SUPPLEMENTARY DATA

Zinc-finger-nucleases mediate specific and efficient excision of HIV-1 proviral DNA from infected and latently infected human T cells

Xiying Qu[1]**, Pengfei Wang[1]**, Donglin Ding[1], Lin Li[1], Haibo Wang[2], Li Ma[2], Xin Zhou[1], Shaohui Liu[1], Shiguan Lin[1], Xiaohui Wang[1], Gongmin Zhang[1], Sijie Liu[1], Lin Liu[1], Jianhua Wang[2], Feng Zhang[1], Daru Lu[1] and Huanzhang Zhu[1]*

[1]State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China.
[2]Key Laboratory of Molecular Virology and Immunology, Institute Pasteur of Shanghai, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200025, China.

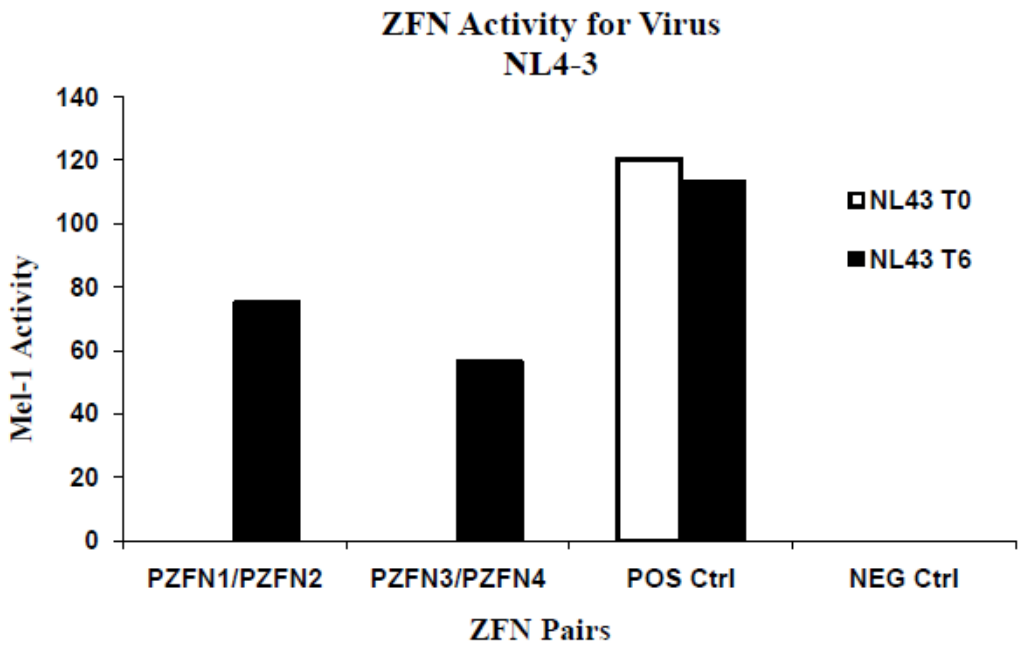
* To whom correspondence should be addressed. Tel: +86-21-60746711; Fax: +86-21-60746711;
Email: hzzhu@fudan.edu.cn

** These authors contributed equally to this work.

Table S1. List of the target sites of originally designed ZFNs.

design #	Target sites on HIV-1 LTR
1	CACACAAGGCTACTTCCCnnATTGGCAGAACTACACACCAG
2	CACACAAGGCTACTTCCCTGATnnGCAGAACTACACACCAG
3	GGCTACTTCCCTGATTGGnnGAACTACACACCAGGGCCAG
4	GGCTACTTCCCTGATTGGnnGAACTACACACCAGGGCCAG
5	CGAGCCCTCAGATGCTAnnTATAAGCAGCTGCT
6	CATATAAGCAGCTGnnTTTTGCCTGTACTGGGT
7	CCAGATCTGAGCCTGGGAGCnCTCTGGCTAACTAG
8	CCAGATCTGAGCCTGGGAGCnCTCTGGCTAACTAGGGA
9	CCACTGCTTAAGCCTCAATAnnGCTTGCCTTGAGTG
10	TGCTTAAGCCTCAATAAnGCTTGCCTTGAGTG
11	CTAGAGATCCCTCAGACCCTnTTAGTCAGTGTGGA
12	GGCTACTTCCCTGAnnnGCAGAACTACACACCAG
13	CACACAAGGCTACTTCCCTGAnnnGCAGAACTACACACCAG
14	GTGGCGAGCCCTCAGATGCTnnnTATAAGCAGCTGCT
15	GGAACCCACTGCTTAAGnnnCAATAAAGCTTGCCTTGAGTG
16	CTAGAGATCCCTCAGACCnnnTTAGTCAGTGTGGA

Figure S1. ZFN activity as measured by the yeast MEL-1 reporter assay.



ZFN activity as measured by the yeast Mel-1 reporter assay (Doyon et al., Nat Biotechnol, 2008, 26(6): 702). ZFN cleavage activity was measured before induction (0h, white bars) and after induction of ZFN expression (6h, black bars). Mel-1 levels positively correlate with ZFN ability to create double strand breaks at the desired target site. ZFNs that show >50% signal relative to the positive control ZFN after induction (6h) are regarded as useful for genome editing experiments.

Figure S2. Annotated ZFN-LTR amino acid sequences.

Full amino acid sequences of ZFN-LTR-L **(A)** and ZFN-LTR-R **(B)**. Both proteins contain one SV40 Nuclear Localization Signal (NLS, purple) and one *Fok*I domain (green). The DNA binding modules of ZFN-LTR-L and ZFN-LTR-R are represented in red.

A. ZFN-LTR-L

3FLAG NLS

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPFQCRICMRNFSRSDH 60

Module L6:TGG Module L5:ATC Module L4:CAG

LSTHIRTHTGEKPFACDICGRKFADSSARKKHTKIHTGSQKPFQCRICMRNFSRSDALSE 120

Module L3:GCT Module L2:CAG

HIRTHTGEKPFACDICGKKFATSSSRKKHTKIHTGSQKPFQCRICMRNFSRSDNLSQHIR 180

Module L1:TCC

THTGEKPFACDICGRKFAASNDRKKHTKIHLRGSQLVKSELEEKKSELRHKLKYVPHEYI 240

*Fok*I

ELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTK 300

AYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYK 360

AQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFRS** 412

B. ZFN-LTR-R

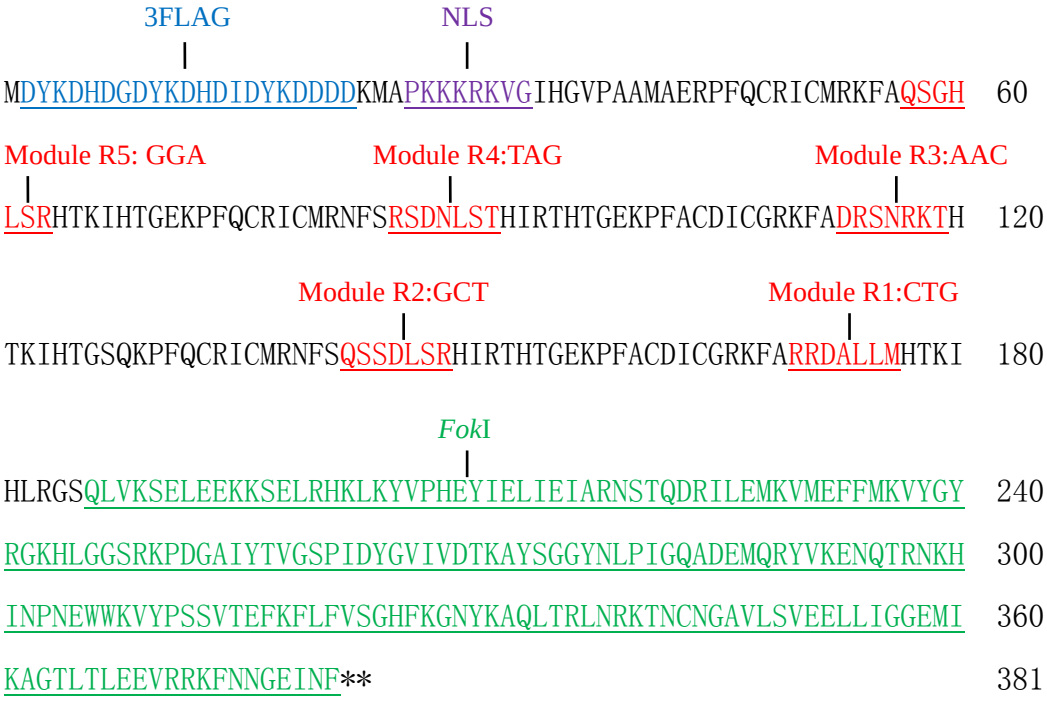

3FLAG
NLS
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPFQCRICMRKFAQSGH 60
Module R5: GGA
Module R4:TAG
Module R3:AAC
LSRHTKIHTGEKPFQCRICMRNFSRSDNLSTHIRTHTGEKPFACDICGRKFADRSNRKTH 120
Module R2:GCT
Module R1:CTG
TKIHTGSQKPFQCRICMRNFSQSSDLSRHIRTHTGEKPFACDICGRKFARRDALLMHTKI 180
FokI
HLRGSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGY 240
RGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKH 300
INPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMI 360
KAGTLTLEEVRRKFNNGEINF** 381

Figure S3. Conservation analysis of ZFN-LTR binding sequence and LoxLTR sequence recognized by Tre-recombinase in different HIV-1 subtypes. (**A**) The ZFN-LTR binding sequence, located at 471-508 bp within the HXB2 reference isolate (GenBank accession number K03455), and (**B**) the LoxLTR sequence, located at 193-226 of the HXB2 reference isolate, were aligned with all HIV-1 genome sequences in the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/) using a web alignment tool (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html). Then the alignments were used to highlight mismatches using the Highlighter for Nucleotide Sequences v2.1.1 online (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/HIGHLIGHT_XYPLOT/highlighter.html). Mismatches are represented in different colors: A-Green, T-Red, G-Orange, C-Light blue, Gaps-Gray. For ZFN-LTR binding sequences, the total number of analyzed ZFN-LTR binding sequences was 344 and the average similarity was 0.937, while for LoxLTR, the total number of analyzed sequences was 269 and the average similarity was 0.775.


A
CCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGA
ZFN-LTR binding sites (HXB2.471-508)
A:Green, T:Red, G:Orange, C:Light blue, IUPAC:Dark blue, Gaps:Gray
0
10
20
30
40
50
Base number

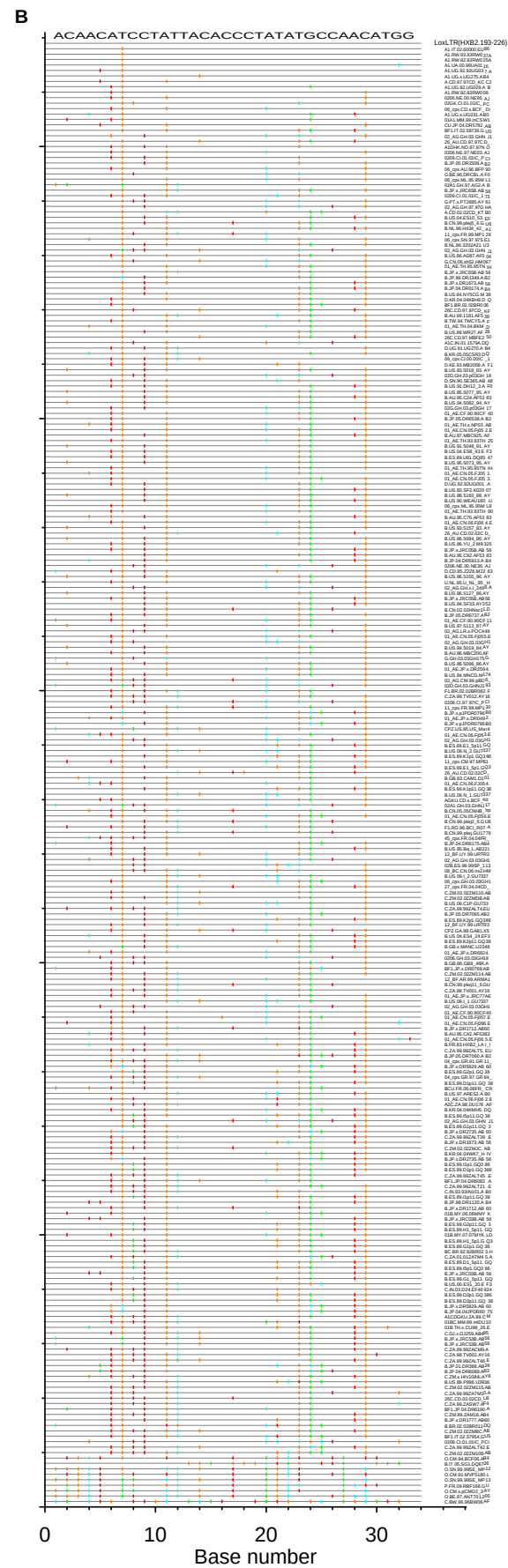
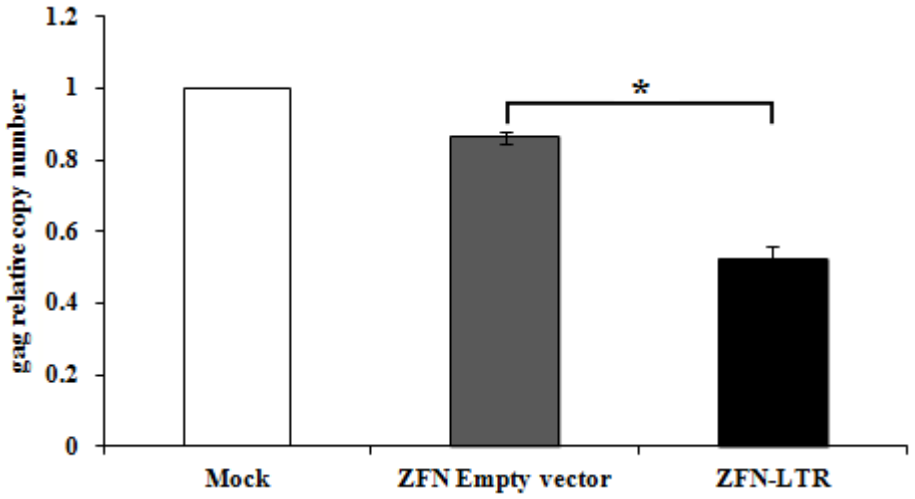
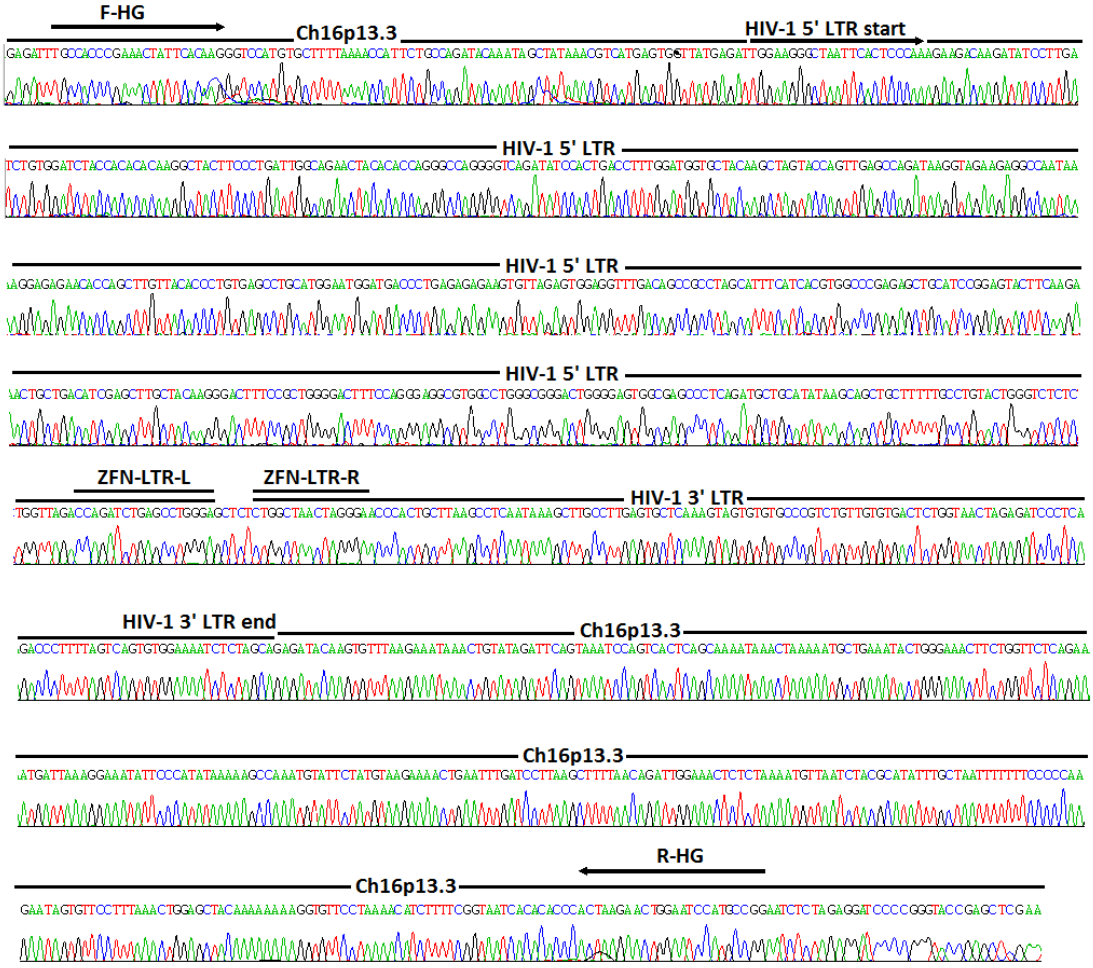

B
ACAACATCCTATTACACCCTATATGCCAACATGG
LoxLTR(HXB2:193-226)
0
10
20
30
Base number

Figure S4. ZFN-LTR induced excision of integrated proviral HIV-1 genome in HIV-based lentiviral infected 293T cells.



HIV-based-lentivirus-infected 293T cells were not transfected (mock) or transfected with ZFN empty vector or ZFN-LTR plasmid. At 3 days post-transfection, genomic DNA was isolated and subjected to quantitative real-time PCR using gene-specific primers for HIV-1 *gag* and human *β-globin*. The relative copy numbers of *gag* were calculated based on the standard curve obtained by serial dilution [10 to 160ng] of an infected cell DNA on the same plate. Normalization was carried out by division of *gag* gene amplicons in mock group. Data are representative of three independent experiments, and error bars represent standard errors (SD). *P< 0.05, **P< 0.01, ***P< 0.001; paired t test.

Figure S5. DNA sequences of PCR products from latently infected C11cells treated with ZFN-LTR.



Latently infected C11cells were transfected with ZFN-LTR, 3 days after transfection, genomic DNA was extracted. PCR was performed using primers located at the two ends of the integrated sites in Ch16p13.3. The PCR products corresponding to the 9.8 kb genomic DNA deletions were cloned and sequenced. ZFN-LTR-L and ZFN-LTR-R indicate ZFN-LTR target sites.