Diseases, Department of Medicine, UCSD Center for AIDS Research, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Received October 06, 2021; Revised April 28, 2022; Editorial Decision May 15, 2022; Accepted June 14, 2022

## ABSTRACT

**The retrovirus human immunodeficiency virus-1 (HIV-1) is the causative agent of AIDS. Although treatment of HIV/AIDS with antiretroviral therapy provides suppression of viremia, latent reservoirs of integrated proviruses preclude cure by current antiviral treatments. Understanding the mechanisms of host–viral interactions may elucidate new treatment strategies. Here, we performed a CRISPR/Cas9 transcriptional activation screen using a high-complexity, genome-wide sgRNA library to identify cellular factors that inhibit HIV-1 infection of human CD4+ T cells. MT4 cells were transduced with a CRISPR/Cas9 sgRNA library and infected with nef-deficient HIV-$1_{NL4-3}$ expressing ganciclovir-sensitive thymidine kinase, thus enabling selection of HIV-1-resistant cells for analysis of enriched sgRNAs. After validation of screen hits, multiple host factors essential for HIV-1 infection were identified, including SET (SET nuclear proto-oncogene) and ANP32A (acidic nuclear phosphoprotein 32A, PP32A), which together form a histone acetylase inhibitor complex. Using multiple human cell lines and peripheral blood mononuclear cells (PBMCs) from healthy donors and HIV-1-infected individuals, we demonstrate that SET depletion increased HIV-1 infectivity by augmenting DNA integration without significantly changing sites of integration. Conversely, SET overexpression decreased HIV-1 integration and infectivity. SET protein expression was significantly reduced in PBMCs from HIV-1-infected individuals and was downregulated by HIV-1 infection of healthy donor cells *in vitro*. Notably, HIV-1-induced downregulation of SET could be alleviated by inhibition of the protease granzyme A. Altogether, we have identified cellular inhibitors of HIV-1 infection on a genome-wide scale, which affords new insight into host–virus interactions and may provide new strategies for HIV-1 treatment.**

## INTRODUCTION

Human immunodeficiency virus-1 (HIV-1) is a retrovirus that causes acquired immunodeficiency syndrome (AIDS). Despite the success of antiretroviral therapy in improving clinical outcomes and limiting HIV-1 transmission, the persistence of a long-lived proviral reservoir hampers efforts to effect a cure. Thus, there is a continuing need to understand host–HIV-1 interactions, particularly the mechanisms underlying viral DNA integration into the human genome, to identify therapeutic and potentially curative targets. TEST

Upon cell infection, the HIV-1 RNA genome is reverse transcribed and viral DNA is subsequently integrated into the host genome by HIV-1 integrase. Integration is a key step in retroviral infection as it establishes a latent pool of proviral DNA. In addition to viral proteins, several cellular proteins have been demonstrated to play essential roles in this step (1). For example, the transcriptional coactivator lens epithelium-derived growth factor (LEDGF)/p75 interacts with lentiviral integrase to target integration into the mid-bodies of active genes (2,3). RAD51, a component of

*To whom correspondence should be addressed. Tel: +1 858 246 1100; Fax: +1 858 246 1600; Email: trana@ucsd.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

the endogenous homologous recombination complex, has been shown to suppress HIV-1 integration by displacing integrase and disrupting the integration complex (4,5). Nevertheless, our understanding of HIV-1 integration remains incomplete, and delineation of the underlying molecular mechanisms may provide new insights into the design of viral integration-blocking agents.

Advances in CRISPR/Cas9-based knockout and activation screening methods have enabled the identification of many host factors that play crucial roles in retrovirus infection and maintenance of the viral life cycle. Zhu *et al*. performed a genome-wide CRISPR/Cas9 knockout screen to identify factors essential for repression of unintegrated retroviral viral DNA in human cells (6). They identified a number of factors, including the large DNA-binding protein NP220, the histone methyltransferase SETDB1 and three proteins of the epigenetic modulator human silencing hub (HUSH) complex (MPP8, TASOR and PPHLN1) as host factors that are required for unintegrated retroviral DNA silencing. Further biochemical analysis by chromatin immunoprecipitation showed that NP220 recruited the HUSH complex, SETDB1 and the histone deacetylases HDAC1 and HDAC4 to silence the unintegrated retroviral DNA (6). Park *et al*. performed a genome-wide CRISPR/Cas9 screen of HIV-1-infected T cells, which identified host dependency factors including the known CD4 and CCR5 co-receptors as well as tyrosylprotein sulfotransferase 2 (TPST2), solute carrier family 35 member B2 (SLC35B2), and activated leukocyte cell adhesion molecule (ALCAM) (7). Another CRISPR/Cas9 screen for HIV-1 inhibitory factors in latently infected J-LAT A2 cells, which leveraged a single-guide RNA (sgRNA) library specifically targeting nuclear proteins, identified the histone demethylase MINA53 as a novel factor promoting HIV-1 latency (8). Thus, CRISPR/Cas9 knockout screens have proven powerful tools to dissect HIV-1–host interactions.

In the present study, we sought to identify host factors that restrict HIV-1 replication in human CD4+ T cells and may therefore be candidate therapeutic targets. We conducted a CRISPR/Cas9 transcriptional activation screen in an HIV-1-susceptible human T cell line using a high-complexity, genome-wide sgRNA library, and we identified several novel host factors that conferred protection from HIV-1 infection. We investigated in detail the mechanism of restriction by the histone acetylase inhibitor SET (SET nuclear proto-oncogene), validated our findings in multiple human cell lines, and confirmed their pathophysiological relevance using peripheral blood cells from healthy donors and HIV-1-infected individuals.

## MATERIALS AND METHODS

### Blood donors and PBMC activation

This study was approved by the University of California, San Diego Institutional Review Board. Blood samples from healthy donors and HIV-1-infected individuals were obtained from the UCSD AntiViral Research Center. Informed consent was obtained from donors or their parent/legal guardian under a protocol approved by the Human Research Protection Program at UCSD. Peripheral blood mononuclear cells (PBMCs) were isolated by density centrifugation of blood samples on Ficoll (Invitrogen) according to the manufacturer's recommendations. The interface layer was removed, washed with phosphate-buffered saline (PBS), and frozen until used. For *in vitro* T cell activation, $10^6$ PBMCs were incubated with 10 μl Immunocult Human CD3/CD28 T Cell Activator (Stemcell Technologies) for 1 week. To ensure robust HIV-1 infection, the cells were re-stimulated by addition of the activator (10 μl/$10^6$ cells) for 24 h before infection.

### Cell culture and virus preparation

The human CD4+ T cell lines MT4, H9, and Jurkat were obtained from ATCC and cultured in RPMI 1640 medium containing 10% fetal bovine serum (FBS, Thermo Fisher) and 50 μM β-mercaptoethanol (Sigma). 293FT cells were obtained from ATCC and cultured in DMEM (Invitrogen) containing 10% FBS. Microglia CHME cells were obtained from Jon Karn lab. For production of HIV-1 particles, 293FT cells were seeded in 10-cm plates at $6 \times 10^5$ cells/ml and transfected with 15 μg pLAI.2, which encodes for replication-competent, X4-tropic HIV-1 strain LAI (HIV-$1_{LAI}$; # 2532, NIH AIDS Reagent Program) or HIV-TK (#98135, NIH AIDS Reagent Program) vectors using Lipofectamine 3000 (Invitrogen). HIV-1 strain Wilmington (#ARP-6914) and strain RF (#ARP-2803) were also provided by the NIH AIDS Reagent Program. For production of HIV-1 pseudovirus, 9 μg of the replication defective, single round construct HIVpp-luc (pNL4-3.Luc.R-E-, #3418, NIH AIDS Reagent Program) was cotransfected with 3 μg pMD2.G into 293FT cells using Lipofectamine 3000. After 2 days of cell culture, virus-containing supernatants were collected, centrifuged at 2000 g for 10 min, and filtered through 0.22 μm filters. Supernatants were incubated with DNase I (100 U/ml, NEB) and RNase A (100 U/ml, Qiagen) at 37°C for 1 h to digest free DNA and RNA. Virus samples were stored at −80°C until use.

### SAM CRISPR/Cas9 screen

The puromycin-resistant pooled sgRNA plasmid library (Addgene #1000000074), Cas9-VP64 fusion plasmid (Addgene #61425), and MS2-P65-HSF1 plasmid (Addgene #61426) were purchased from Addgene and amplified in *E.coli* cells. To produce the lentivirus bearing the sgRNA library, 293FT cells were re-suspended in Opti-MEM medium and co-transfected with 12 μg of the plasmid library, 9 μg psPAX2, and 6 μg pMD2.G using Lipofectamine 3000 (Invitrogen). After 6 h, the medium was exchanged for fresh DMEM medium and the cells were incubated for 48 h. The virus-containing supernatant was harvested, centrifuged at 2000 g at 4°C for 10 min, filtered through a 0.22 μm filter, and stored at −80°C. Viral titer was determined by measuring puromycin resistance of MT4 cells transduced with 10-fold serial dilutions of virus (9). To generate MT4 cells stably expressing dCas9-VP64 and MS2-p65-HSF1, cells were transduced with dCas9-VP64 and MS2-p65-HSF1 lentivirus and selected with hygromycin B (200 μg/ml) and blasticidin (5 μg/ml) for 2 weeks. Thereafter, MT4 cells stably expressing dCas9-VP64 and MS2-p65-HSF1 ($3 \times 10^8$ cells) were transduced with

the sgRNA library lentivirus at an MOI of 0.3 (corresponding to ∼30% puromycin selection survival rate) to reach a coverage of 500 for each sgRNA. Transduced cells were selected with 1 μg/ml puromycin for 7 days, with splitting every 2 days. The cells were then infected with HIV-TK ($6 \times 10^7$ cells/sample) at an MOI of 1 for 4 days and GCV was then added at 250 μg/ml for an additional 3 days. Surviving cells were isolated by sorting using a viable cell separation kit (Miltenyi Biotec). The cells were placed back in culture, expanded for an additional 1 week in GCV-containing medium, infected again with HIV-TK at an MOI of 1 for 4 days, and treated again with GCV for 3 days. The surviving cells, designated HIV-resistant cells, were separated using the kit and collected for sequencing. For mock infection, samples of 30 million cells were collected for sequencing.

### Quantification of sgRNA enrichment by deep sequencing

Sequencing was performed as previously described (10). Briefly, genomic DNA was extracted from MT4 cells and sgRNA integration was quantified using a two-step nested PCR process. In PCR#1, 10 μg genomic DNA per reaction was used to amplify the lentiviral sequence containing the 20 bp sgRNA cassette. A total of 20 reactions were performed (200 μg genomic DNA) to maintain full library representation. In PCR#2, 10 μl of pooled PCR1 product was used for each reaction, with one reaction per $10^4$ constructs (seven reactions per sample). PCR products were purified and sequenced using an Illumina NextSeq system.

### Knockdown and overexpression vectors

All sequences are listed in Supplementary Table S1. Short-hairpin (sh) RNAs were purchased from the La Jolla Institute for Immunology. SET was cloned into the pLVX vector using an In-Fusion kit (Takara). CRISPR vectors were generated using the Zhang lab protocol (http://sam.genome-engineering.org/protocols/). For shRNA, sgRNA, and overexpression lentivirus production, 293FT cells were seeded at $6 \times 10^5$/well into six-well plates for 12 h and then transfected with 1.8 μg of the appropriate lentiviral vector mixed with two packaging vectors (1.2 μg psPAX2 and 0.6 μg pMD2.G) in Opti-MEM using Lipofectamine 3000. At 6 h after transfection, the Opti-MEM was exchanged for DMEM containing 10% FBS and the cells were incubated for an additional 48 h. Supernatants were collected, centrifuged at 2000 g for 10 min, and stored at –80°C until use. For cell transduction, MT4 or Jurkat cells were seeded at $4 \times 10^5$ cells/well in 12-well plates and 500 μl of virus supernatant plus 4 μg/ml Polybrene were added per well. The plates were centrifuged at 750 g for 45 min at room temperature and the cells were incubated at 37°C. After 24 h, the medium was exchanged and the cells were cultured for at least 3 days before analysis.

### CRISPR/Cas9-mediated SET knockout or activation

For CRISPR-mediated SET knockdown (KD), three SET-targeting sgRNAs (Supplementary Table S1) or control (11) RNAs were synthesized and annealed for cloning into lenti-CRISPR v2 (Addgene #52961). Positive clones containing sgRNAs were sequenced. To produce the lentiviruses bearing the sgRNA, 293FT cells were seeded at $6 \times 10^5$/well into six-well plates for 12 h and then transfected with 1.8 μg the sgRNA vectors along with two packaging vectors (1.2 μg psPAX2 and 0.6 μg pMD2.G) in Opti-MEM using Lipofectamine 3000. MT4 or Jurkat cells were transduced with sgSET or sgNC lentiviruses and selected with puromycin (1 μg/ml) for 7 days. For generation of an sgSET Jurkat clone, single cells were sorted into 96-well plates and cultured for 2 weeks. SET KD clones were identified by western blot analysis with anti-human SET antibody (Abcam, ab181990, 1:2000) and the cells were sequenced to confirm biallelic modification of the SET locus.

For CRISPR-mediated SET overexpression, the two most significantly enriched SET sgRNAs obtained in the MT4 screen (Supplementary Table S2) were synthesized and annealed for cloning into lenti sgRNA(MS2)_puro plasmid (Addgene, #73795). Lentiviruses were produced and transduced into dCas9- and MS2-expressing MT4 cells as described above. Cells were selected in puromycin (1 μg/ml) and increased expression was confirmed by Western.

### Expression of *in vitro*-assembled Cas9-RNP complex in human PBMCs

For *in vitro* assembly of the sgRNA-Cas9-RNP complex, Cas9 (Alt-R® S.p. HiFi Cas9 Nuclease V3), SET-specific crRNA (GAGCAGCACCAUGUCGGCGCGUUUUAGAGCUAUGCU) or a non-targeting control crRNA (11), tracrRNA (Alt-R® CRISPR-Cas9 tracrRNA), and electroporation enhancer (Alt-R® Cas9 Electroporation Enhancer) were purchased from Integrated DNA Technologies. Electroporation kit was purchased from Invitrogen. The Alt-R CRISPR-Cas9 crRNA or the Alt-R CRISPR-Cas9 tracrRNA was dissolved in IDTE Buffer to final concentrations of 200 μM. To generate the sgRNA, the crRNA (NC or SET-specific crRNA) was mixed with the tracrRNA in equimolar concentrations in a sterile PCR tube to a final duplex at concentration of 100 μM. To anneal the two RNA oligos into sgRNA, the two RNA oligo mixture was heated at 95°C for 5 min, removed from heat, and allowed to cool to room temperature. To prepare the RNP complexes, 5 μl Cas9 protein (at final concentration of 5 μM), 5 μl guide RNAs (final concentration of 50 μM), and 1 μl electroporation enhancer (final concentration of 10 μM) were incubated at room temperature for 10–20 min. To deliver the RNP into the PBMC cells, activated PBMC cells were washed and re-suspended in electroporation buffer R at a density of $2 \times 10^7$ cells/ml. RNP complexes (10 μl), added into 50 μl cells, and electroporated into cells at 1600 V, 10 ms, 3 plus. Electroportaed cells were quickly added into 500 μl pre-warmed culture media in 24-well plate and cells were re-suspended by gently pipetting up and down. The cells were cultured for 6 days to enable sufficient editing and SET expression was detected by Western blotting.

### Integration site sequencing and bioinformatics

Sites of HIV-1 integration in human genomic DNA were determined from raw Illumina sequence reads using previ-

ously published pipelines (12–17). Control or SET-depleted MT4 or PBMCs ($2 \times 10^6$) were infected with HIVpp-luc for 72 h, and genomic DNA extracted using a DNeasy Blood and Tissue Kit (Qiagen) was digested with MseI and BglII and purified using a QIAquick PCR Purification Kit (Qiagen). Linker DNA was ligated to the digested cellular DNA overnight at 16°C, and Illumina P7 and P5 adapter sequences were added by semi-nested PCR. PCR products were purified and sequenced at Genewiz using the Illumina NextSeq platform.

Following HIV-1 U5 and linker sequence identification and demultiplexing, the trimmed sequences were aligned to human genome build hg19 by BWA-MEM with pair-end option. From the aligned reads, HIV-1 integration site coordinates were determined in bed format as previously described (12–17). The bedfiles were used to determine HIV-1 integration site distributions with respect to defined genomic features in the human genome including Refseq genes and gene-dense regions of chromosomes (12–17). Peak coordinates for histone marks H3K4me1, H3K36me3, and H3K27ac were downloaded for hg19 and Dnd41 cells from UCSC genome Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables). HIV-1 integration site proximity to histone markers was calculated by BEDtools (18). Coordinates of 10 SPIN (Spatial Position Inference of the Nuclear genome) states, which were downloaded from https://github.com/ma-compbio/SPIN, were converted to the hg19 genome as described previously (19). Genomic locations of speckle-associated domains (SPADs) were as previously described (12,13). Lamina-associated domain (LAD) coordinates were provided by the authors (20). Percent integration frequencies with respect to SPADs, LADs and SPIN states were calculated using BEDtools. Random integration control (RIC) values were generated following *in silico* digestion of the hg19 reference genome with MseI and BglII (12).

### Western blot analysis

MT4, H9 or PBMC lysates were prepared in IP lysis buffer (Thermo Fisher) and centrifuged. Supernatants were collected and protein concentrations were measured using DC Protein Assay (Bio-rad). Proteins (25 μg) were separated using 4–12% pre-cast SDS-PAGE gels (Invitrogen) and transferred to PVDF membranes. The membranes were blocked in 5% FBS at room temperature for 1 h, washed, and incubated with primary antibodies at 4°C overnight or at room temperature for 2 h. The membranes were washed and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (Pierce Fast Western Blotting Blot Kit) for 15 min at room temperature. Finally, the blots were washed three times and proteins were visualized using ECL Substrate (Pierce Fast Western Blotting Blot Kit). The antibodies and dilutions used include: anti-HIV p24 (1:1000, Abcam, ab9071), HRP-conjugated anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH-HRP; 1:5000, Proteintech, HRP-60004), anti-SET (1:2000, Abcam, ab181990), anti-p53 (1:1000, Thermo Fisher, MS-104-P0), and anti-tubulin-HRP (1:5000, Proteintech, HRP-66031).

### Flow cytometry

For analysis of intracellular SET expression in human PBMCs, cryopreserved PBMCs from HIV-infected individuals or healthy donors were rapidly thawed at 37°C and resuspended in RPMI 1640 medium supplemented with 10% FBS and 100 U/ml DNase I (Stem Cell Technologies). After washing, the cells were incubated with the viability dye Zombie Aqua, AF700-conjugated anti-CD3 (clone OKT3) and APC-Cy7-conjugated anti-CD4 (clone RPA-T4), all from BioLegend. The cells were washed again, fixed, and permeabilized using BioLegend staining buffer, FluoroFix buffer, and permeabilization wash buffer. The cells were then incubated with rabbit anti-SET antibody EPR12973 (Abcam, ab181990 or ab237445, 1:40). The cells were finally washed with staining buffer and analyzed on a FACSCanto flow cytometer (BD Biosciences, San Jose, CA, USA). For analysis of GZMA in HIV-infected PBMCs, PBMCs were stimulated with Immunocult Human CD3/CD28 T Cell Activator (Stemcell Technologies) and incubated for 1 week. The cells were stimulated with T cell activator again and infected with HIV-$1_{LAI}$. Three days post-infection (p.i.), T cells were collected and incubated with the viability dye Zombie Aqua, pacific blue-conjugated anti-CD3 (clone OKT3), APC-conjugated anti-CD4 (clone RPA-T4), and PerCP/cy5.5-conjugated anti-CD8 (clone RPA-T8). The cells were washed, fixed, and permeabilized. Cells were then incubated with FITC-conjugated anti-GZMA antibody (clone CB9, 507212, BioLegend). The cells were finally washed with staining buffer. Gating was performed using the fluorescence-minus-one or isotype-stained control methods, and compensation was performed using UltraComp beads (Thermo Fisher) individually coupled to each antibody. Data analysis was performed using FlowJo v10 software (FlowJo LLC).

### Quantitative reverse-transcription PCR (RT-qPCR)

To detect SET or HIV *env* expression, total cellular RNA was extracted with TRIzol (Invitrogen) according to the manufacturer's instructions. Cellular RNA (1 μg) was reverse transcribed using iScript cDNA Synthesis Kit (Bio-Rad) and RT-qPCR was performed using 2× SYBR Green mix (Bio-Rad) on a LightCycler480 System (Roche Life Science). PCR cycling conditions were 95°C for 5 min followed by 40 cycles of 95°C for 10 s, 60°C for 10 s and 72°C for 10 s. RT-qPCR primer sequences are listed in Supplementary Table S1.

### Late RT qPCR and Alu qPCR

To analyze the early steps of the HIV-1 life cycle (late RT and integration), we used the methods developed by the Bushman group (21). Genomic DNA was extracted from MT4 cells or PBMCs infected with HIVpp-luc at various times p.i. as described above. Late RT was detected using MH531 and MH532 primers. Integration was detected by Alu-qPCR. Alu-Gag segments in 100 ng genomic DNA were first amplified using a forward Alu primer (5′-GCCTCCCAAAGTGCTGGGATTACA-3′) and reverse

Gag primer (5′-GTTCCTGCTATGTCACTTCC-3′). HIV-1 Gag DNA that had integrated proximal to Alu elements was then quantified by qPCR using U5 primers. DNA levels were normalized to genomic GAPDH.

### Detection of HIV-1 infection by p24 ELISA and luciferase assay

For MT4 cells or PBMCs infected with HIV-$1_{LAI}$, culture supernatants were collected at the indicated times p.i. and p24 levels were detected using an HIV-1 p24 ELISA kit (RETROtek) according to the manufacturer's protocol. For cells infected with HIVpp-luc, the cells were collected at the indicated times p.i., washed with PBS, and lysed with Passive Lysis Buffer (Promega). Luciferase expression was measured using a Luciferase Assay kit (Promega).

### Chromatin immunoprecipitation (ChIP)–qPCR

H3ac-ChIP and H3-ChIP were performed to assess the association between HIV-1 DNA and acetylated H3 (H3ac) and total H3. Briefly, MT4 cells ($2 \times 10^7$) were infected with HIV-$1_{LAI}$ at an MOI of 0.2 for 3 days and then harvested and fixed with 1% formaldehyde for 10 min at room temperature. The reaction was quenched by addition of glycine at a final concentration of 125 mM. Fixed cells were lysed and sonicated to yield ∼400 bp fragments. Protein G Magnetic Beads (Cell Signaling Technology, #70024) were pre-blocked in 0.5% BSA buffer. Samples of 20 μg chromatin were diluted in 1 mL Dilution buffer (20 mM Tris–Cl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% TritonX-100, 0.01% SDS, 1 mM PMSF) and pre-cleared by 20 μl pre-blocked beads for 1 h at 4°C. After removing the beads from the chromatin, the pre-cleared chromatin was incubated with 3 μg of control rabbit IgG antibody (ab4729, Abcam), rabbit anti-H3 antibody (#1791, Abcam), or rabbit anti-H3ac antibody (#47915, Abcam) for 2 h at 4°C. And then, antibody-chromatin complexes were incubated with 30 μl pre-blocked beads for 2 h at 4°C, washed two times with low salt wash buffer (20 mM Tris–Cl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% TritonX-100, 0.1% SDS, 1 mM PMSF), two times with high salt wash buffer (20 mM Tris–Cl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% TritonX-100, 0.1% SDS, 1mM PMSF), once with LiCl wash buffer ((20 mM Tris–Cl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP40 10%, 1% Na-deoxycholate, 1 mM PMSF) and once with TE buffer (10 mM Tris–Cl, pH 8.0, 1 mM EDTA, 1 mM PMSF). DNA was eluted from the immunoprecipitates using 100 μl of IP elution buffer (50 mM NaHCO3, 1% SDS 10%) for 30 min at room temperature. 8 μl of 4 M NaCl was added into the 100 μl eluent (final concentration of 0.3 M) and incubated at 67°C for 8 h to reverse crosslink genomic DNA. The genomic DNA was then purified using PCR purification kit (Qiangen). HIV-1 DNA was quantified by qPCR using the primers listed in Supplementary Table S1.

### Statistical analysis

Data were analyzed using Prism version 8.0 software (GraphPad). Data are presented as the mean ± standard deviation (SD) unless indicated, and group mean differences were analyzed using Student's *t* test. $P < 0.05$ was considered statistically significant.

Fisher's exact test was used to calculate *P* values associated with statistically significant differences in integration distribution with respect to genomic features, except for gene-density comparisons, where the Wilcoxon-sum rank test was applied.

## RESULTS

### Identification of HIV-1 restriction factors using a genome-wide CRISPR-Cas9 transcriptional activation screen

To identify cellular factors restricting HIV-1 replication, we performed a previously described CRISPR/Cas9 gain-of-function transcriptional activation method, which is based on synergistic activation mediator complex (SAM) (9). The three components of SAM-mediated activation are inactive dCas9-VP64 fusion protein, the MS2-P65-HSF1 activation helper protein, and the sgRNA library harboring MS2 RNA aptamers. The human genome-wide sgRNA library was designed to activate expression of all known coding isoforms from the RefSeq database (23,430 isoforms: 3 sgRNAs per gene, 70,290 guides total) (9). To obtain HIV resistant cells, we chose a highly permissive MT4 T cell line, which could be infected at ∼100% efficiency (22). To improve the efficiency of selection of HIV-resistant cells, we used an engineered nef-deficient HIV-$1_{NL4-3}$ construct (referred to as HIV-TK), which expresses the ganciclovir (GCV)-sensitive herpes simplex virus thymidine kinase and therefore allows depletion of all HIV-1-infected cells (23) while preserving HIV-1-resistant cells. The experimental strategy is outlined in Figure 1A.

We first verified whether this infection strategy could enhance the selection of HIV-1-resistant cells. MT4 cells were transduced with non-targeting sgRNA (sgNC) or sgRNA targeting the HIV-1 co-receptor CXCR4 and the cells were then infected with HIV-TK for 4 days. GCV was then added to the cells at various concentrations and viability assays were performed 3 days later. This analysis demonstrated that HIV-TK infection alone caused the death of ∼90% of sgNC-expressing cells by day 7, while the combination of HIV-TK infection and 250 μg/ml GCV treatment resulted in almost 100% loss of viability (Supplementary Figure S1A). In contrast, CXCR4-edited MT4 cells were virtually unaffected by HIV-TK infection in the presence or absence of GCV, as expected (Supplementary Figure S1A). Thus, HIV-resistant MT4 cells could be selected with high efficiency when using HIV-TK infection at an MOI of 1 for 4 days followed by GCV treatment at a concentration of 250 μg/ml for 3 days.

For the sgRNA screen, MT4 cells stably transduced with dCas9-VP64 and MS2-p65-HSF1 lentiviruses were selected for 2 weeks, and then transduced with the sgRNA lentiviral library for 7 days, infected with HIV-TK (or mock-infected) for 4 days, and treated with GCV 250 μg/ml for an additional 3 days (Figure 1A). The surviving cells were then sorted by using a viable cell separation kit, expanded for 1 week, and then subjected to a second round of HIV-TK infection for 4 days followed by GCV selection for 3 days. The surviving cells, considered to be HIV-1-resistant, were sorted and genomic DNA was extracted for PCR amplifi-

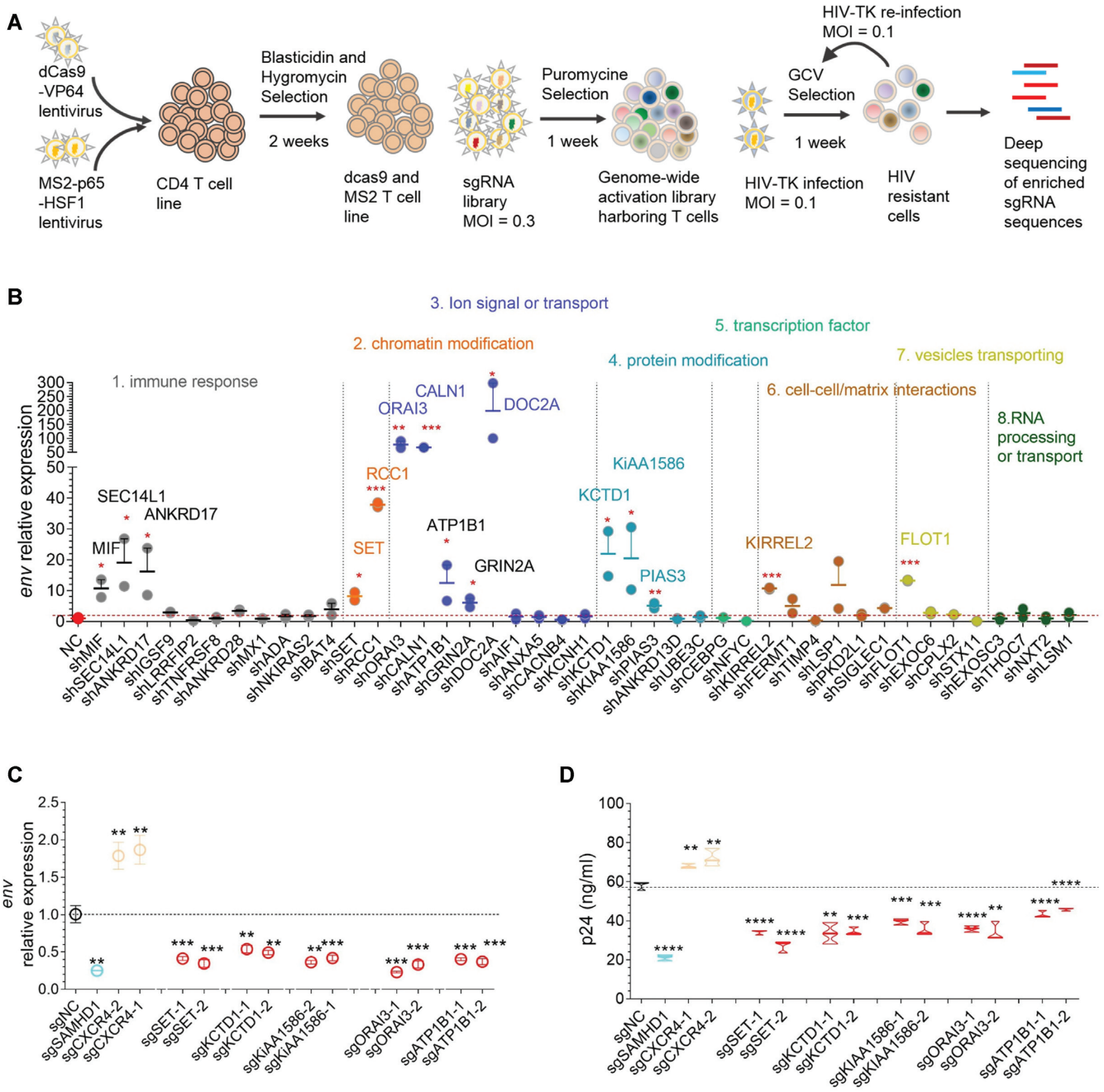


**Figure 1.** Identification of HIV inhibitory factors by genome-wide SAM CRISPR/Cas9 screening. (**A**) Schematic representation of the genome-wide SAM CRISPR/Cas9 screen of human MT4 cells. A stable MT4 cell line expressing dCas9-VP64 and MS2-p65-HSF1 was transduced with the sgRNA library, infected with HIV-TK, and then treated with ganciclovir (GCV) to further deplete HIV-infected cells. The HIV-TK infection and GCV selection cycle was repeated, and surviving HIV-resistant cells were sorted for analysis. sgRNA libraries were prepared from genomic DNA by nested PCR and sequenced. Differentially enriched sgRNAs were identified and the genes were further analyzed as candidate HIV-1 restriction factors. (**B**) Target hit validation. MT4 cells were transduced with control shRNA (shNC) or the indicated gene-specific shRNAs for 2 days followed by infection with HIV-TK (MOI 0.1) for 3 days. Cells were harvested to extract the RNA, and RT-PCR was performed to quantify the expression of HIV-1 *env* relative to GAPDH. Mean ± SD of $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$, by Student's *t* test. (**C**, **D**) SAM mediated activation of SET, KCTD1, KiAA1586, ORAI3, and ATP1B1 inhibit HIV-1 replication. dCas9 and MS2 MT4 cells were transduced with lentivirus expressing control sgRNA (sgNC) or sgRNA of indicated genes. After 2 days, cells were infected with HIV-$1_{LAI}$ at an MOI of 0.1 for 3 days. Cellular RNA was extracted, and RT-qPCR was performed to quantify the expression of HIV-1 *env* relative to GAPDH (C). The release of HIV-1 particles in the supernatant was detected by p24 ELISA (D). Mean ± SD of $n = 3$. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, by Student's *t* test.

cation and deep sequencing of sgRNAs. The screening results were processed using the comprehensive STARS analysis, which ranks screen hits with consistent enrichment among multiple sgRNAs in HIV-infected cells compared with mock-infected cells (11).

In total, we obtained 2431 candidate genes that were represented by enrichment of at least two of the three targeting sgRNAs per gene (Supplementary Table S2). Included among them was the nucleotide phosphohydrolase SAMHD1 (SAM domain and HD domain-containing protein 1), which is a well-known HIV-1 restriction factor (24,25) and, accordingly, all three SAMHD1-targeting sgRNAs were significantly enriched in HIV-resistant MT4 cells (Supplementary Figure S1B). As expected, the screen effectively counter selected many known HIV-1-dependency factors, including CXCR4, CD4, and three proteins (ALCAM, SLC35B and TPST2) identified by a previous CRISPR/Cas9 knockout screen (7) (Supplementary Figure S1B). These findings served to validate the screening approach and confirmed the ability to enrich for sgRNAs activating expression of HIV-1 restriction factors.

From the list of candidate genes, we chose 46 with the highest STARS scores (>1) for functional validation (Supplementary Table S2), preferentially selecting those annotated in pathways that may affect HIV-1 proliferation, including immune response, chromatin modification, ion signaling, protein modification, transcriptional regulating, cell-to-cell interactions, vesicle transporting and RNA processing or transport (26). To validate the hits, MT4 cells were transduced with three targeting shRNAs per gene, infected with HIV-TK, and then analyzed for HIV-1 infection by measuring viral *env* expression by RT-qPCR (Figure 1B). We found that KD of 15 of the 46 genes resulted in significantly increased HIV-1 *env* expression (Figure 1B and Supplementary Figure S1C). Of these, SET, ORAI3 (ORAI calcium release-activated calcium modulator 3), ATP1B1 (ATPase $Na^+/K^+$ transporting subunit beta 1), KCTD1 (potassium channel tetramerization domain-containing 1), and KIAA1586 (E3 SUMO-protein ligase) were further validated by shRNA-mediated KD in a second human T cell line, H9, which confirmed that their expression restricted HIV-1 *env* expression, suggesting that these factors act in cell line independent manners (Supplementary Figure S1D, E). We selected these five genes for further confirmation using SAM directed activation. Two of the enriched sgRNAs from the screen were used to activate gene expression for each gene, while expression activation of SAMHD1 and CXCR4 was used as positive and negative controls for virus restriction, respectively. HIV-$1_{LAI}$ was used to infect cells for gain-of-function experiments. As expected, *env* expression and p24 levels in infected cell supernatants were suppressed by SAMHD1 overexpression but enhanced by CXCR4 overexpression. HIV-$1_{LAI}$ infection of MT4 cells was significantly inhibited by SET, KCTD1, KIAA1586, ORAI3 and ATP1B1 overexpression, as measured by RT-qPCR analysis of *env* expression and ELISA quantification of HIV p24 protein level in the cell supernatants (Figure 1C and D). Thus, this genome-wide CRISPR/Cas9 transcriptional activation screen identified a series of novel inhibitory factors for HIV-1 infection in human CD4+ T cells.

**SET restricts HIV-1 replication in human T cell lines and PBMCs**

From the chromatin modification category of cell factors, SET robustly inhibited HIV-1 infection (Figure 1B–D) in both MT4 and H9 cells (Supplementary Figure S1D and E). The SET gene encodes four isoforms, of which only isoform 2 was identified as a potential HIV-1 inhibitory factor in our CRISPR/Cas9 activation screen. SET isoforms 1 and 2 are inhibitors of the histone acetylases (HATs) EP300 and GCN5 (27,28). EP300 and GCN5 have been shown to impact HIV-1 replication, although their mechanism of action is unclear (29–32). Thus, we investigated the mechanism of HIV-1 restriction by SET in more detail. To deplete SET, we designed three sgRNAs to target the transcriptional start site (TSS), exon1 and exon3 (Figure 2A). Western blot analysis confirmed that SET protein expression was significantly depleted in sgSET-expressing MT4 cells compared with the control cells (Figure 2B). SET edited MT4 cells were infected with HIV-$1_{LAI}$ at an MOI of 0.1 for a time course analysis of HIV-1 gene expression. A significant increase in p24 was evident (∼10-fold) at 2 days p.i., which reached over 30-fold at 3 days p.i. (Figure 2D). To obtain a T cell line with complete SET KD, we separated single colonies of Jurkat T cells after transduction with sgRNA1, and one clone was characterized as completely depleted for SET expression at the protein level (Figure 2C). This colony was identified to harbor two heterozygous editing sites (Supplementary Figure S2A). To hone in on the early events of HIV-1 infection including RT, integration, and gene expression from the viral *env* position, we used the single-cycle reporter construct HIVpp-luc. Similar to the results observed in bulk MT4 cell cultures, RT-qPCR analysis of *env* revealed that infection by the HIVpp-luc was increased in the SET depleted Jurkat cell clone compared with control cells, which was evident as early as day 1 p.i (Figure 2E).

As mentioned above, SET isoform 2 was the only isoform identified in our screen. To confirm that the effects of SET observed here were indeed mediated by isoform 2, we overexpressed SET isoform 2 in MT4 cells. As expected, infection with HIV-$1_{LAI}$ was reduced by ∼50% in SET isoform 2-overexpressing cells compared with cells overexpressing the control protein GFP (Figure 2F–H). Altogether, these results demonstrate that SET restricts HIV-1 infection.

Next, we sought to establish the pathophysiological relevance of our results by examining HIV-1 infection of human PBMCs. To activate T cells, PBMCs from three healthy donors were incubated *in vitro* with antibodies to CD3 and CD28 and then transduced with three SET-specific shRNAs, each of which reduced SET mRNA and protein levels by ∼80% (Supplementary Figure 2B–G). SET KD cells were then infected with HIV-$1_{LAI}$ for 3 days. Expression of SET shRNAs increased HIV-$1_{LAI}$ infection by about 3–5-fold, 9–22-fold and 4–12-fold in cells from donors 1, 2, and 3, respectively (Figure 2I–K), which was consistent with the effects of SET KD in MT4 and Jurkat cells. Taken together, these results demonstrated that SET, most likely isoform 2, negatively regulates HIV-1 replication not only in human CD4+ T cell lines, but also in PBMCs.

To ascertain the functionality of SET-based restriction beyond lab-adapted strains such as LAI and NL4-3, we next

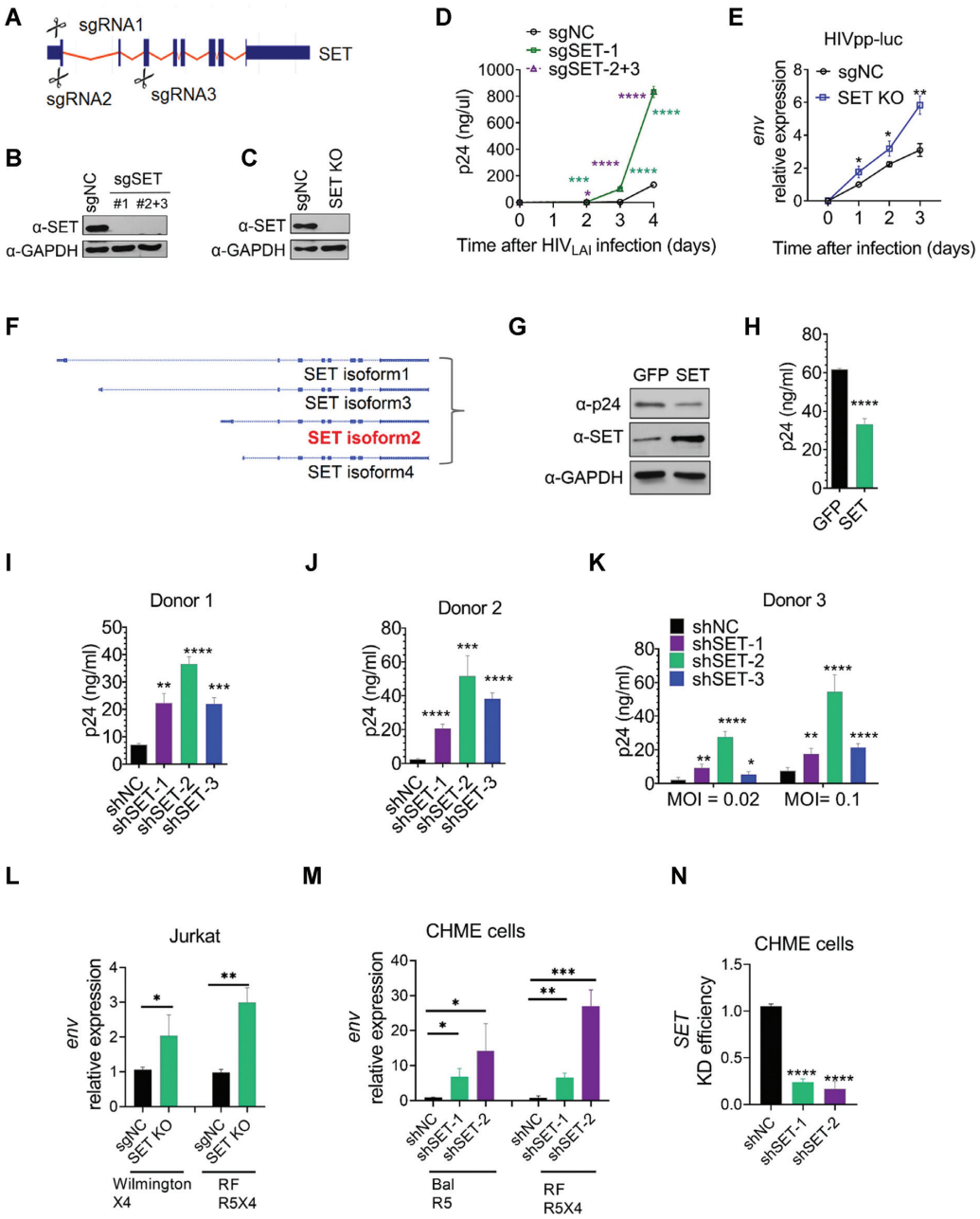


**Figure 2.** SET restricts HIV-1 replication in human CD4+ T cell lines and human PBMCs. (A–E) Knockout of SET enhances HIV-1 replication in CD4+ T cells. (**A**) Illustration of the three sgRNAs used for CRISPR/Cas9-mediated SET KO. (**B**) SET protein expression was analyzed in control sgRNA (sgNC) or SET sgRNA (sgSET #1, or #2 + 3) transduced MT4 cells by Western blotting. (**C**) Characterization of single Jurkat T cell clone knocked out for SET expression. Jurkat cells were transduced with sgSET#1 and SET expression was determined by Western blotting. (**D**) Control or SET depleted cells from panel B were infected with HIV-1$_{LAI}$ at an MOI of 0.01 for 4 days. Time dependent p24 released in the supernatant was detected by p24 ELISA. (**E**) Control or SET knockout cells from panel C were infected with HIV pseudovirus (HIVpp-luc, MOI = 0.2) for 3 days. RNA was collected at the indicated times and HIV-1 *env* expression was quantified by RT-qPCR and normalized to GAPDH. (F–H) Ectopic expression of SET isoform 2 decreases HIV-1 replication. (**F**) Schematic representation of the 4 isoforms of SET. (**G**) SET isoform 2 was overexpressed in MT4 cells and SET expression was determined by Western blotting. (**H**) The cells were infected with HIVpp-luc at an MOI of 0.2. After 2 days, luciferase levels were measured. (**I–K**) KD of SET enhances HIV-1 replication in primary PBMCs from three different donors. Primary PBMCs were isolated from three healthy donors. T cells were activated using CD3 and CD28 antibodies. PBMCs were transduced with lentivirus vectors expressing non-targeting shRNA (shNC) or three SET specific shRNAs (shRNA1-3), followed by infection with HIV-1$_{LAI}$ at an MOI of 0.02 or 0.1 for 3 days. p24 released in the supernatant was detected by p24 ELISA. (**L**) Control or SET depleted cells from panel C were infected with an X4-tropic stain Wilmington and a dual-tropic strain RF at an MOI of 0.01 for 3 days. RNA was collected and HIV *env* mRNA was quantified by RT-qPCR and normalized to GAPDH expression. (**M, N**) KD of SET enhances HIV-1 replication in microglia cells. Lentivirus expressing control shRNA or two SET specific shRNA were transduced into microglia cells (CHME) followed by puromycin selection for 7 days. Stable cell lines were infected with an R5-tropic stain Bal and a dual-tropic strain RF at an MOI of 0.2 for 3 days. *env* mRNA (M) and SET mRNA (N) was detected by RT-qPCR and normalized to GAPDH expression. Data in (B), (C) and (G) are representative of at least two independent experiments. GAPDH was used as a loading control. Mean ± SD of n = 3. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by Student's *t* test.

examined the primary X4 strain Wilmington, the R5 strain Bal, and the R5X4 dual-tropic strain RF. SET knockout increased Wilmington and RF infection of Jurkat T cells (Figure 2L). Moreover, HIV-1 Bal and RF infection was increased significantly in microglial CHME cells following SET KD (Figure 2M, N). These results demonstrated that SET can restrict infection of multiple HIV-1 strains.

### SET restricts HIV-1 infection at the DNA integration step

We next sought to identify the stage of the HIV-1 life cycle at which SET was required. Reverse transcription (RT) and integration of HIV-1 DNA was examined by qPCR and Alu-qPCR, respectively. To restrict this analysis to a single-round of HIV-1 infection, MT4 cells overexpressing SET isoform 2 or GFP were infected with HIVpp-luc and total DNA was extracted at various times up to 48 h p.i. (Figure 3A, B). Interestingly, while SET overexpression in MT4 cells had little effect on the magnitude or kinetics of HIV-1 RT (Figure 3A), DNA integration was significantly reduced by ∼40–50% at 24–48 h p.i. (Figure 3B). Levels of unintegrated 2-LTR circles were analyzed to determine whether the integration block occurred pre- or post-nuclear import. SET overexpression was found to dramatically increase the levels of unintegrated 2-LTR circles, suggesting that HIV-1 nuclear import is unaffected and that SET blocks post-nuclear entry DNA integration (Supplementary Figure S3A). Moreover, because HIV-1 infectivity, as assessed by HIVpp-luc-mediated luciferase expression, was also reduced by ∼40% upon SET overexpression (Figure 3C), we considered it unlikely that the post-integration steps of HIV-1 infection were affected by SET overexpression. To test this, we performed a side-by-side comparison of HIV-1 DNA integration and HIVpp-luc infectivity in MT4 cells with CRISPR/Cas9-mediated SET overexpression. We found that HIV-1 DNA integration was decreased by ∼70% and ∼50% in cells expressing two different SET-targeting sgRNAs (Figure 3D, Supplementary Figure S3B), with parallel and comparable reductions in infectivity (Figure 3E). To directly investigate whether HIV-1 transcription was affected by SET, we employed an HIV-1 promoter reporter assay in control and SET KD cells, which revealed statistically insignificant differences among the various cell types (Supplementary Figure S3E). These data suggest that SET blocks HIV-1 infection by interfering with the DNA integration step.

Our experiments thus far have modulated SET expression using sgRNAs and shRNAs delivered by lentiviral vectors, which have backbones based on HIV-1. To rule out any influence on our findings by integration of the lentiviral vectors, we expressed shNC or shSET in MT4 cells using Moloney murine leukemia virus (MLV)-based vectors, which are not detected by the Alu-qPCR assay for HIV-1 DNA integration due to sequence differences between MLV and HIV-1. We confirmed that MLV-mediated shSET expression effectively reduced SET protein and mRNA expression in these cells (Supplementary Figure 3C, D), and then examined HIVpp-luc infectivity and DNA integration. Notably, both integration and infectivity were enhanced by about 1.5–2-fold in SET KD cells compared with control cells (Figure 3F, G). Thus, the involvement of SET in HIV-1 integration was not affected by the viral vector employed to modulate SET expression.

### SET does not alter HIV-1 integration site preference and restricts HIV-1 integration by inhibiting histone acetylation

As noted above, SET inhibits acetylation of histones, including H3 and H4, by inhibiting the function of the HATs EP300 and GNC5, and has been proposed to act as part of a HAT inhibitor complex (INHAT) with ANP32A (27,28). Integration of HIV-1 in the human genome is nonrandom, with specific preferences for active genes (33) and associated chromatin, including regions enriched for activating epigenetic marks such as acetylated H3 and H4 (34). Therefore, we hypothesized that SET may inhibit HIV-1 integration *via* its activity as part of the INHAT complex. We first determined whether ANP32A and EP300 are also involved in HIV-1 integration by examining MT4 cells with shRNA-mediated ANP32A or EP300 KD. Indeed, HIVpp-luc DNA integration and infection were amplified by ANP32A KD, a phenotype similar to that seen in SET-depleted cells (Figure 3H, I and Supplementary Figure 3F, G), and conversely, were inhibited by EP300 KD (Figure 3H and I, Supplementary Figure S3H). These results are consistent with a positive role for EP300 in promoting HIV-1 integration and an inhibitory role for SET/ANP32A in preventing HIV-1 integration (29).

We envisioned a possible scenario for SET-mediated inhibition of HIV-1 integration where SET expression would decrease accessibility to favorable integration sites via downregulation of activating epigenetic marks. We further envisaged that sites of HIV-1 integration in the human genome would alter under this scenario. To determine whether SET alters HIV-1 integration site preferences, we performed genomic integration site sequencing in SET depleted PBMCs and MT4 cells. PBMCs isolated from the blood of healthy donors and activated by antibodies to CD3 and CD28 for 3 days were electroporated with ribonucleoprotein complexes (Cas9-RNPs) that consisted of the Cas9 nuclease bound to a SET CRISPR RNA (crRNA-SET) and the trans-activating crRNA (tracrRNA) (Figure 4A). After 6 days in culture to allow for depletion of the targeted gene, the cells were infected for 3 days to enable efficient integration of single cycle HIV-1 (HIVpp-luc). In the two donors, SET Cas9-RNP knocked out SET expression (Figure 4B and C). Consistent with our results in Figure 3, depletion of SET increased integrated proviral HIV-1 DNA by ∼1.5-fold (Figure 4B and C). Integration sites were also determined in MT4 cells knocked down for SET expression using MLV-based shRNA. Cellular DNA from infected control or SET depleted cells was extracted, and HIV-1 integration sites were amplified by ligation-mediated PCR and sequenced on the Illumina platform (12–17). Previous integration site analyses showed that HIV-1 prefers to integrate into active chromatin including speckle-associated genomic domains (SPADs) and disfavors integration into transcriptionally-repressed chromatin including lamina-associated domains (LADs) (12,14). Integration sites were initially mapped to RefSeq genes, surrounding gene density, SPADs, and LADs (Supplementary Table S3). In MT4 cells, around 75% of integrations were within genes (Figure 4D); the average

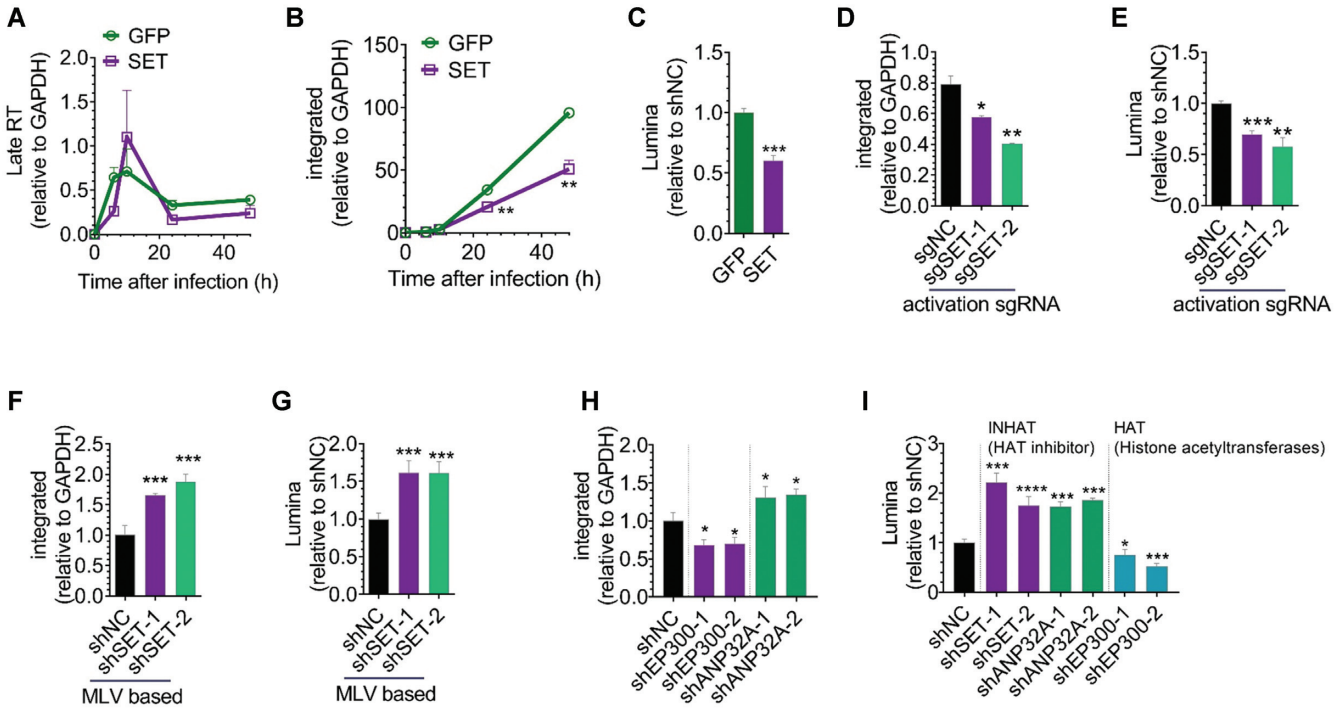


**Figure 3.** SET represses HIV-1 integration. (**A–C**) SET overexpression represses the integration of HIV-1 DNA into cellular DNA. Late RT products (A), HIV integration (B) and infectivity (C) were measured by qPCR analysis, Alu-qPCR analysis, and luciferase assay, respectively, of MT4 cells overexpressing GFP or SET and infected with HIVpp-luc (MOI = 0.2) for 6, 10, 24 or 48 h. Mean ± SD of $n = 3$. $^{**}P < 0.01$, $^{***}P < 0.001$ by Student's $t$ test. (**D, E**) SAM mediated activation of SET represses HIV-1 integration. HIV-1 Integration (D) and infectivity (E) were measured by Alu-qPCR and luciferase assay, respectively, of MT4 cells expressing sgNC or two sgSETs and infected with HIVpp-luc for 2 days. Mean ± SD of $n = 3$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by Student's $t$ test. (**F, G**) KD of SET using MLV based shRNAs represses the integration of HIV-1 DNA into cellular DNA. HIV-1 Integration (F) and infectivity (G) were measured by Alu-qPCR and luciferase assay, respectively, of MT4 cells subjected to MLV-mediated transduction of shNC or shSETs and infected with HIVpp-luc (MOI = 0.2) for 2 days. Mean ± SD of $n = 3$. $^{***}P < 0.001$ by Student's $t$ test. (**H, I**) EP300 promotes and ANP32A inhibits HIV-1 integration. Lentivirus expressing control shRNA or specific shRNAs against SET, the histone acetylase (HAT) EP300, or the HAT inhibitor (INHAT) ANP32A was transduced into MT4 cells. KD cells were selected with puromycin for 4 days, and then infected with HIVpp-Luc for 2 days. Integration by Alu qPCR (H) and HIV-Luc infectivity (I) were quantified. Mean ± SD of $n = 3$. $^{*}P < 0.05$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by Student's $t$ test. Integrated HIV DNA in (B), (D), (F) and (H) were normalized to genomic GAPDH.

gene density surrounding integration sites was about 10.6 genes/Mb (Figure 4E); and about 7% and 44% of integrations occurred within SPADs and LADs, respectively (Figure 4F, G). In primary PBMCs, we observed that genes and gene-enriched regions were also favorable targets, with ∼86% of integrations within genes (Figure 4D) and the average gene density surrounding integration sites increased to about 22 genes/Mb (Figure 4E). Moreover, about 31% and 15% of integrations occurred within SPADs and LADs, respectively (Figure 4F, G, Supplementary Table S3). While integration frequencies in transcription units in PBMCs and MT4 cells were similar to prior reports (35), MT4 cells supported uncharacteristically low levels of integration into gene dense regions and SPADs (12) and, concordantly, LAD-proximal integration was uncharacteristically high in MT4 cells (14). By contrast, the integration site profile of HIV-1 in PBMCs was consistent with prior reports (35). Although it is unclear why MT4 cells support significantly less integration in active regions of chromatin that are typically targeted by HIV-1, their unusual behavior afforded a unique opportunity to test the role of SET in integration in two different cell types that presented rather different basal integration targeting profiles.

SET depletion failed to appreciably affect integration targeting frequencies in MT4 cells and PBMCs (Figure 4D–G). Based on these findings, we expanded our analyses to other features known to be associated with HIV-1 integration site targeting, including histone modifications associated with active chromatin H3K4me1 (Supplementary Figure S4A), H3K36me3 (Supplementary Figure S4B), and H3K27Ac (Supplementary Figure S4C). We additionally analyzed spatial position inference of the nuclear genome or SPIN states, which are 10 regions of the human genome segmented based on radial position and transcriptional activity (36). HIV-1 integration greatly prefers interior SPIN states Speckles and Interior Active 1, and disfavors nuclear membrane proximal regions such as Lamina (19,35). Similar to our analyses of SPAD and LAD-proximal integration targeting, HIV-1 in MT4 cells targeted regions associated with activating epigenetic marks less so than it did in PBMCs (Supplementary Figure S4A–C), and only marginally favored the Speckle SPIN state and uncharacteristically favored Near Lamina 2 chromatin (Supplementary Figure S4D, E). Regardless, these integration targeting metrics remained largely unaffected by SET depletion in both MT4 cells and PBMCs. Based on these findings, we concluded

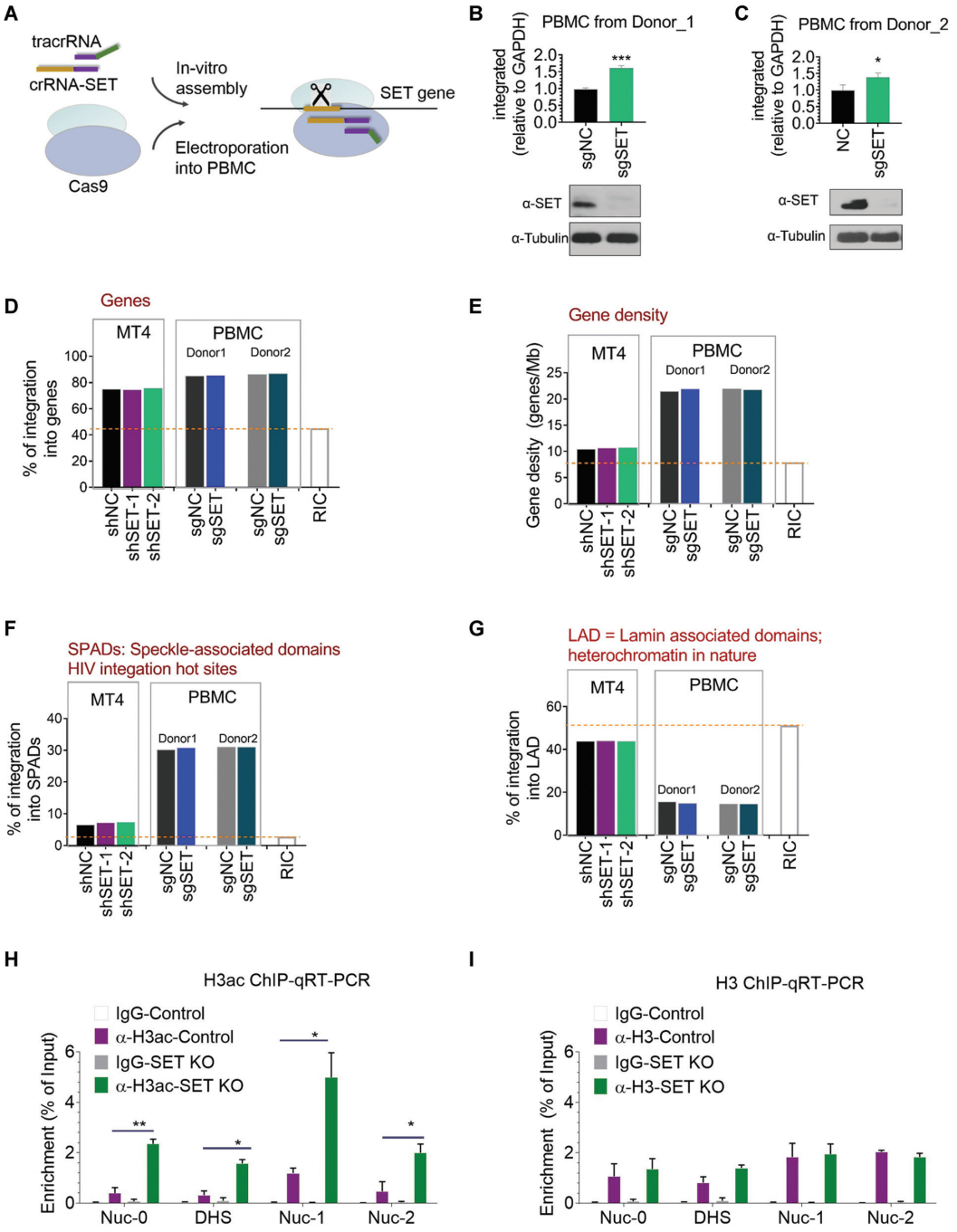


**Figure 4.** SET represses HIV-1 integration but does not alter integration sites. (A–C) SET knockout in PBMCs enhances the integration of HIV-1. (**A**) Illustration of *in vitro* ribonucleoprotein complexes (Cas9-RNPs) formation and delivery into PBMCs. PBMCs isolated from the blood of healthy donors were activated by using CD3 and CD28 antibodies. Then, cells were electroporated with Cas9-RNPs, which consisted of the Cas9 nuclease bound to a SET CRISPR RNA (crRNA-SET) and the trans-activating crRNA (tracRNA). (**B**, **C**) PBMC of two healthy donors were delivered with the SET Cas9-RNPs. After 6 days in culture to allow for depletion of the targeted gene, the cells were infected with HIVpp-luc (MOI = 2) for 3 days to enable efficient integration of the single cycle reporter virus. Integrated HIV-1 DNA was quantified by Alu qPCR, and SET expression detected using Western blot. Integrated HIV-1 DNA was normalized to genomic GAPDH. Mean ± SD of $n = 3$. *$P < 0.05$, ***$P < 0.001$ by Student's $t$ test. Western blot data are representative of at least two independent experiments. Tubulin was used as a loading control. (**D–G**) SET depletion does not appreciably alter sites of HIV-1 integration. HIV-1 integration site sequencing was performed in HIVpp-luc infected (MOI = 0.2) MT4 cells depleted for SET by the MLV KD system. Meanwhile, two SET depleted sets of PBMCs prepared in panels (B), (C) were subjected to integration site sequencing. Cellular DNA from infected control or SET depleted cells were extracted. Sites of HIV-1 integration, which were amplified by ligation-mediated PCR, were sequenced using Illumina. Integration site analysis of HIV-1 DNA integration sites mapped with respect to RefSeq genes (D), surrounding gene density (E), SPADs (F) and LADs (G). RIC, random integration control. (**H, I**) SET inhibits H3ac-associated HIV-1 DNA. Acetylated H3 ChIP (H3ac-ChIP, H) and total H3 ChIP (I) assays of Jurkat cells expressing sgNC (control) or sgSET (SET KO) and infected with HIV-$1_{LAI}$ (MOI = 1) for 2 days. Immunoprecipitates were subjected to qPCR with primers specific for four regions of HIV-1 DNA: Nucleosome (Nuc)-0, Nuc-1, Nuc-2 and DHS (DNase I highly sensitive). Mean ± SD of $n = 2$. *$P < 0.05$, **$P < 0.01$ by Student's $t$ test.

that SET plays a significant role to regulate HIV-1 integration without appreciably affecting integration site targeting.

To further investigate the impact of SET expression on HIV-1-specific as well as genome-wide histone acetylation, we next measured histone acetylation modifications by ChIP. H3ac-ChIP and H3-ChIP were performed in HIV-1-infected SET knockout cells at several time points that spanned from before detectable integration (6 h) through peak values at 48 h p.i.. Jurkat cells harboring sgNC or sgSET were infected with HIV-$1_{LAI}$ and levels of total H3 and acetylated H3 (H3ac) were detected by ChIP, followed by qPCR specific for nucleosomes (Nuc) 0, Nuc-1 and Nuc-2, and NHS (DNase I hypersensitive sites) associating HIV-1 DNA (37,38). SET knockout upregulated levels of H3ac-associated HIV-1 DNA both at 6 h p.i. (Supplementary Figure S4F) and at 48 h p.i. (Figure 4H). By contrast, total levels of H3-associated HIV-1 DNA were unchanged under these conditions (Supplementary Figure S4G, Figure 4I). These data indicated that H3ac-associated levels of unintegrated and integrated HIV-1 DNA were increased by SET knockout. We next asked if this change of acetylation was specific to HIV-1 DNA or due to global increases of H3ac levels. We found that overall histone acetylation, including levels associated with host genes such as GAPDH (Supplementary Figure S4F, G), were upregulated in SET knockout cells, suggesting that the upregulation of HIV-1-DNA-associated histone acetylation was not specific to the virus. Taken together, SET inhibits global H3ac levels and restricts HIV-1 integration while not specifically altering HIV-1 integration site preferences.

### SET expression is downregulated by HIV-1 infection *in vivo* and is restored by inhibition of granzyme A in infected cells

Finally, we examined whether SET expression might correlate with HIV-1 pathogenesis in people lived with HIV/AIDS (PLWH). We first investigated SET mRNA expression levels in PBMCs that were infected *ex vivo* with HIV-$1_{LAI}$ at different time points, which revealed that robust HIV-1 infection did not affect SET mRNA levels (Figure 5A, Supplementary Figure S5A).

We next analyzed SET mRNA expression in single-cell RNA sequencing data obtained from PBMC from two healthy subjects (15121 cells) or six HIV-infected donors (28,610 cells) (39). Our analysis showed that SET mRNA expression was not changed in CD4+ T cells of HIV-1 infected individuals as compared to healthy donors (Supplementary Figure S5B). Additionally, in 24 HIV-1-infected samples, mRNA expressions were variable among different donors, and the average-line of expression was not significantly altered when compared to the average-line of the 10 healthy donors (Figure 5B).

SET protein levels, however, were significantly decreased among 16 HIV-1-infected donors (Figure 5C, Supplementary Table S4). By contrast, SET protein levels remained constant in MT4 and H9 T cells infected with HIV-$1_{LAI}$ (Supplementary Figure S5C, D). To determine if the reduction in SET protein observed in HIV patients was cell-type dependent, we infected primary PBMC, which revealed a significant reduction of SET at 3–4 days p.i. (Figure 5D, Supplementary Figure S5E, F). We considered that SET degradation under these conditions was likely to be indirectly mediated by the virus. Previous studies suggested that Granzyme A (GzmA) cleaved SET to disrupt its nucleosome assembly activity (40). We accordingly pretreated PBMC with either GzmA inhibitor, proteasome inhibitor (MG132), or a lysosome inhibitor (bortezomib). The level of SET protein was rescued by GzmA inhibitor across donors (Figure 5E, F). In contrast, neither MG132 nor bortezomib treatment rescued the decrease in SET protein instilled by HIV-1 infection (Supplementary Figure S5G). Additionally, GZMA mRNA and protein were both upregulated in CD8+ T and CD4+ T cells upon HIV-1 infection (Supplementary Figure S6A–C). These data suggest that GzmA, activated by HIV-1 infection, mediated SET degradation in PBMC.

In summary, we have performed a genome-wide functional screen in CD4+ T cells to identify cellular restriction factors of HIV-1. We revealed that one such factor, SET, restricted H3ac, the consequences of which inhibited HIV-1 integration without significantly affecting integration site preferences. SET protein expression correlated with HIV-1 pathogenesis. During HIV-1 infection, GZMA was activated, which mediated the degradation of SET (Supplementary Figure S7).

## DISCUSSION

As a retrovirus, the life cycle of HIV-1 involves reverse transcription of genomic RNA into DNA followed by integration into the host genome to form a provirus. Understanding how HIV-1 integration is established and regulated by host proteins may inform novel therapeutic strategies to prevent virus infection and potentially inform HIV cure strategies. For example, treatment of cells with LEDGINs, a class of small molecules that can inhibit the interaction between integrase and cellular LEDGF/p75, can impinge transcriptional reactivation of resultant proviruses (41). Here, we employed a high-complexity genome-wide CRISPR/Cas9 transcriptional activation screen in human CD4 + T cells to identify host factors that restrict HIV-1 infection. We identified a novel series of host proteins that block viral integration and additionally revealed that the SET/ANP32A INHAT complex plays a role in preventing HIV-1 integration by regulating histone acetylation.

The novel inhibitors of HIV-1 infection identified here included KCTD1, KIAA1586, ORAI3, ATP1B1, and SET. KCTD1 is a soluble non-channel member of the potassium channel tetramerization domain family and functions as a Cullin3-dependent E3 ligase. KIAA1586 is an E3 SUMO-protein ligase that facilitates UBE2I/UBC9-mediated conjugation of the small ubiquitin-like modifier SUMO2 to target proteins (42). ORAI3 encodes a subunit of $Ca^{2+}$ release-activated Ca2+-like (CRAC-like) channel and mediates $Ca^{2+}$ influx. ATP1B1 is a member of the family of $Na^+/K^+$ and $H^+/K^+$ ATPases beta chain proteins and is responsible for establishing and maintaining electrochemical $Na^+$ and $K^+$ gradients across the plasma membrane. Finally, we focused our study mainly on the histone acetylation inhibitor SET, which we considered highly likely to play a regulatory role in HIV-1 infection (27,28).

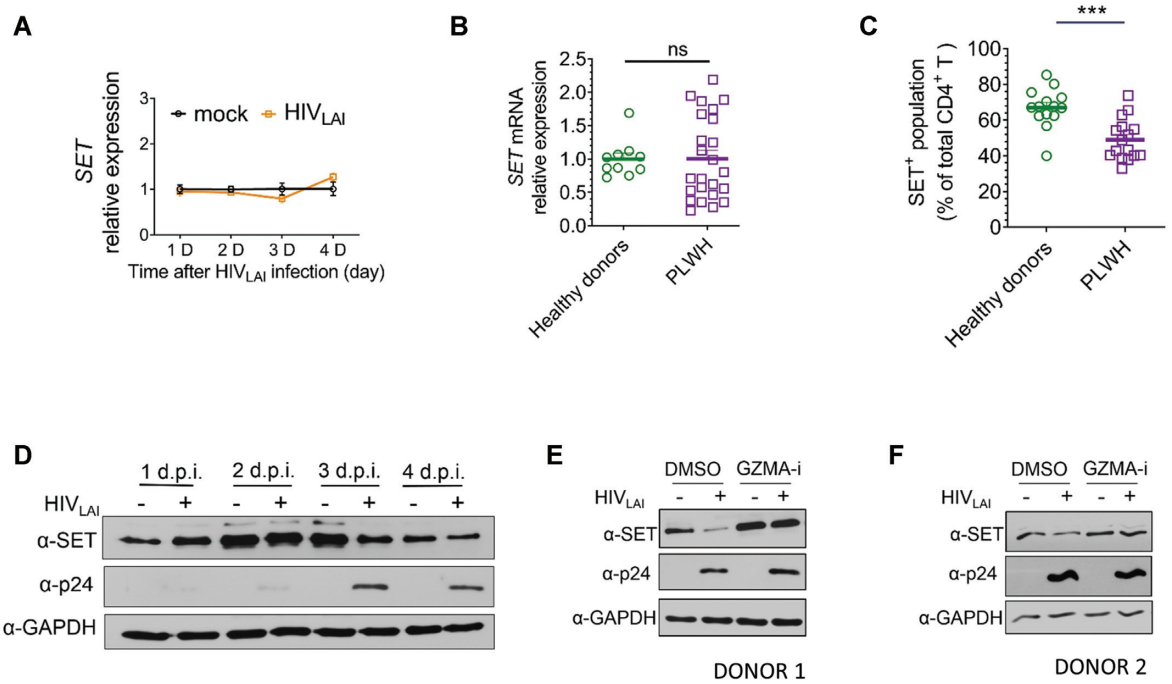


**Figure 5.** Downregulation of SET in PBMCs by HIV-1 infection is partially reversed by inhibition of granzyme A. (**A**) *SET* mRNA levels are not changed by HIV infection. RT-qPCR analysis of SET mRNA in activated PBMCs from healthy donors infected with HIV-$1_{LAI}$ (MOI 0.1) for 4 days. Mean ± SD, $n = 3$. *SET* mRNA was normalized to *GAPDH* expression. (**B**) *SET* mRNA levels are not changed in people living with HIV (PLWH). RT-qPCR analysis of *SET* mRNA in PBMCs isolated from 24 PLWH and 10 healthy donors. *SET* mRNA was normalized to *18S* RNA expression. (**C**) Significant decreases in SET protein levels in PLWH. Flow cytometric analysis of peripheral blood CD4+ T cells from 16 PLWH and 14 healthy donors after staining of intracellular SET protein. Horizontal bars indicate the mean. ***$P < 0.001$, by Student's $t$ test. (**D**) HIV-1 infection reduces SET protein expression. Western blot analysis of SET and p24 proteins in activated PBMCs from healthy donors mock-infected (−) or infected (+) with HIV-$1_{LAI}$ (MOI 1) for 4 days. (**E, F**) Granzyme A inhibitor counteracts the decrease of SET protein expression upon HIV-1 infection. Western blot analysis of SET and p24 proteins in activated PBMCs from healthy donors mock-infected (−) or infected (+) with HIV-$1_{LAI}$ (MOI 1) for 3 days. Cells were incubated with DMSO or 50 μM nafamostat mesylate, a granzyme A inhibitor (GZMA-i), 1 day before lysis. Western blot data in D-F are representative of at least two independent experiments. GAPDH was used as a loading control.

Our results imply that SET inhibits HIV-1 integration by reducing histone acetylation of integrase and/or HIV-1 DNA. SET has been shown to regulate the HATs EP300 and GCN5 (27,28), and our finding that EP300 KD in MT4 cells inhibited both HIV-1 DNA integration and replication is consistent with previous findings (29,30). These authors demonstrated that EP300 and GCN5 enhanced HIV-1 integration by catalyzing acetylation of HIV-1 integrase. A later study from the same group showed that Tripartite motif-containing 28 (TRIM28/KAP1) binds to acetylated HIV-1 integrase and recruited HDAC1, followed by deacetylation of HIV-1 integrase (43). However, it remains unclear whether acetylation of HIV-1 integrase directly affects its enzymatic activity, and evidence currently exists both for and against this association. Cereseto *et al.* showed that point mutations in the EP300 acetylation sites of HIV-1 integrase impaired its integration activity (29), although some concern has been raised that the position of a C-terminal HA tag close to the integrase acetylation sites may have influenced the results, particularly because Muesing and colleagues reported that similar point mutations in untagged integrase failed to significantly effect HIV-1 integration (32). Furthermore, a more recent study showed that mutation of four acetylated lysines in the C-terminal domain of HIV-1 integrase affected proviral transcription but not the preceding integration step (31). Based on these studies, it seems unlikely that acetylation of HIV-1 integrase directly affects DNA integration. Further work will be necessary to uncover the mechanism(s) by which EP300- and GCN5-mediated acetylation of integrase promotes HIV-1 replication. Of note, our ChIP studies demonstrated that SET inhibits acetylation of histones associated with HIV-1 DNA. We cannot exclude the possibility that SET may additionally act by directly inhibiting integrase acetylation, but the preceding discussion suggests that this would be unlikely to affect integration.

SET is known to form two complexes: a nuclear complex composed of SET and ANP32A and an endoplasmic reticulum (ER)-associated complex composed of SET, the base excision repair endonuclease APE1, the exonuclease TREX1, and the endonuclease NM23-H1. In the present study, we showed that shRNA-mediated KD of ANP32A phenocopied the effects of SET KD on HIV-1 integration and infection, suggesting that the SET/ANP32A complex mediates the observed effects of SET, although we cannot rule out a role for the second SET complex. Previously, the ER-associated complex was implicated in HIV-1 integration by suppressing suicidal autointegration (44).

In summary, our genome-wide CRISPR/Cas9 transcriptional activation screen proved to be an effective method

for identifying host factors that control HIV infection of human CD4+ T cells and enabled the discovery of several novel proteins that prevent HIV-1 infection. Our results may open up new avenues of research into the development of therapeutic agents that modulate host factors involved in HIV-1 life cycle.

## DATA AVAILABILITY

The CRISPR sequencing data reported in this paper is uploaded in NCBI GEO (GSE185343). The integration sequencing data is uploaded into BioProject (PRJNA768457). The Flow-Cytometry raw data and analysis files are available via FlowRepository (FR-FCM-Z4L4).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr Kristen Jepsen of the Institute of Genomic Medicine at UCSD for help with the sequencing and members of the Rana lab for helpful discussions and advice. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 system that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929). We thank Dr. Jonathan Karn for the microglia CHME cell line, and the NIH AIDS Reagent Program for the $HIV_{LAI}$, $HIV_{NL4-3}$-TK and HIVpp-luc constructs.

## FUNDING

National Institutes of Health [CA177322, DA039562, DA046171, AI052014, AI125103, in part]. Funding for open access charge: NIH.

*Conflict of interest statement.* A.N.E. has consulted for ViiV Healthcare, Co. on unrelated work. T.M.R. is a founder of ViRx Pharmaceuticals and has an equity interest in the company. The terms of this arrangement have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies. All other authors have no competing interests to declare.

## REFERENCES

1. Goff,S.P. (2007) Host factors exploited by retroviruses. *Nat. Rev. Microbiol.*, **5**, 253–263.
2. Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J.R. and Bushman,F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
3. Singh,P.K., Plumb,M.R., Ferris,A.L., Iben,J.R., Wu,X., Fadel,H.J., Luke,B.T., Esnault,C., Poeschla,E.M., Hughes,S.H. *et al.* (2015) LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev.*, **29**, 2287–2297.
4. Cosnefroy,O., Tocco,A., Lesbats,P., Thierry,S., Calmels,C., Wiktorowicz,T., Reigadas,S., Kwon,Y., De Cian,A., Desfarges,S. *et al.* (2012) Stimulation of the human RAD51 nucleofilament restricts HIV-1 integration in vitro and in infected cells. *J. Virol.*, **86**, 513–526.
5. Desfarges,S., San Filippo,J., Fournier,M., Calmels,C., Caumont-Sarcos,A., Litvak,S., Sung,P. and Parissi,V. (2006) Chromosomal integration of LTR-flanked DNA in yeast expressing HIV-1 integrase: down regulation by RAD51. *Nucleic Acids Res.*, **34**, 6215–6224.
6. Zhu,Y., Wang,G.Z., Cingoz,O. and Goff,S.P. (2018) NP220 mediates silencing of unintegrated retroviral DNA. *Nature*, **564**, 278–282.
7. Park,R.J., Wang,T., Koundakjian,D., Hultquist,J.F., Lamothe-Molina,P., Monel,B., Schumann,K., Yu,H., Krupzcak,K.M., Garcia-Beltran,W. *et al.* (2017) A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat. Genet.*, **49**, 193–203.
8. Huang,H.C., Kong,W.L., Jean,M.M., Fiches,G., Zhou,D.W., Hayashi,T., Que,J.W., Santoso,N. and Zhu,J. (2019) A CRISPR/Cas9 screen identifies the histone demethylase MINA53 as a novel HIV-1 latency-promoting gene (LPG). *Nucleic Acids Res.*, **47**, 7333–7347.
9. Konermann,S., Brigham,M.D., Trevino,A.E., Joung,J., Abudayyeh,O.O., Barcena,C., Hsu,P.D., Habib,N., Gootenberg,J.S., Nishimasu,H. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583–588.
10. Wang,S., Zhang,Q., Tiwari,S.K., Lichinchi,G., Yau,E.H., Hui,H., Li,W., Furnari,F. and Rana,T.M. (2020) Integrin alphavbeta5 internalizes zika virus during neural stem cells infection and provides a promising target for antiviral therapy. *Cell Rep.*, **30**, 969–983.
11. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
12. Francis,A.C., Marin,M., Singh,P.K., Achuthan,V., Prellberg,M.J., Palermino-Rowland,K., Lan,S., Tedbury,P.R., Sarafianos,S.G., Engelman,A.N. *et al.* (2020) HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat. Commun.*, **11**, 3505.
13. Chen,Y., Zhang,Y. and Wang,Y. (2018) Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.*, **217**, 4025–4048.
14. Achuthan,V., Perreira,J.M., Sowd,G.A., Puray-Chavez,M., McDougall,W.M., Paulucci-Holthauzen,A., Wu,X., Fadel,H.J., Poeschla,E.M., Multani,A.S. *et al.* (2018) Capsid-CPSF6 interaction licenses nuclear HIV-1 trafficking to sites of viral DNA integration. *Cell Host Microbe*, **24**, 392–404.
15. Sowd,G.A., Serrao,E., Wang,H., Wang,W., Fadel,H.J., Poeschla,E.M. and Engelman,A.N. (2016) A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1054–E1063.
16. Serrao,E., Cherepanov,P. and Engelman,A.N. (2016) Amplification, Next-generation sequencing, and genomic DNA mapping of retroviral integration sites. *J. Vis. Exp.*, https://doi.org/10.3791/53840.
17. Matreyek,K.A., Wang,W., Serrao,E., Singh,P.K., Levin,H.L. and Engelman,A. (2014) Host and viral determinants for MxB restriction of HIV-1 infection. *Retrovirology*, **11**, 90.
18. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
19. Singh,P.K., Bedwell,G.J. and Engelman,A.N. (2022) Spatial and genomic correlates of HIV-1 integration site targeting. *Cells*, **11**, 655.
20. Meuleman,W., Peric-Hupkes,D., Kind,J., Beaudry,J.B., Pagie,L., Kellis,M., Reinders,M., Wessels,L. and van Steensel,B. (2013) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.*, **23**, 270–280.
21. Butler,S.L., Hansen,M.S. and Bushman,F.D. (2001) A quantitative assay for HIV DNA integration in vivo. *Nat. Med.*, **7**, 631–634.
22. Gyuris,A., Vajda,G. and Foldes,I. (1992) Establishment of an MT4 cell line persistently producing infective HIV-1 particles. *Acta Microbiol. Hung.*, **39**, 271–279.
23. Smith,S.M., Markham,R.B. and Jeang,K.T. (1996) Conditional reduction of human immunodeficiency virus type 1 replication by a gain-of-herpes simplex virus 1 thymidine kinase function. *Proc. Nat. Acad. Sci. U.S.A.*, **93**, 7955–7960.
24. Hrecka,K., Hao,C., Gierszewska,M., Swanson,S.K., Kesik-Brodacka,M., Srivastava,S., Florens,L., Washburn,M.P. and Skowronski,J. (2011) Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, **474**, 658–661.
25. Laguette,N., Sobhian,B., Casartelli,N., Ringeard,M., Chable-Bessia,C., Segeral,E., Yatim,A., Emiliani,S., Schwartz,O. and Benkirane,M. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by vpx. *Nature*, **474**, 654–657.

26. Freed,E.O. (2015) HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.*, **13**, 484–496.
27. Saavedra,F., Rivera,C., Rivas,E., Merino,P., Garrido,D., Hernandez,S., Forne,I., Vassias,I., Gurard-Levin,Z.A., Alfaro,I.E. *et al.* (2017) PP32 and SET/TAF-Ibeta proteins regulate the acetylation of newly synthesized histone h4. *Nucleic Acids Res.*, **45**, 11700–11710.
28. Seo,S.B., McNamara,P., Heo,S., Turner,A., Lane,W.S. and Chakravarti,D. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell*, **104**, 119–130.
29. Cereseto,A., Manganaro,L., Gutierrez,M.I., Terreni,M., Fittipaldi,A., Lusic,M., Marcello,A. and Giacca,M. (2005) Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J.*, **24**, 3070–3081.
30. Terreni,M., Valentini,P., Liverani,V., Gutierrez,M.I., Di Primio,C., Di Fenza,A., Tozzini,V., Allouch,A., Albanese,A., Giacca,M. *et al.* (2010) GCN5-dependent acetylation of HIV-1 integrase enhances viral integration. *Retrovirology*, **7**, 18.
31. Winans,S. and Goff,S.P. (2020) Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. *PLoS Pathog.*, **16**, e1009147.
32. Topper,M., Luo,Y., Zhadina,M., Mohammed,K., Smith,L. and Muesing,M.A. (2007) Posttranslational acetylation of the human immunodeficiency virus type 1 integrase carboxyl-terminal domain is dispensable for viral replication. *J. Virol.*, **81**, 3012–3017.
33. Schroder,A.R., Shinn,P., Chen,H., Berry,C., Ecker,J.R. and Bushman,F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.
34. Wang,G.P., Ciuffi,A., Leipzig,J., Berry,C.C. and Bushman,F.D. (2007) HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.*, **17**, 1186–1194.
35. Bedwell,G.J., Jang,S., Li,W., Singh,P.K. and Engelman,A.N. (2021) rigrag: high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. *Nucleic Acids Res.*, **49**, 7330–7346.
36. Wang,Y., Zhang,Y., Zhang,R., van Schaik,T., Zhang,L., Sasaki,T., Peric-Hupkes,D., Chen,Y., Gilbert,D.M., van Steensel,B. *et al.* (2021) SPIN reveals genome-wide landscape of nuclear compartmentalization. *Genome Biol.*, **22**, 36.
37. Rafati,H., Parra,M., Hakre,S., Moshkin,Y., Verdin,E. and Mahmoudi,T. (2011) Repressive LTR nucleosome positioning by the BAF complex is required for HIV latency. *PLoS Biol.*, **9**, e1001206.
38. Verdin,E., Paras,P. Jr and Van Lint,C. (1993) Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J.*, **12**, 3249–3259.
39. Wang,S., Zhang,Q., Hui,H., Agrawal,K., Karris,M.A.Y. and Rana,T.M. (2020) An atlas of immune cell exhaustion in HIV-infected individuals revealed by single-cell transcriptomics. *Emerg. Microb. Infect.*, **9**, 2333–2347.
40. Beresford,P.J., Zhang,D., Oh,D.Y., Fan,Z., Greer,E.L., Russo,M.L., Jaju,M. and Lieberman,J. (2001) Granzyme a activates an endoplasmic reticulum-associated caspase-independent nuclease to induce single-stranded DNA nicks. *J. Biol. Chem.*, **276**, 43285–43293.
41. Debyser,Z., Bruggemans,A., Van Belle,S., Janssens,J. and Christ,F. (2021) LEDGINs, inhibitors of the interaction between HIV-1 integrase and LEDGF/p75, are potent antivirals with a potential to cure HIV infection. *Adv. Exp. Med. Biol.*, **1322**, 97–114.
42. Eisenhardt,N., Chaugule,V.K., Koidl,S., Droescher,M., Dogan,E., Rettich,J., Sutinen,P., Imanishi,S.Y., Hofmann,K., Palvimo,J.J. *et al.* (2015) A new vertebrate SUMO enzyme family reveals insights into SUMO-chain assembly. *Nat. Struct. Mol. Biol.*, **22**, 959–967.
43. Allouch,A., Di Primio,C., Alpi,E., Lusic,M., Arosio,D., Giacca,M. and Cereseto,A. (2011) The TRIM family protein KAP1 inhibits HIV-1 integration. *Cell Host Microbe*, **9**, 484–495.
44. Yan,N., Cherepanov,P., Daigle,J.E., Engelman,A. and Lieberman,J. (2009) The SET complex acts as a barrier to autointegration of HIV-1. *PLoS Pathog.*, **5**, e1000327.