![PNAS www.pnas.org]

Supplementary Information for

Species-specific KRAB-ZFPs function as repressors of retroviruses by targeting PBS regions

Bo Yang[a], Lu Fang[b], Qianqian Gao[a], Ce Xu[a], Junqin Xu[a], Zhen-Xia Chen[c,d], Yixuan Wang[b,1], Peng Yang[a,1]

[a]Shanghai Key Laboratory of Maternal and Fetal Medicine, Clinical and Translational Research Center of Shanghai First Maternity & Infant Hospital, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China.

[b]Translational Medical Center for Stem Cell Therapy, Institute for Regenerative Medicine of Shanghai East Hospital, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China.

[c]Hubei Hongshan Laboratory, College of Biomedicine and Health, Huazhong Agricultural University, Wuhan 430070, China

[d]Hubei Key Laboratory of Agricultural Bioinformatics, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China.

[1]Correspondence:

Peng Yang (peng.yang@tongji.edu.cn) or Yixuan Wang (wangyixuan@tongji.edu.cn)

This PDF file includes:

Materials and Methods

Fig. S1. Generation of *Zfp961-GFP*^{*flox/flox*} mESCs cell lines.

Fig. S2. ChIP-seq analysis shows ZFP961 as a potential ERVK repressor.

Fig. S3. ZFP961 targets PBS-Lys sites for transcriptional repression.

Fig. S4. Deletion of *Zfp961* leads to increased expression of genes and repeats elements.

Fig. S5. Schematic overview of HIV-1 pseudo-viral infection repression assay.

Fig. S6. Human ZNF417 and ZNF587 target PBS-Lys for repression.

Fig. S7. Human ZNF417 and ZNF587 repress human ERVK.

Fig. S8. Strategy and validation of *ZNF417/ZNF587* KO cell lines.

Fig. S9. ZNF417 and ZNF587 restrict HIV-1 infectivity in human primary CD4+ T cells.

Fig. S10. ZFP961, ZNF417 and ZNF587 restrict HIV-1 transcription and integration.

Dataset S1. RNA-seq of *Zfp961-GFP*^{*flox/flox*} mESCs.

Dataset S2. ZNF ChIP peaks overlap with PBS sites.

Dataset S3. RNA-seq of *ZNF417 ZNF587* dKO and control K562 cells.

Dataset S4. HIV-1 integration sites in *ZNF417 ZNF587* dKO and control HEK293T cells.

Dataset S5. Classification of ERVs by PBS type.

Dataset S6. PCR primers and promoters mentioned in this paper.

References for SI

Materials and Methods

Mouse lines and cell culture

Specific-pathogen-free (SPF) mice were housed in the animal facility at Tongji University, Shanghai, China. All animal maintenance and experimental procedures were performed according to the Tongji University Guide for the use of laboratory animals.

Zfp961-GFP^flox/flox; Rosa26-CreERT2 mouse ESCs were cultured in ESC medium containing DMEM (Gibco), 15% fetal bovine serum (FBS; Hyclone), 10 mM NEAA (Millipore), 100 µM β-mercaptoethanol (Sigma), 2 mM L-glutamine (Millipore), and 10 ng/mL LIF (Millipore), 2i (3 µM CHIR-99021, 1 µM PD0325901, Selleck). HEK293T cells were cultured in DMEM High Glucose (Gibco) containing 10% FBS (Gibco) with 2 mM L-glutamine (Millipore). Jurkat E6.1 and K562 cells were cultured in RPMI1640 (Gibco) containing 10% FBS (Gibco) and 2 mM L-glutamine (Millipore).To induce *Zfp961* KO in ESCs, *Zfp961*-GFP^flox/flox^; Rosa26-CreERT2 ESCs were treated with 4-OHT for 48 h and then cultured in ESC medium for another 72 h before experiments.

Generation of *Zfp961* KO mouse lines

C57BL/6 female mice (6–8 weeks old) were super-ovulated by injection with 6 IU of pregnant mare serum gonadotropin (PMSG), followed by injection of 8 IU of human chorionic gonadotropin (hCG) (San-Sheng Pharmaceutical Co., Ltd.) 48 h later. The super-ovulated female mice were mated with C57BL/6 male mice and zygotes were collected 20 h later. Cas9 mRNA (100 ng/µL) and 3 sgRNAs (25 ng/µL each) were injected into zygotes using a Piezo-driven micromanipulator and then cultured in G1 medium with amino acids at 37°C in a humidified atmosphere of 5% CO_2. At ~28 h after this, 2-cell stage embryos were transferred into the oviducts of pseudo-pregnant ICR mothers. The mMESSAGE mMACHINE T7 Ultra Kit (Ambion, Thermo Fisher Scientific, AM1345) and MEGA shortscrip T7 Transcription Kit (Ambion, Thermo Fisher Scientific, AM1354) were used for the synthesis of Cas9 mRNA and sgRNAs, respectively. All Cas9 mRNA and sgRNAs were purified according to a standard protocol by phenol:chloroform extraction and ethanol precipitation, and then dissolved in DNase/RNase-free water. For genotyping, tail tips from *Zfp961*-KO mice were lysed in SENT lysis buffer (1% SDS, 5 mM EDTA, 0.4 M NaCl, 20 mM Tris-HCl, and 400 µg/mL Proteinase K) and genomic DNA was extracted as a template for PCR. The purified PCR products with large fragment deletion from *Zfp961*-KO mice were cloned into pLB vectors using the Lethal Based Fast Cloning Kit (Tiangen) to sequence. The sgRNAs and genotyping PCR primers used in this study are listed in *SI Appendix* Dataset S6.

Generation of *Zfp961-GFP^flox/flox^* mESCs

pX330 plasmids encoding hCas9 were purchased from Addgene (#42230), and sgRNAs were cloned into the *BbsI* cut site. For *Zfp961-GFP^flox/flox^* mESCs, the donor sites (HAL: chr8:71,966,757-71,967,934; HAR: chr8:71,969,308-71,970,098; mm10) were cloned from mouse genomic DNA and in-fusion cloned into a pUC backbone. Donor plasmid and pX330 plasmids were

co-electroporated into mouse Rosa26-CreERT2 ESCs using Lonza 4D-nucleofactor units (Lonza, program no. A-023) with a puromycin resistance plasmid. mESCs were selected using puromycin, and colonies were picked after 5 – 7 days for subsequent expansion and genotyping. The sgRNAs and genotyping PCR primers used in this study are listed in *SI Appendix* Dataset S6.

Construction of *ZNF417* and *ZNF587* knockout cell lines.

To construct *ZNF417* and *ZNF587* double knockout (dKO) cell lines and control cell lines, pX330 with specific sgRNAs or scrambled sgRNAs (CGGGTCTTCGAGAAGACCT) and GFP vectors were co-transfected into HEK293T, Jurkat or K562 cells following Lonza 4D-nucleofector instructions. A BD FACS Aria II was used to isolate single GFP positive cells for subsequent expansion and genotyping. The sgRNAs and genotyping PCR primers used in this study are listed in *SI Appendix* Dataset S6.

RNA extraction and mRNA-seq analysis

RNA was prepared from cells or embryos using the Eastep Super RNA extraction kit (Promega). RNA-seq libraries were prepared from total RNA from cells or embryos using the TruSeq stranded mRNA sample preparation kit (Illumina) and sequenced on a NovaSeq6000 (Illumina). Approximately 16 - 20 million Q30 >90% trimmed and filtered high-quality 150 bp paired-end reads per sample were mapped to the mouse (mm9) or human genome (hg38) using TopHat v2.1.1 (1) with default parameters. Gene counts were overlapped with the exons from ReFlat (UCSC) and ERVs counts were overlapped using RepeatMasker (UCSC) and then normalized to FPKM by Cufflinks (2) for differential gene expression analysis. Expression visualized tracks were generated with bigwig files generated by bamCoverage (3) and tdf files from igvtools (IGV 2.8.6).

ChIP-seq and ChIP-qPCR analysis

Zfp961-GFP^flox/flox mESCs were used to perform ZFP961, KAP1, H3K9me3 and H3K27ac ChIP, and 4-OHT-treated *Zfp961-GFP^flox/flox* mESCs were used as a control. For human ZNFs, GFP-fused ZNF417 and ZNF587 were cloned into a piggyBac transposon vector and transfected into HEK293T cells. Anti-GFP ChIP experiments were performed 48 h after transfection.

For ChIP experiments, $2–4 \times 10^7$ cells were cross-linked for 10 minutes at room temperature by the addition of formaldehyde solution at final concentration of 1% followed by quenching with glycine. Cells were washed twice with PBS, then the supernatant was aspirated and the cell pellet was stored at −80°C. Pellets were then lysed, resuspended in 1 mL of LB1 on ice for 10 min (50 mM HEPES-KOH pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% Glycerol, 0.5% NP40, 0.25% Tx100, and protease inhibitors), then after centrifugation resuspend in LB2 on ice for 10 min (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors). After centrifugation, cells were resuspended in LB3 (10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% NaDOC, 0.1% SDS and protease inhibitors) for histone marks and SDS shearing buffer (10 mM Tris pH8, EDTA 1 mM, SDS 0.15% and protease inhibitors) for transcription factors and sonicated (Covaris settings: 5% duty, 200 cycle, 140 PIP, 20 min), yielding genomic DNA fragments with a bulk size of 100 – 300

bp. Coating of beads with specific antibodies (GFP (Invitrogen A-6455) for ZFPs, KAP1 (Abcam ab22553), H3K9me3 (Abcam ab8898) and H3K27ac (Active Motif 39133) for histone marks) was carried out during the day at 4°C, then chromatin was added to the beads overnight at 4°C for histone marks while antibodies for transcription factors were first incubated with chromatin with 1% Triton and 150 mM NaCl. Subsequently, washes were performed with 2x Low Salt Wash Buffer (10 mM Tris pH 8, 1 mM EDTA, 150 mM NaCl, 0.15% SDS), 1x High Salt Wash Buffer (10 mM Tris pH 8, 1 mM EDTA, 500 mM NaCl, 0.15% SDS), 1x LiCl buffer (10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA, 250 mM LiCl, 1% NP40, 1% NaDOC) and 1 with TE buffer. Final DNA was purified with QIAGEN Elute Column. The ChIPed DNA was then subjected to NGS sequencing or qPCR analysis.For ChIP-seq, up to 10 nanograms of ChIPed DNA or input DNA (Input) was subjected to sequencing library preparation. Sequencing was performed on a Novaseq6000 (Illumina) following library construction using the KAPA Hyper Prep Kit (Roche). Next, 150-bp paired-end reads per ChIP-seq were mapped to the mouse (mm9) genome or human (hg38) genome using Bowtie2 (4) with default parameters. Peaks were called in MACS using GFP as control and calling p values smaller than 1e-6 significant (5). Consensus target motifs were derived from peak regions using MEME-ChIP (https://meme-suite.org/meme/tools/meme-chip). ZFP binding peak analysis was performed using BedTools (6), and the matrix between CHIP peaks and repeats was calculated using ComputeMatrix. Heatmaps and profile plots were then generated by plotHeatmap or plotProfile using the calculated matrix generated by ComputeMatrix (3).

The following public ChIP-seq data sets were used to generate heatmaps: mouse KAP1 (Sequence Read Archive [SRA] accession: SRR611529), mouse SETDB1 (SRA accession: SRR031683), human H1 ESCs KAP1 (SRR3178875), and HEK293T H3K9me3 (SRR8983692).

For ChIP-qPCR to test the binding of ZFPs to HIV-1 sequences, GFP-fused ZNF417, ZNF587 and ZFP961 were cloned into a piggyBac transposon vector, respectively, and transfected into HEK293T cells. Then cells were infected with pNL4.3-RFP pseudo-virus virus after 4 – 6 h of transfection. Cells were harvested and crosslinked after 48 h of infection. The chromatin was sonicated and IPed with GFP antibody (Invitrogen A11122). For ChIP-qPCR to test the enrichments of POLII, H3K9me3 and H3K27ac on HIV-1 LTR, *ZNF417 ZNF587* dKO HEK293T cells or control HEK293T cells (generated by scrambled sgRNAs) were subjected to ChIP with anti-RNA POLII (CTD4H8, Millipore), anti-H3K9me3 (Abcam ab8898), and anti-H3K27ac (Active Motif 39133) antibodies, respectively, after 48 h of viral infection. ChIPed DNA was quantified by real-time PCR with specific primers listed in *SI Appendix* Dataset S6.

Co-immunoprecipitation (coIP) and Western blotting

For coIP, HEK293T cells were co-transfected with Flag-tagged ZFP-KRABs, HA-tagged SETDB1 and Myc-tagged KAP1 plasmids. After 48 h, cells were pelleted and washed twice with cold PBS, and proteins were extracted in RIPA buffer or IP buffer (20 mM Tris (PH7.5), 150 mM NaCl, 1% Triton X-100, and proteinase inhibitor cocktail (PIC) (Roche)). Cell extracts were incubated with protein A/G

magnetic beads (ThermoFisher) crosslinked with anti-HA (C29F4 Cell Signaling Technology), anti-FLAG (Sigma F1804), or anti-Myc (BBI) antibodies at 4°C overnight, respectively. Beads were then washed three times with BC150 buffer (10 mM Tris pH 7.8, 0.5 mM EDTA, 10% glycerol, 150 mM NaCl, PIC).

For Western blotting, the primary antibodies α-GFP (Invitrogen A11122), α-GAPDH (Proteintech Cat# 60004-1-Ig), α-ZNF587/ZNF417 (Abcam ab111697), α-HIV-1 GAG-p24 (Abcam ab9071), and α-RFP (Abcam ab183628) were used to test specific proteins.

Luciferase assays

For motif binding assays, annealed oligos (different PBS sequences) were cloned downstream of the pGL3-SV40 vector (Promega). For native promoters, ETn or IAP LTR was cloned from mESCs genomic DNA and the HIV-1LTR was cloned from the pNL4.3 vector (NIH AIDS Reagent Program). These native promoters were cloned into the pGL3-basic (Promega) vector by in-fusion cloning (Tiangen). These pGL3 reporter plasmids were co-transfected with different ZNF genes (cloned into the pcDNA3.1 vector or piggyBac-GFP vector) into HEK293T cells using Vigofect (Invitrogen) together with the Renilla luciferase-expressing pRL-SV40p (Addgene #27163) vector for internal normalization. Luciferase activity was measured 2 days after transfection using this dual-luciferase reporter assay system (Promega). Sequences of promoters and LTRs are listed in *SI Appendix* Dataset S6.

Virus production

Pseudo type HIV-1 and pSicoR-RFP pseudo-viral stocks were generated by co-transfecting 10 cm dishes of HEK293T cells with 10 μg of pSicoR-mCherry (Addgene #31847), 6 μg of psPAX2(Addgene #12260) and 4 μg of pMD2.G (Addgene #12259).For pNL4.3-RFP pseudo-virus virus, pNL4.3-RFP vector was generated from the HIV-1 full length vector pNL4.3 with *Env* deletion and replaced with a mCherry-T2A-Luciferase cassette. 16 μg of pNL4.3-RFP and 4 μg of pMD2.G were co-transfected into 10 cm dishes of HEK293T cells.The medium was changed 6–8 h after transfection, and supernatants were collected after 48 – 72 h, filtered through a 0.45 μm membrane to clear cell debris, and then collected for infection or concentrated by PEG8000 addition at a 10% final concentration rolling overnight at 4°C, followed by 3500 g centrifuged for 30 min.

Cell infection

HEK293T cells were infected with pSicoR-RFP pseudo-virus supernatants diluted 1:30 for ZFP OE cells or 1:50 for ZFP KO cells. K562 and Jurkat cells were directly infected with pSicoR-RFP pseudo-virus supernatants without dilution.

For preparing pNL4.3-RFP pseudo-virus, four 10 cm dishes of HEK293T supernatants were harvest and concentrated to 1 mL in proper culture medium. $1 - 2 \times 10^5$ cells were infected with 200 μL of virus for 24 – 48 h.

Zfp961-GFP^flox/flox mESCs were treated with 4-OHT for 48 h and cultured for another 1 - 2 passages, then infected with concentrated HIV-1 virus.

Mouse PBLs infection

Approximately 4-week-old *Zfp961* WT or KO mice were anesthetized and whole blood was collected immediately after removing the eyeball. PBLs were separated and isolated with a Peripheral Blood Lymphocyte Isolation Kit (Solarbio). PBLs were cultured in RPMI1640 medium containing 10% FBS and 100 U/mL IL-2 and infected with HIV-1 pseudo-virus at the same viral concentration as K562 or Jurkat cells mentioned previously.

qPCR quantitation of viral DNA copies

For viral DNA copy quantitation, genomic DNA from infected cells was harvested and purified using phenol chloroform 12 to 96 h post-infection, respectively. Early HIV-1 reverse transcripts were quantitated with primers qHIVU5-F/R targeting the 5'LTR U5 of the pSicoR-RFP and pNL4.3-RFP pseudo-virus. Late reverse transcripts were quantitated with primers specific to the mCherry gene in the pSicoR-mCherry and pNL4.3-RFP vectors. Standards curves were generated with a gradient dilution of pSicoR-mCherry and pNL4.3-RFP vectors for each primer.100 ng of cell DNA was used to perform qPCR using MagicSYBR Mixture (CWBio CW3008).

For integrated DNA quantitation of virus, HEK293T cells were infected with lentivirus pCW57.1-GFP as a control to generate an integrated HIV-1 DNA standard. Infected cells were then cultured for several weeks in the presence of puromycin (1 μg/mL). After passaging, unintegrated DNA was degraded. Thus, the copy numbers of integrated viral DNA matched the total HIV-1 DNA copy number. After puromycin selection, DNA from infected cells was purified and diluted with DNA from uninfected cells to prepare a standard. Standard DNA was first quantified by total HIV-1 DNA copy number using U5 primers. Then, an Alu-based two round nested PCR was performed to quantitate integrated viral DNA. First, PCR reaction mixtures containing 100 ng standards or test DNA samples, primers L-M667 and Alu1 Alu2 and Super Taqman Mixture (CWBio CW2698) were used. PCR was for 2 min at 95°C, followed by 20 cycles of 15 s at 95°C, 10 s at 60°C and 3 min at 72°C. PCR products were diluted 1/10 for a second nested real-time PCR using primers Lambda T and AA55M, probe MH603 and Super Taqman Mixture.

The PCR primers are listed in *SI Appendix* Dataset S6.

Identification of human ZNFs for PBS targeting

PBS sequences were extracted from human ERV published dataset (7). PBS motifs were generated according to the types of PBS listed by MEME (Multiple Em for Motif Elicitationhttps://meme-suite.org/). PBS motif-containing regions in human genome were identified by FIMO (Find Individual Motif Occurences, https://meme-suite.org/). These regions were overlapped with 3173 human ERV locus by BEDTools. Human ZNF ChIP peaks (GSE78099) were overlapped with the above PBS motif-containing ERV regions using BEDTools (6).

Viral titer calculation assays

pLenti-CMV-GFP-Neo (Addgene #17447) was mutated at its PBS site by PCR. Briefly, wild type PBS-Lys3 was changed into PBS-Lys3 mutant (TGGCGCCtGAACAGGGAC), PBS-Pro (TGGGGGCTCGTCCGGGAT) or PBS-Phe (TGGTGCCGAAACCCGGGA), respectively. Virus was produced by co-transfection of pLenti-GFP vector with psPAX2 and pMD2.G into HEK293T cells. Supernatants were harvested 48 h after transfection and diluted from 10−1 to 10−3. HEK293T cells were infected for 48 h in 96 well dishes. GFP+ cells were counted to calculate viral titers.

Isolation and infection of primary CD4⁺ T Cells

Primary human T cells were isolated from PBMCs from healthy, anonymous donor blood and processed within 12 h. Primary CD4⁺ T cells were harvested by positive selection using anti-human CD4 microbeads (Miltenyi Biotec). Isolated CD4⁺ T cells were suspended in complete Roswell Park Memorial Institute (RPMI) media, consisting of RPMI-1640 (Gibco) supplemented with 10% FBS (Hyclone) and 2 mM L-glutamine. T cells were activated as previously described (8) with some modifications. T cells were stimulated on anti-CD3 coated plates (coated 2 h at 4°t with 10 µg/mL anti-CD3 in the presence of 5 µg/mL soluble anti-CD28 (Invitrogen) and 20 U/ml IL-2 (R&D Systems).

To KO *ZNF417* and *ZNF587* in activated T cells, electroporation was performed using the Amaxa P3 Primary Cell 96-well Nucleofector kit and a 4D-Nucleofecter (Lonza). Approximately 5 µg of pX330 plasmids containing sgRNAs targeting *ZNF417* and *ZNF587* or pX330 vector containing control sgRNAs (scrambled sgRNAs) were transfected into 106 activated T cells, respectively. Cells were then cultured in RPMI1640 medium containing 20 U/mL IL-2 for 3 - 5 days. 2x105 cells from the ZNF KO group or control group were infected with concentrated pNL4.3-RFP pseudo-virus virus for 48 h, respectively. Infected T cells were harvested and stained with anti-CD4-FITC (BD 555346) and anti-CD25-APC (BD 555434) and were then monitored by flow cytometry for RFP signals for validation of viral infection efficiencies. Genomic DNA was also extracted for sequencing to confirm the knockout of *ZNF417* and *ZNF587*.

Genome-wide sequencing of retroviral integration sites

We infected *ZNF417 ZNF587* dKO HEK293T or control HEK293T cells with NL4.3 pseudo-virus for 48 h. Then genomic DNA was purified from cell lysis using phenol chloroform. Sequencing libraries were prepared from 10 µg of extracted DNA according to the protocol in a previous report with slight modifications (9). DNA was sheared into 300 - 500 bp random fragments using a Covaris Adaptive Focused Acoustics sonicator (Covaris, Woburn, MA). The sheared DNA fragments were end-repaired and a single dA was added to the 3' ends using the KAPA DNA Hyper Prep kit. A partially double stranded linker (5'GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACT3', 5'–PO4-GTCCCTTAAGCGGAG-3'-C6) with a one nucleotide 3' T overhang was ligated to these genomic DNA fragments. The 3' dA overhang on the genomic DNA fragments and the 3' T overhang on the linker DNA prevented genomic-genomic DNA and linker-linker ligation, increased the efficiency

of the desired ligation, and reduced the artifacts associated with cross-ligation commonly seen with restriction-enzyme-based linker-mediated PCR (LM-PCR). The integration junctions were selectively amplified using one primer that matched the HIV-1LTR sequence and the other primer that matched the single stranded portion of the linker. Nested PCR was performed with an inner primer pair; one primer matched the sequence of the LTR and the other primer matched the linker (outer: HIV-3LTR, 5'TGTGACTCTGGTAACTAGAGATCCCTC3'; Inner: HIV-3LTRnest, 5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNN NCCCTTTTAGTCAGTGTGGAAAATC3'; outer: HIV-5LTR,TCAGGGAAGTAGCCTTGTGTGTGGT; inner: HIV-5LTRnest,

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNN NGTCTTTTTTGGGACCAAATTAGCCC; outer: Linker-P1, 5'GTAATACGACTCACTATAGGGC3'; inner: PE2_Linkernest, 5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNN NNNNAGGGCTCCGCTTAAGGGAC3'; NNNNNN stands for barcodes).

Sequencing was performed on a Novaseq6000 (Illumina). Processing, except for the merging of overlapping paired-end read sequences and read mapping, were performed using in-house Perl scripts (available at https://github.com/odehir/VINSSRM). The primer sequences for the LTR and linker are found in each 5′-end of the original paired-end read sequences (read1 and read2). Each set of overlapping paired-end read sequences was merged into one read (read3). Then, the merged read sequences and paired-end read sequences were subjected to trimming of linker sequences and subsequent split-read mapping onto the human genome (hg38) and HIV-1 reference sequences (NL4.3), respectively. The integration sites and breakpoints were assigned by nucleobase positions onto the human genome. The breakpoints were used to evaluate whether the integration sites   were identified from clonally expanded cells. The LTR ends were assigned by nucleobase positions onto an HIV-1 reference genome. The nucleobase positions of the LTR ends were used to evaluate whether the viral ends were intact or aberrant.
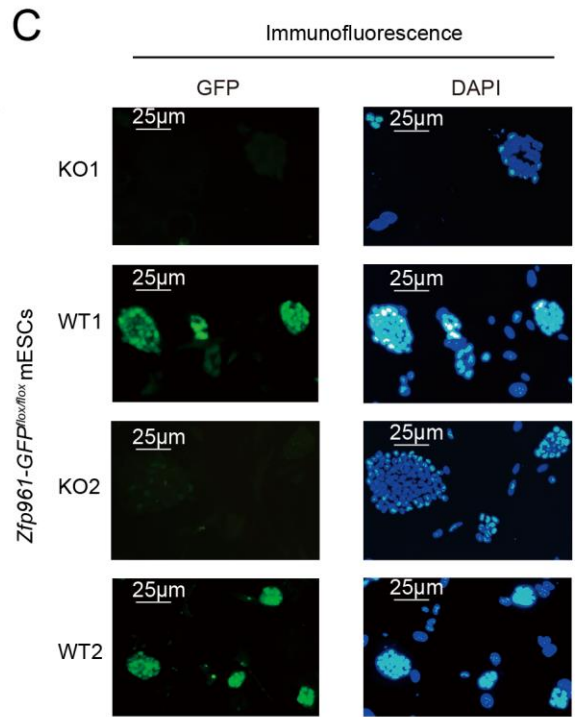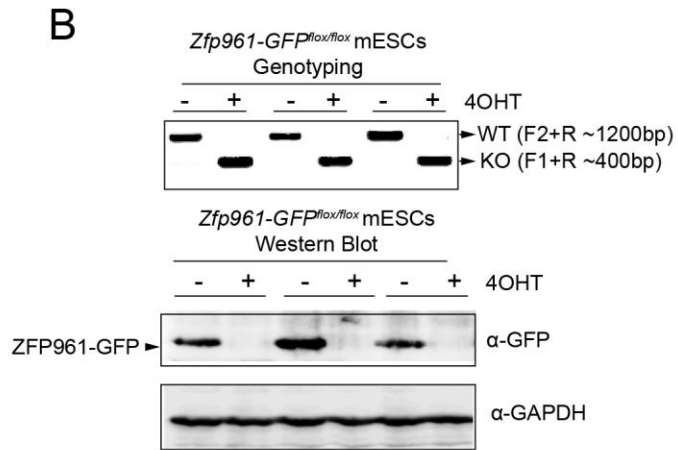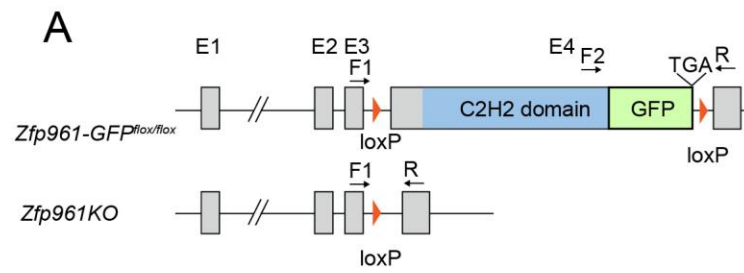
A

Zfp961-GFP^flox/flox

E1 E2 E3 F1 loxP C2H2 domain E4 F2 GFP TGA R loxP

Zfp961KO

F1 R loxP

B

Zfp961-GFP^flox/flox mESCs
Genotyping

- + - + - + 4OHT

► WT (F2+R ~1200bp)
► KO (F1+R ~400bp)

Zfp961-GFP^flox/flox mESCs
Western Blot

- + - + - + 4OHT

ZFP961-GFP ► α-GFP

α-GAPDH

C

Immunofluorescence

GFP   DAPI

KO1 25μm 25μm

WT1 25μm 25μm

KO2 25μm 25μm

WT2 25μm 25μm

Zfp961-GFP^flox/flox mESCs

Fig. S1. Generation of *Zfp961-GFP^{flox/flox}* mESCs cell lines.

(A) Schematic overview of *Zfp961-GFP^{flox/flox}* targeting strategy in mESCs.

(B) Genotyping PCR (upper panel) and Western blot (lower panel) analysis to confirm deletion of *Zfp961* alleles and loss of ZFP961-GFP protein upon 4-OHT treatment of *Zfp961-GFP^{flox/flox}*; Rosa26-CreERT2 mESCs.

(C) Immunofluorescent images confirming loss of ZFP961-GFP protein upon 4-OHT treatment of *Zfp961-GFP^{flox/flox}*; Rosa26-CreERT2 ESCs. *Zfp961* WT and KO mESCs were stained with DAPI and GFP antibodies. Scale bars, 25 µm.
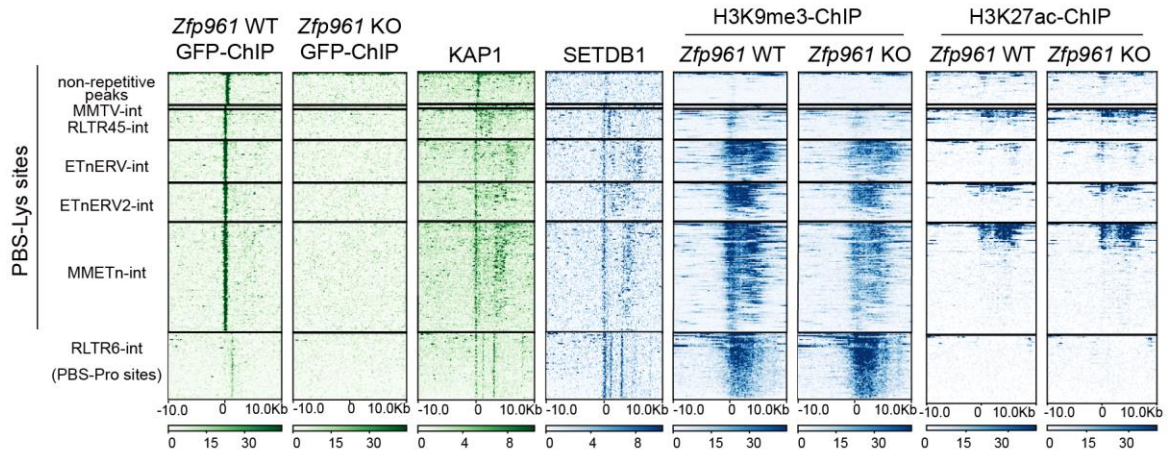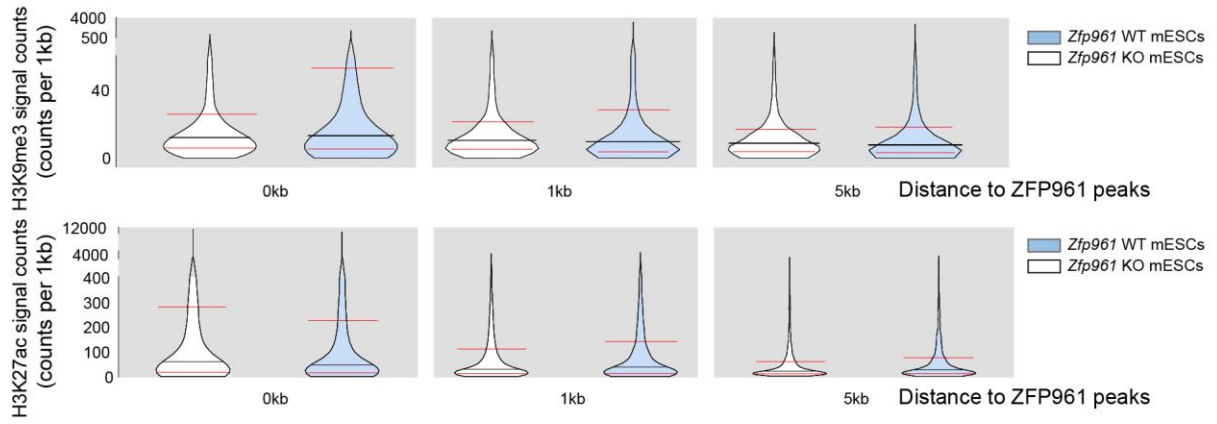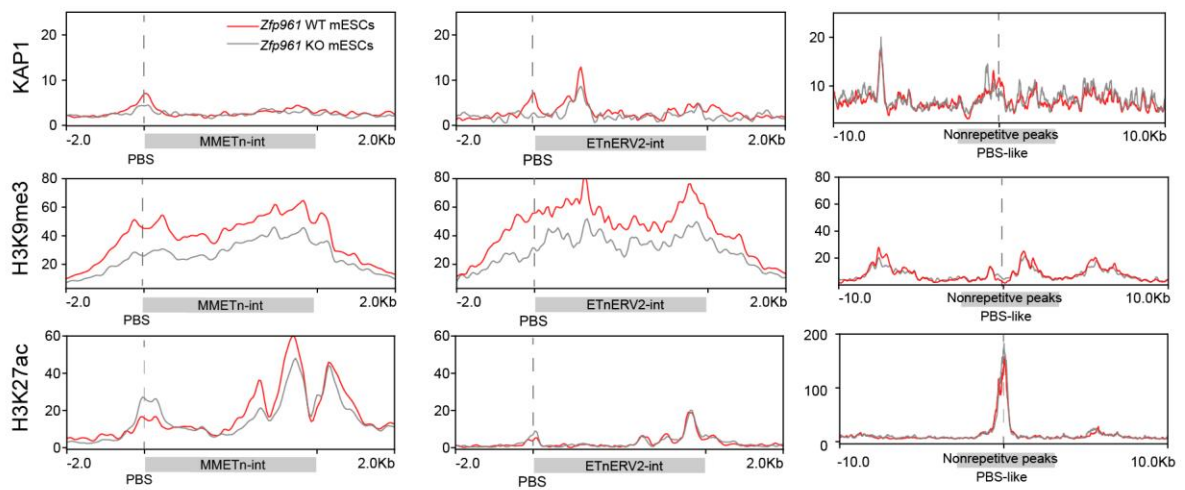
A

Zfp961 WT GFP-ChIP · Zfp961 KO GFP-ChIP · KAP1 · SETDB1 · H3K9me3-ChIP (Zfp961 WT, Zfp961 KO) · H3K27ac-ChIP (Zfp961 WT, Zfp961 KO)

PBS-Lys sites:
- non-repetitive peaks
- MMTV-int
- RLTR45-int
- ETnERV-int
- ETnERV2-int
- MMETn-int
- RLTR6-int (PBS-Pro sites)

-10.0 0 10.0Kb

B

H3K9me3 signal counts (counts per 1kb)
4000 / 500 / 40 / 0

H3K27ac signal counts (counts per 1kb)
12000 / 4000 / 400 / 300 / 100 / 0

0kb   1kb   5kb   Distance to ZFP961 peaks

Zfp961 WT mESCs
Zfp961 KO mESCs

C

KAP1 · H3K9me3 · H3K27ac

Zfp961 WT mESCs
Zfp961 KO mESCs

-2.0 PBS MMETn-int 2.0Kb
-2.0 PBS ETnERV2-int 2.0Kb
-10.0 Nonrepetive peaks PBS-like 10.0Kb

Fig. S2. ChIP-seq analysis shows ZFP961 as a potential ERVK repressor.

(A) Heatmaps of ZFP961, KAP1, SETDB1, H3K9me3 and H3K27ac occupancy at ERV-associated PBS-Lys sites and strong non-repetitive ZFP961 ChIP-seq peak regions. 20 kb regions are displayed with the PBS-Lys sequence or non-repetitive peaks located at the center. PBS-Pro sites-containing RLTR6-int are shown as control.

(B) Violin plots of H3K9me3 and H3K27ac signals at different distances to ZFP961 peaks. Counts are normalized to 1kb region. Quartiles are marked in red.

(C) Profiles of KAP1, H3K9me3 and H3K27ac signals at representative ERVK regions and nonrepetitive ZFP961 peaks in *Zfp961* WT (red) and KO (grey) mESCs.
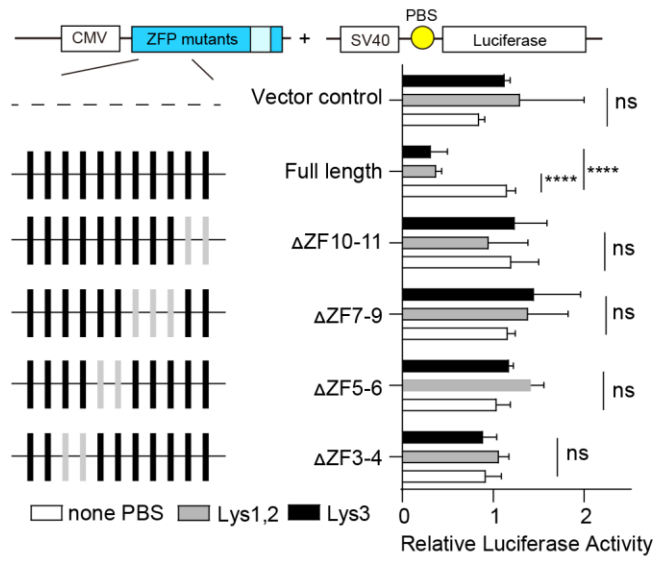
A
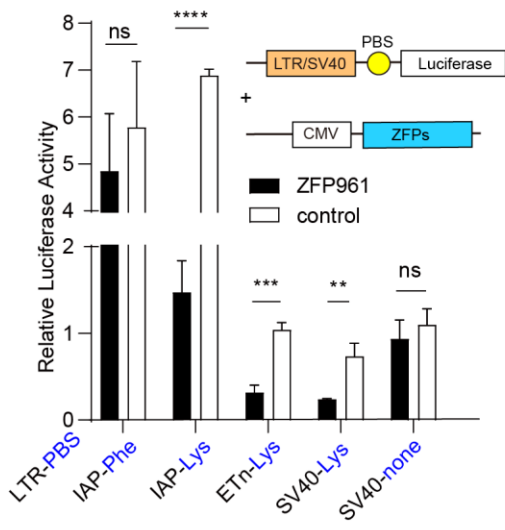
B

Fig. S3. ZFP961 targets PBS-Lys sites for transcriptional repression.

(A) Relative luciferase activity of HEK293T cells overexpressing ZFP961 or control plasmid (pCDNA3.1+) and IAP, ETn, SV40 promoter-driven luciferase plasmid containing two tRNA PBS: PBS-Phe or PBS-Lys. t test: Error bars indicate standard deviation; ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns $p>0.05$; n=3.

(B) Relative luciferase activity of HEK293T cells overexpressing SV40 promoter-driven luciferase plasmid containing the PBS-Lys1,2 or PBS-Lys3 and ZFP961 with the indicated zinc finger deletions marked in gray. t test: Error bars indicate standard deviation; **** $p<0.0001$, ns $p>0.05$; n=3.
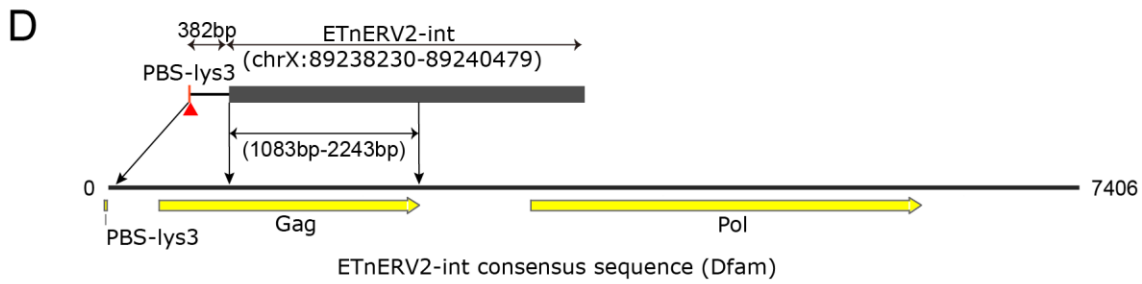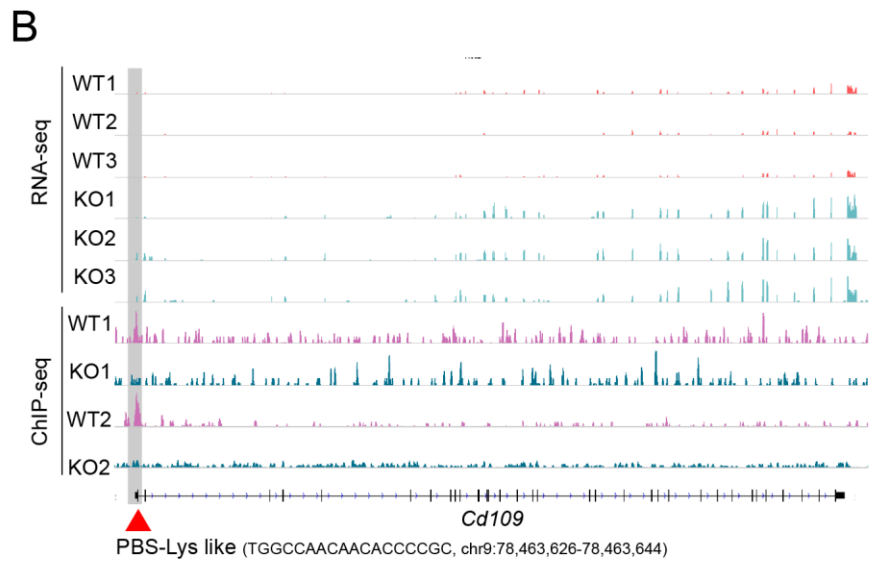
A

KO1 0-250
WT1 0-250
KO2 0-250
WT2 0-250

RNA-seq

Zfp961

C

ERVs expression test
Zfp961 WT VS KO

$-Log_{10}$(P Value)

ETnERV2-int
(FC:1.26348,
PV:0.0049)

Log_2(Fold Change)

B

WT1
WT2
WT3
KO1
KO2
KO3

RNA-seq

WT1
KO1
WT2
KO2

ChIP-seq

Cd109

PBS-Lys like (TGGCCAACAACACCCCGC, chr9:78,463,626-78,463,644)

D

382bp
ETnERV2-int
(chrX:89238230-89240479)

PBS-lys3

(1083bp-2243bp)

0   7406

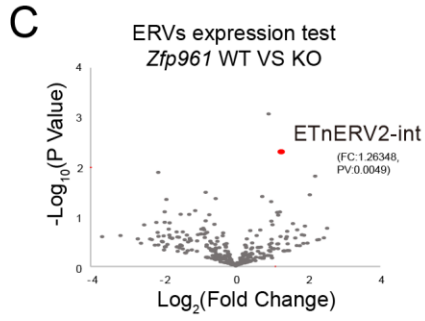PBS-lys3   Gag   Pol

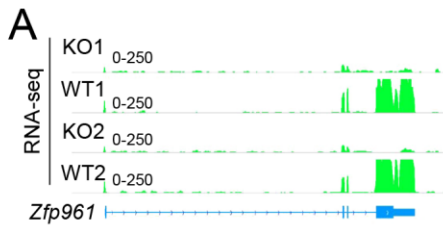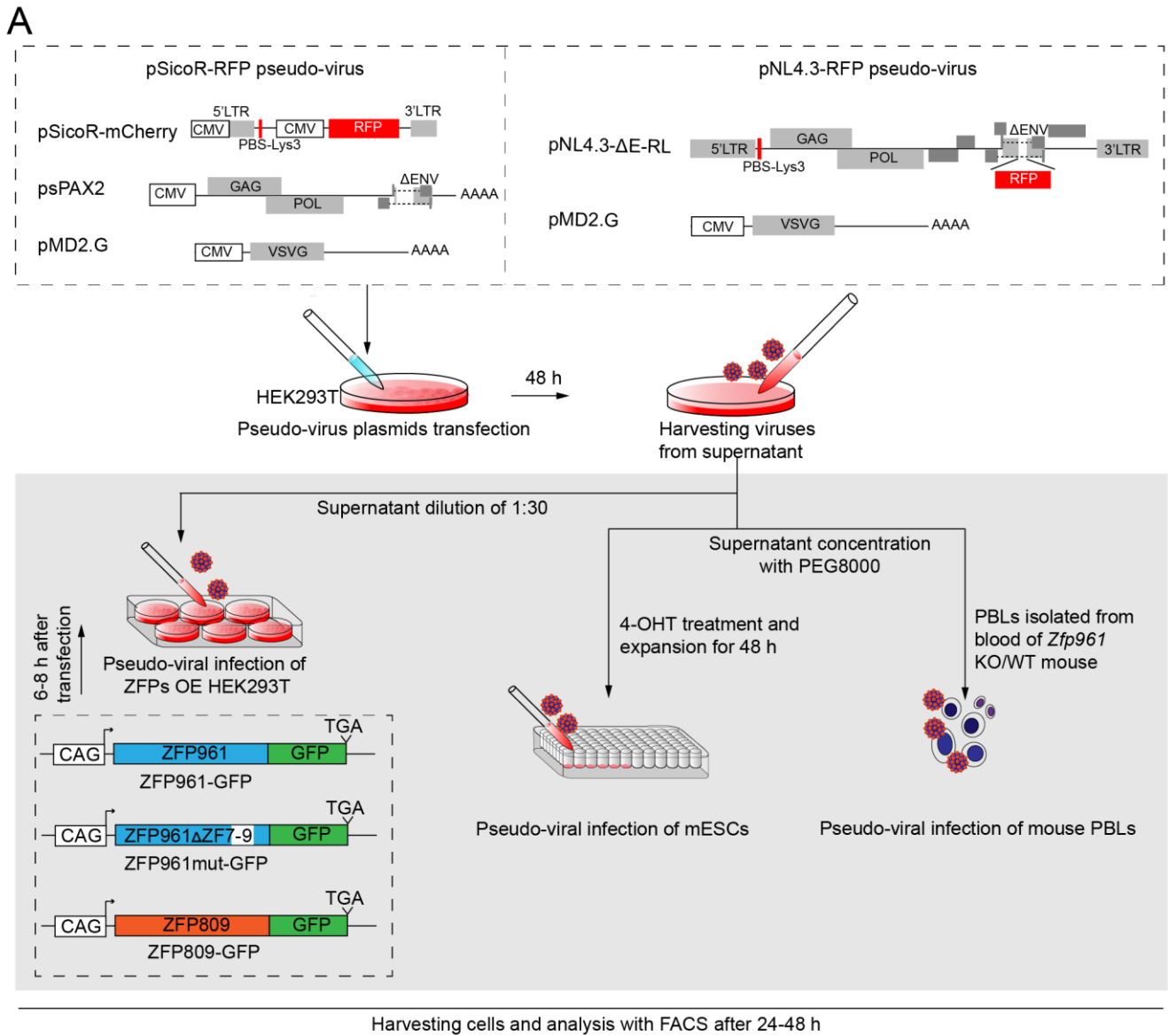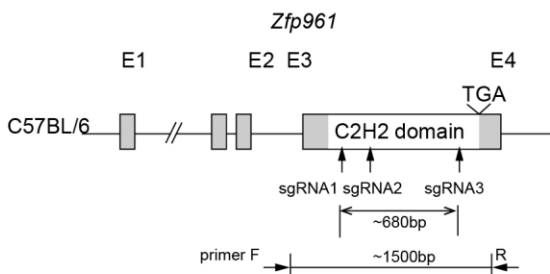ETnERV2-int consensus sequence (Dfam)

Fig. S4. Deletion of *Zfp961* leads to increased expression of genes and repeat elements.

(A) RNA-seq signal in *Zfp961* KO or WT mESCs at the *Zfp961* locus.

(B) RNA-seq and ChIP-seq signal in *Zfp961* KO or WT mESCs at the *Cd109* locus.

(C) Volcano plot of repeats expression in *Zfp961* KO or WT mESCs. Red dots present upregulated repeat elements.

(D) Schematic overview of a reactivated ETnERV2-int locus compared to consensus ETnERV2-int sequence.

A

pSicoR-RFP pseudo-virus

pSicoR-mCherry | 5'LTR CMV | CMV | RFP | 3'LTR
PBS-Lys3

psPAX2 | CMV GAG POL ΔENV AAAA

pMD2.G | CMV VSVG AAAA

pNL4.3-RFP pseudo-virus

pNL4.3-ΔE-RL | 5'LTR GAG POL ΔENV 3'LTR
PBS-Lys3 RFP

pMD2.G | CMV VSVG AAAA

HEK293T
Pseudo-virus plasmids transfection

48 h

Harvesting viruses
from supernatant

Supernatant dilution of 1:30

Supernatant concentration
with PEG8000

6-8 h after transfection

Pseudo-viral infection of
ZFPs OE HEK293T

CAG | ZFP961 | GFP TGA
ZFP961-GFP

CAG | ZFP961ΔZF7-9 | GFP TGA
ZFP961mut-GFP

CAG | ZFP809 | GFP TGA
ZFP809-GFP

4-OHT treatment and
expansion for 48 h

Pseudo-viral infection of mESCs

PBLs isolated from
blood of *Zfp961*
KO/WT mouse

Pseudo-viral infection of mouse PBLs

Harvesting cells and analysis with FACS after 24-48 h

B

Zfp961

E1   E2 E3   E4
                                      TGA
C57BL/6 | C2H2 domain
sgRNA1 sgRNA2   sgRNA3
          ~680bp
primer F   ~1500bp   R

C

Zfp961 KO mouse genotyping

WT1 KO1 KO2 WT2 KO3

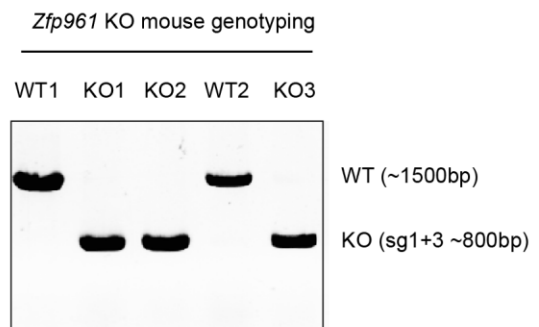WT (~1500bp)

KO (sg1+3 ~800bp)

Fig. S5. Schematic overview of HIV-1 pseudo-viral infection repression assay.

(A) Schematic overview of HIV-1 pseudo-viral infection repression assay. An HIV-1 pseudo-viral plasmid pSicoR-mCherry or pNL4.3-RFP, together with respective packing plasmids, were transfected into HEK293T cells to generate viral particles, which were then infected with HEK293T cells overexpressing ZFPs, *Zfp961-GFP^flox/flox* mESCs before and after 4-OHT treatment, or PBLs isolated from *Zfp961* WT and KO mice. Cells were harvested and subjected to FACS analysis 48 hours after infection.

(B) Strategy to generate *Zfp961* KO mice using CRISPR/Cas9.

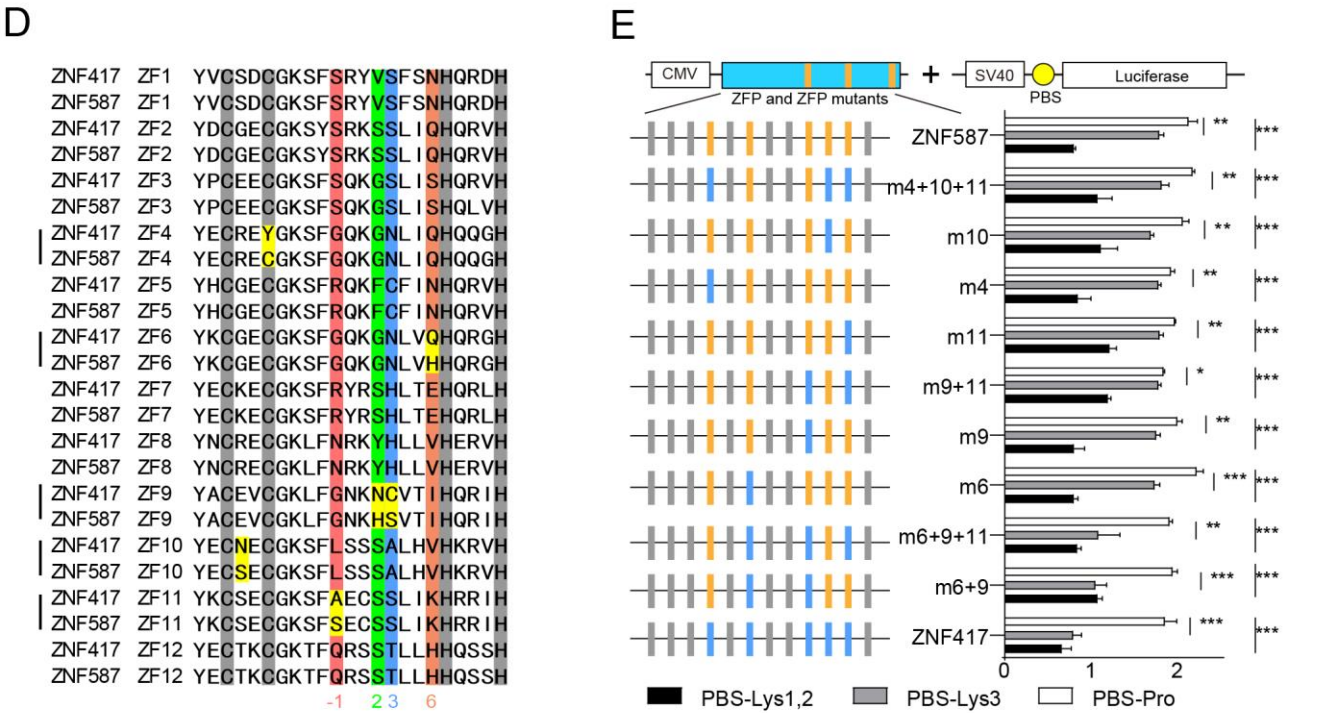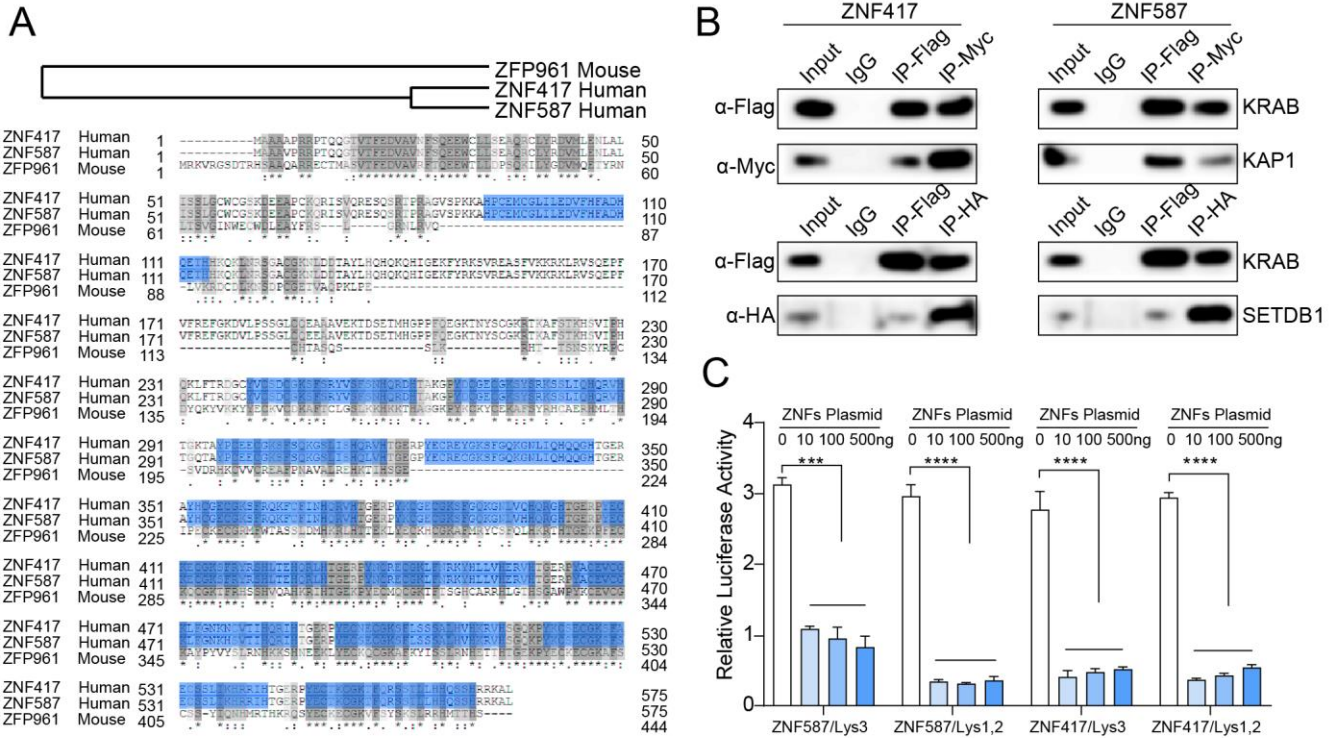(C) Genotyping PCR analysis to confirm the deletion of *Zfp961* alleles in *Zfp961* KO mice.

Fig. S6. Human ZNF417 and ZNF587 target PBS-Lys for repression.

(A) Amino acid sequence alignments of the ZNF417, ZNF587 and ZFP961 proteins.

(B) Co-IP and Western blotting of Flag-ZNF417 KRAB domain or Flag-ZNF587 KRAB domain with myc-KAP1 and HA-SETDB1.

(C) Relative luciferase activity assays to test the binding preference of ZNF417 and ZNF587 at PBS-Lys1, 2 and PBS-Lys3 sites. t test: Error bars indicate standard deviation. *** p<0.001, **** p < 0.0001 and n = 3.

(D) Amino acid sequence alignments of the ZNF417 and ZNF587 zinc fingers. The 'C2H2' amino acid residues are marked in grey. The fingerprint amino acid residues at the -1, 2, 3 and 6 positions are marked in red, green, blue and orange, respectively. The potential mutation sites that may influence binding capacity of PBS-Lys are marked in yellow.

(E) Relative luciferase activity of HEK293T cells overexpressing ZNF417, ZNF587 or ZNF587 zinc fingers mutants and an SV40 promoter-driven reporter containing PBS-Lys1,2, PBS-Lys3 or PBS-Pro as a control. Zinc fingers specific for ZNF587 are marked in yellow, and zinc fingers specific for ZNF417 are marked in blue. t test: Error bars indicate standard deviation. *** p < 0.001, **** p < 0.0001, ns p > 0.05, and n = 3.
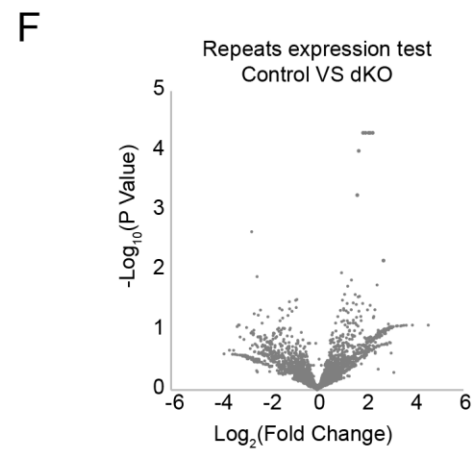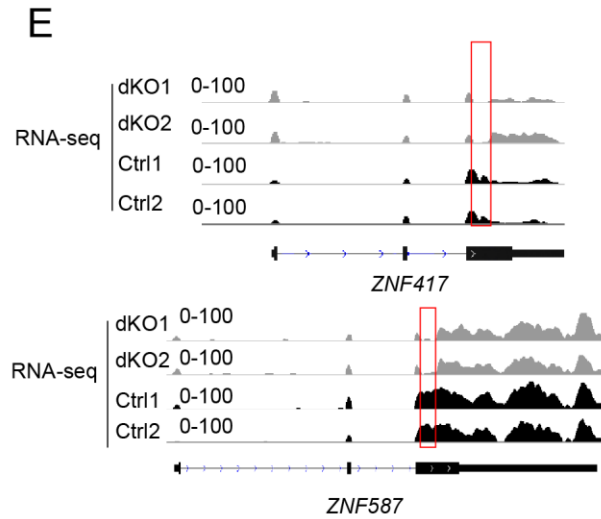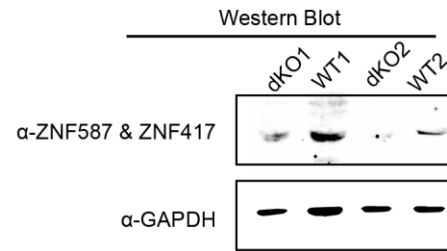
A — Color Key (1 3). Heatmap rows: ZNF749, ZNF789, ZNF44, ZNF568, ZNF350, ZNF212, ZNF785, ZNF506, ZNF271, ZNF574, ZNF362, ZNF497, ZNF76, ZNF486, ZNF281, ZNF281, ZNF417 &ZNF587. Columns: Oocyte, Zygote, 2-cell, 4-cell, 8-cell, Morulae, Blastocyst.

B — Color Key (0 2 4). Heatmap rows: MER11B, HERVK3–int, LTR14, LTR13, LTR13_, MER11C, LTR5A, LTR5B, LTR3A, HERVK22–int, LTR22C0, LTR22B2, LTR22, HERVK14C–int, HERVK11D–int, LTR13A, HERVK13–int, LTR22A, LTR14B, HERVK14–int, LTR14A, LTR3B_, LTR3B, LTR22C2, LTR22B1, MER11D, LTR3, MER9a3, LTR14C, HERVK9–int, MER9a1, LTR22E, LTR22B, LTR22C, MER9B, MER9a2, LTR5, LTR5_Hs, MER11A, HERVKC4–int, HERVK11–int, HERVK–int. Columns: Oocyte, Zygote, 2-cell, 4-cell, 8-cell, Morulae, Blastocyst.

C — ZNF417 and ZNF587 gene structure: sgRNAs, ~500bp, ~2500bp, E1, E2, E3, C2H2 domain.

D — K562. Genotyping: ZNF417 (WT1, KO1, WT2, KO2), ZNF587 (WT1, KO1, WT2, KO2). WT(~2500bp), KO(~2000bp). Western Blot: dKO1, WT1, dKO2, WT2. α-ZNF587 & ZNF417, α-GAPDH.

E — RNA-seq. ZNF417: dKO1 0-100, dKO2 0-100, Ctrl1 0-100, Ctrl2 0-100. ZNF587: dKO1 0-100, dKO2 0-100, Ctrl1 0-100, Ctrl2 0-100.

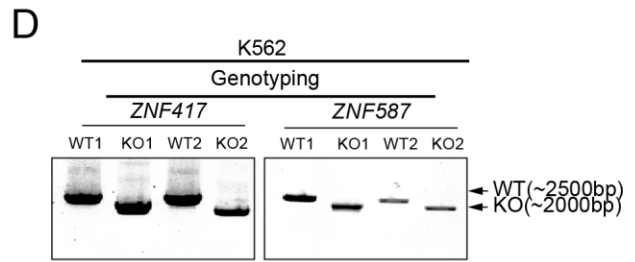**F** — Repeats expression test Control VS dKO. $-\text{Log}_{10}$(P Value) vs Log_2(Fold Change).
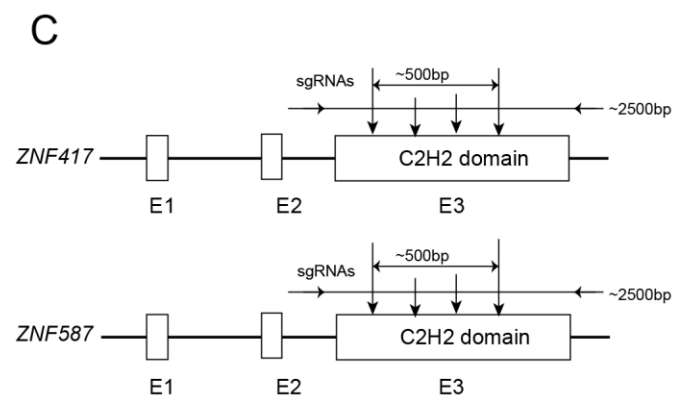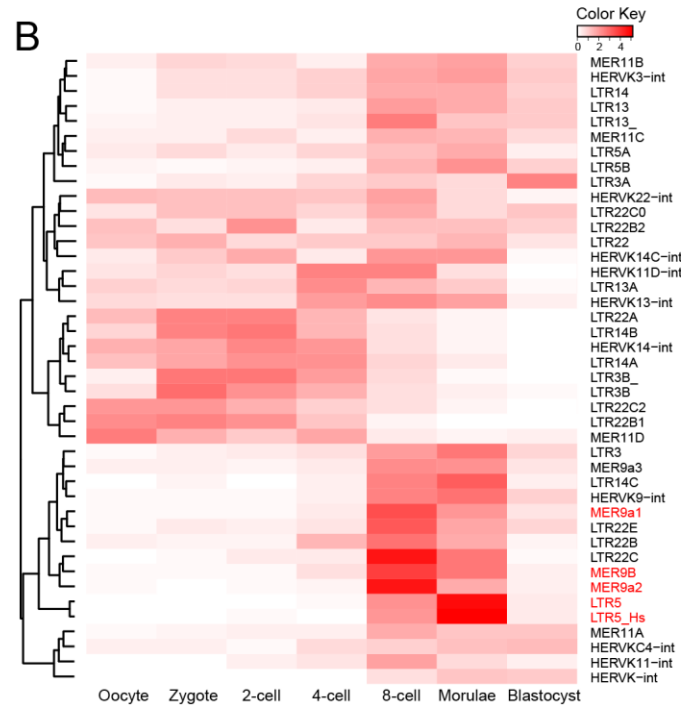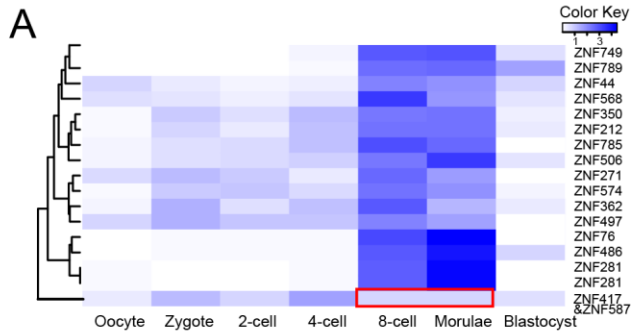
Fig. S7. Human ZNF417 and ZNF587 repress human ERVK.

(A-B) Heatmaps showing the expression levels of representative ZNFs (A) and human ERVKs (B) during different early embryonic developmental stages. Each row represents a ZNF or an ERVK region and each column represents a developmental stage. Embryonic genome activation-related ERVKs are marked in red.

(C) Strategy to knockout *ZNF417/ZNF587* using CRISPR/Cas9 in K562 cells.

(D) Genotyping PCR (upper panels) and Western blot (lower panels) analysis to confirm deletion of *ZNF417* and *ZNF587* alleles and loss of ZNF417 and ZNF587 proteins in K562 cells.

(E) RNA-seq signal at the *ZNF417* and *ZNF587* loci in *ZNF417 ZNF587* dKO or control K562 cells.

(F) Volcano plots of repeat expression in *ZNF417 ZNF587* dKO cells compared with control K562 cells. Red dots represent upregulated repeat elements.
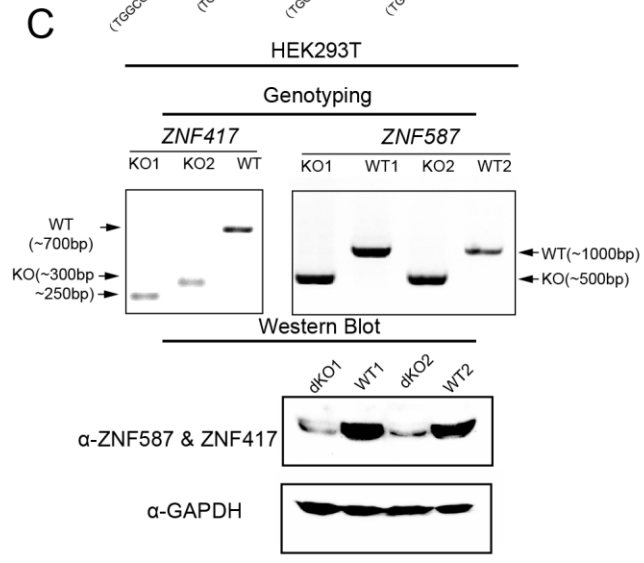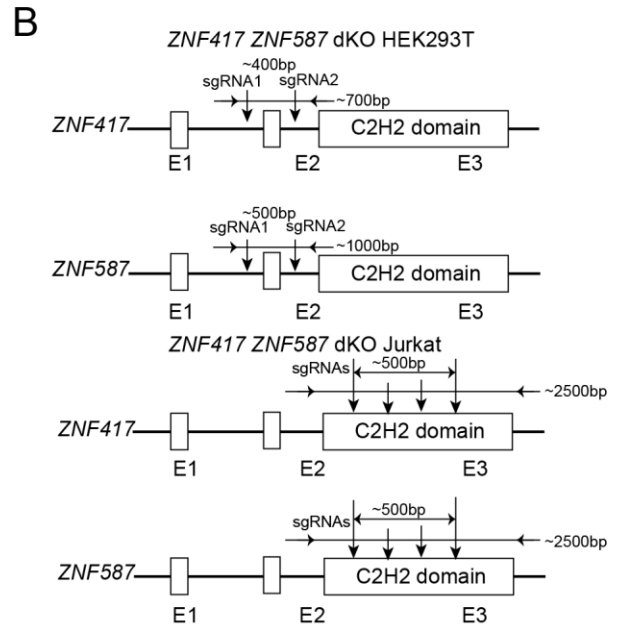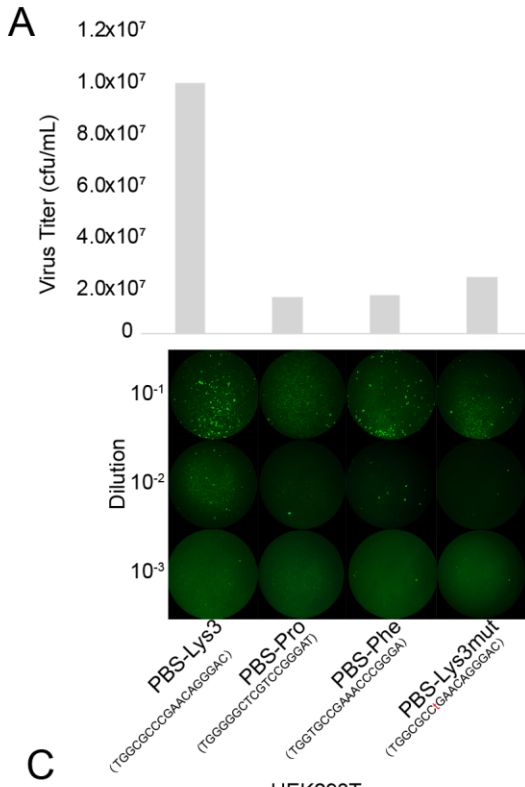
A

Virus Titer (cfu/mL)

1.2x10⁷
1.0x10⁷
8.0x10⁷
6.0x10⁷
4.0x10⁷
2.0x10⁷
0

Dilution

10⁻¹
10⁻²
10⁻³

PBS-Lys3
(TGGCGCCCGAACAGGGAC)

PBS-Pro
(TGGGGGCTCGTCCGGGAT)

PBS-Phe
(TGGTGCCGAAACCCGGGA)

PBS-Lys3mut
(TGGCGCCGGAACAGGGAC)

B

ZNF417 ZNF587 dKO HEK293T

~400bp
sgRNA1   sgRNA2
ZNF417 ▭——↓—▭—↓——[C2H2 domain]——— ~700bp
        E1   E2   E3

~500bp
sgRNA1   sgRNA2
ZNF587 ▭——↓—▭—↓——[C2H2 domain]——— ~1000bp
        E1   E2   E3

ZNF417 ZNF587 dKO Jurkat

sgRNAs   ~500bp
ZNF417 ▭———▭——[C2H2 domain]——— ~2500bp
        E1   E2   E3

sgRNAs   ~500bp
ZNF587 ▭———▭——[C2H2 domain]——— ~2500bp
        E1   E2   E3

C

HEK293T

Genotyping

ZNF417   *ZNF587*
KO1 KO2 WT   KO1 WT1 KO2 WT2

WT
(~700bp) →

KO(~300bp) →
~250bp →

← WT(~1000bp)

← KO(~500bp)

Western Blot

dKO1 WT1 dKO2 WT2

α-ZNF587 & ZNF417

α-GAPDH

D

Jurkat

Genotyping

ZNF417   *ZNF587*
KO1 WT1 KO2 WT2   KO1 WT1 KO2 WT2

← WT(~2500bp)
← KO(~2000bp)

Western Blot

dKO1 WT1 dKO2 WT2

α-ZNF587 & ZNF417

α-GAPDH

Fig. S8. Strategy and validation of *ZNF417/ZNF587* KO cell lines.

(A) Fluorescent images (lower panel) and histogram (upper panel) for viral titers of HIV-1 pseudo-virus variants containing PBS-Pro, PBS-Phe, or PBS-Lys3 mutant. The unaltered HIV-1 pseudo-virus has a PBS-Lys3 site.

(B) Strategy to knockout *ZNF417/ZNF587* using CRISPR/Cas9.

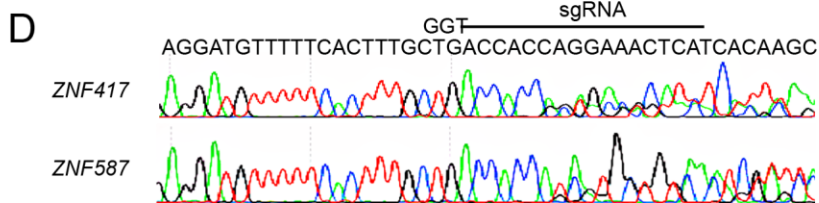(C) Genotyping PCR (upper panels) and Western blot (lower panels) analysis to confirm deletion of *ZNF417* and *ZNF587* alleles and loss of ZNF417 and ZNF587 proteins in HEK293T cells.

(D) Genotyping PCR (upper panels) and Western blot (lower panels) analysis to confirm deletion of *ZNF417/ZNF587* alleles in Jurkat cells.
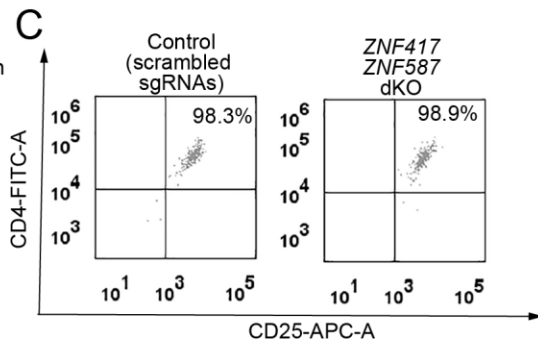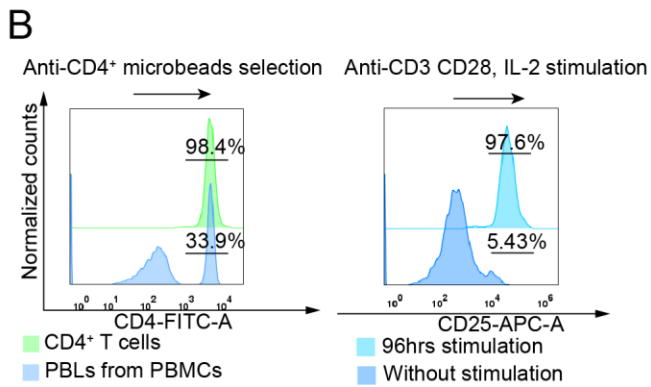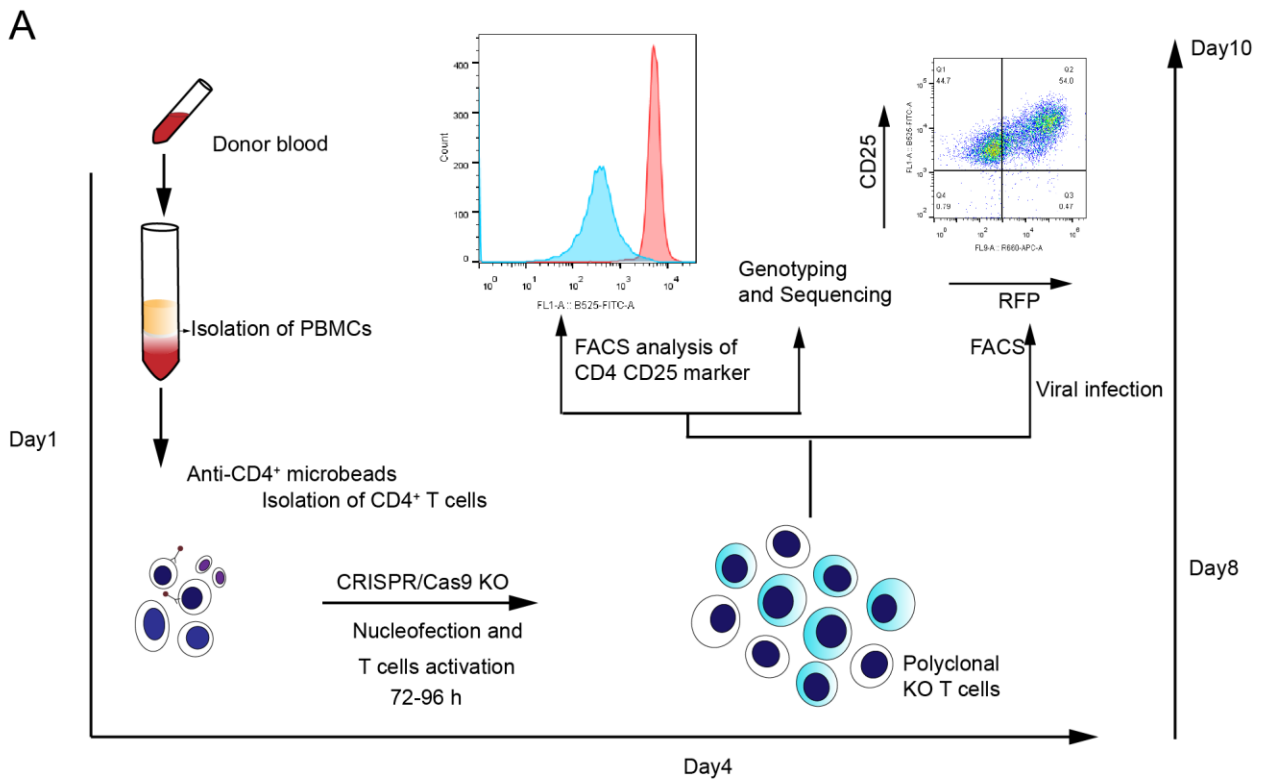
A

Day1

Donor blood

-Isolation of PBMCs

Anti-CD4⁺ microbeads
Isolation of CD4⁺ T cells

CRISPR/Cas9 KO

Nucleofection and
T cells activation
72-96 h

FACS analysis of
CD4 CD25 marker

Genotyping
and Sequencing

FACS

Viral infection

Polyclonal
KO T cells

CD25

RFP

Day10

Day8

Day4

B

Anti-CD4⁺ microbeads selection

Normalized counts

98.4%

33.9%

CD4-FITC-A

CD4⁺ T cells
PBLs from PBMCs

Anti-CD3 CD28, IL-2 stimulation

97.6%

5.43%

CD25-APC-A

96hrs stimulation
Without stimulation

C

Control
(scrambled
sgRNAs)

98.3%

CD4-FITC-A

ZNF417
ZNF587
dKO

98.9%

CD25-APC-A

D

GGT   sgRNA

AGGATGTTTTTCACTTTGCTGACCACCAGGAAACTCATCACAAGC

ZNF417

ZNF587

Fig. S9. ZNF417 and ZNF587 restrict HIV-1 infectivity in human primary CD4⁺ T cells.

(A) Schematic overview of HIV-1 infection in human primary CD4⁺ T cells before and after *ZNF417/ZNF587* deletion.

(B) FACS analysis of human CD4⁺ T cells separated from peripheral blood followed by activation.

(C) FACS analysis of activated CD4⁺ T cells in *ZNF417 ZNF587* dKO and control human active CD4⁺ T cells.

(D) Sanger sequencing to confirm deletion of *ZNF417* and *ZNF587* in activated CD4⁺ T cells.
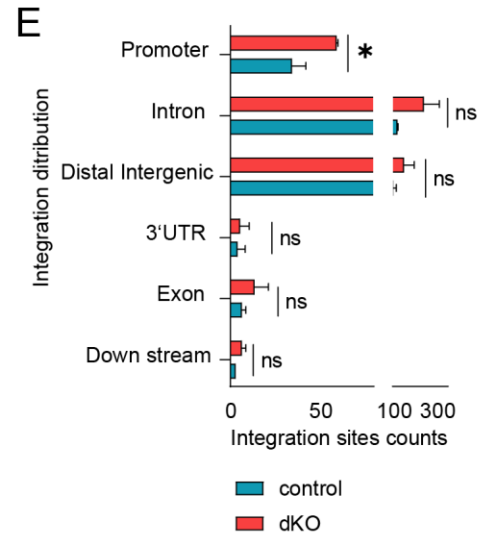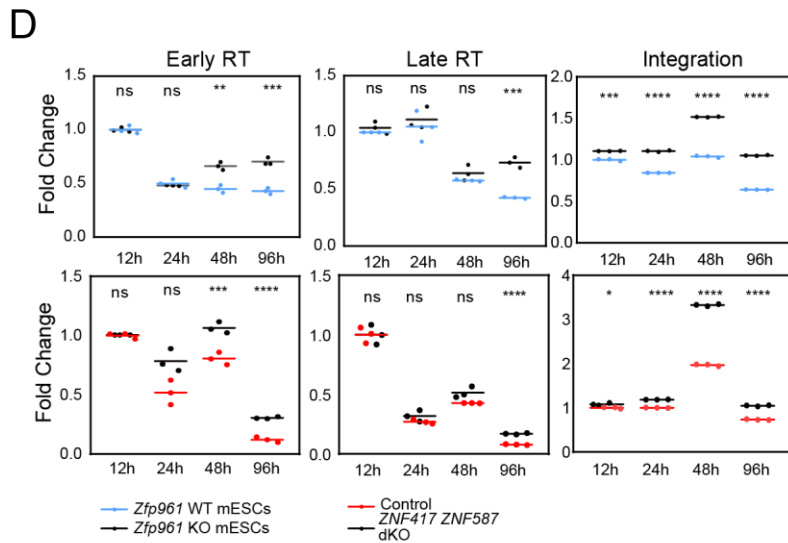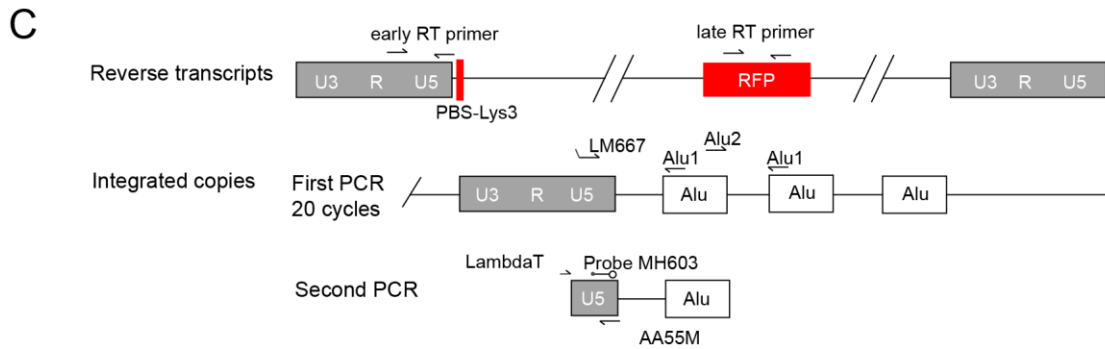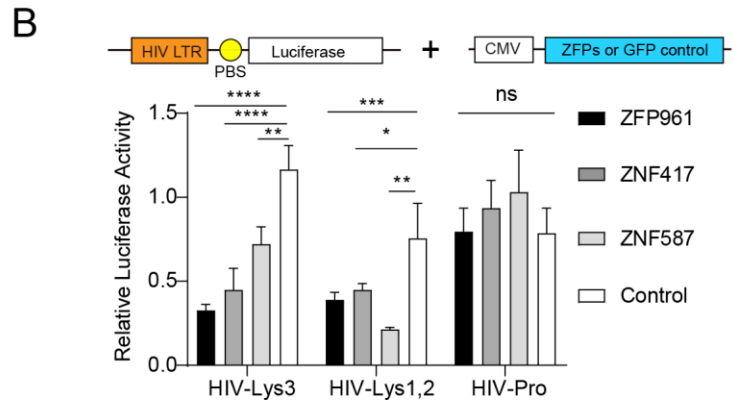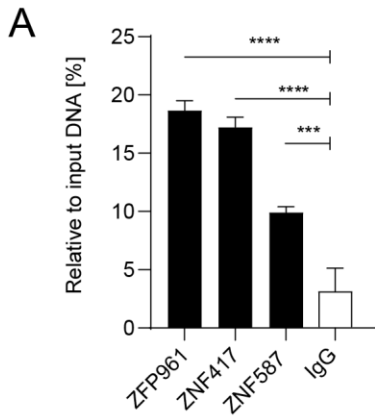
Fig. S10. ZFP961, ZNF417 and ZNF587 restrict HIV-1 transcription and integration.

(A) ChIP assays with antibodies against GFP and immunoglobulin G (IgG) as a control using GFP-fused ZFP961 ZNF417/ZNF587 and GFP vector overexpressing HEK293T cells, followed by qPCR using primers specific for HIV-1 LTR Nuc1. t test: Error bars indicate standard deviation. **** $p < 0.0001$, *** $p < 0.001$, and n = 3.

(B) Relative luciferase activity of HEK293T cells overexpressing ZNF417, ZNF587, and ZFP961 and an HIV-1 LTR-driven luciferase plasmid containing PBS-Lys1,2, PBS-Lys3, and PBS-Pro. t test: Error bars indicate standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns $p > 0.05$, and n = 3.

(C) Schematic of viral DNA copy assay for reverse transcription and integration processes based on viral plasmid structures. Early reverse transcribed HIV-1 DNA quantifications (early RT) were performed by real-time PCR using primers annealing to the R and the U5 regions of the 5'LTR. Total HIV-1 DNA quantification (late RT) in cells carrying pSicoR-RFP or pNL4.3-RFP lentiviral DNAs were performed using primers specific for the RFP gene. Detection of integrated proviruses was accomplished by amplification with primers complementary to the HIV-1 LTR and chromosomal Alu repeats. Specific PCR primers are labeled.

(D) Dot charts of viral DNA copy numbers in *Zfp961* WT and KO mESCs (upper panels), or *ZNF417 ZNF587* dKO and control HEK293T cells (lower panels) from 12 to 96 hours post-infection with pNL4.3-RFP pseudo-virus. t test: Error bars indicate standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns $p > 0.05$, and n = 3.

(E) Histograms showing the integration feature distributions of HIV-1 pseudo-virus integration sites in control and *ZNF417 ZNF587* dKO HEK293T cells. t test: Error bars indicate standard deviation. * $p < 0.05$, ns $p > 0.05$, and n = 2.

Dataset S1- RNA-seq of *Zfp961-GFP^{flox/flox}* mESCs.

We performed RNA−seq on *Zfp961-GFP^{flox/flox}* mESCs before and after 4OHT treatment. We used three independent replicates for each group and the RNA−seq read counts of 4OHT-treated mESCs (KO sample_2) and EtOH-treated (WT sample_1) were normalized (FPKM) and averaged (value_1 and 2). Log2 fold_changes, p values and q values show the test differences in the summed FPKM of transcripts sharing each gene_id (status OK).

Dataset S2- ZNF ChIP peaks overlap with PBS sites.

We compared PBS-containing ERVs regions with published ZNFs ChIP−seq peaks. All ERV regions are displayed in sheet1. BED format information of ZNF peaks overlapped with different PBS types are shown separately.

Dataset S3- RNA-seq of *ZNF417 ZNF587* dKO and control K562 cells.

We performed RNA−seq on *ZNF417* and *ZNF587* dKO and control K562 cells. The RNA−seq read counts of *ZNF417* and *ZNF587* dKO cells (ko sample_2) and scrambled sgRNAs treated control K562 cells (wt sample_1) were normalized (FPKM) and averaged (value_1 and 2). Log2 fold_changes, p values and q values show the test differences in the summed FPKM of transcripts sharing each gene_id (status OK).

Dataset S4- HIV-1 integration sites in *ZNF417 ZNF587* dKO and control HEK293T cells.

We performed high-throughput sequencing of HIV-1 integration sites in *ZNF417 ZNF587* dKO and control HEK293T cells. The integration sites and the nearest genes close to them are displayed.

Dataset S5- Classification of ERVs by PBS type.

We sorted representative PBS-containing ERVs and categorized them by PBS type. Potential binding ZFPs are also listed.

Dataset S6- PCR primers and promoters mentioned in this paper.

The PCR primers and promoters of luciferase reporters are listed in this Dataset.

References for SI

1. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).
2. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-515 (2010).
3. F. Ramírez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160-W165 (2016).
4. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
5. Y. Zhang *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
6. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
7. L. Vargiu *et al.*, Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **13**, 7 (2016).
8. J. F. Hultquist *et al.*, CRISPR-Cas9 genome engineering of primary CD4(+) T cells for the interrogation of HIV-host factor interactions. *Nature protocols* **14**, 1-27 (2019).
9. F. Maldarelli *et al.*, HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179-183 (2014).