OMTN, Volume *35*

Supplemental information

Engineering miniature CRISPR-Cas Un1Cas12f1 for efficient base editing

Yueer Hu, Linxiao Han, Qiqin Mo, Zengming Du, Wei Jiang, Xia Wu, Jing Zheng, Xiao Xiao, Yadong Sun, and Hanhui Ma

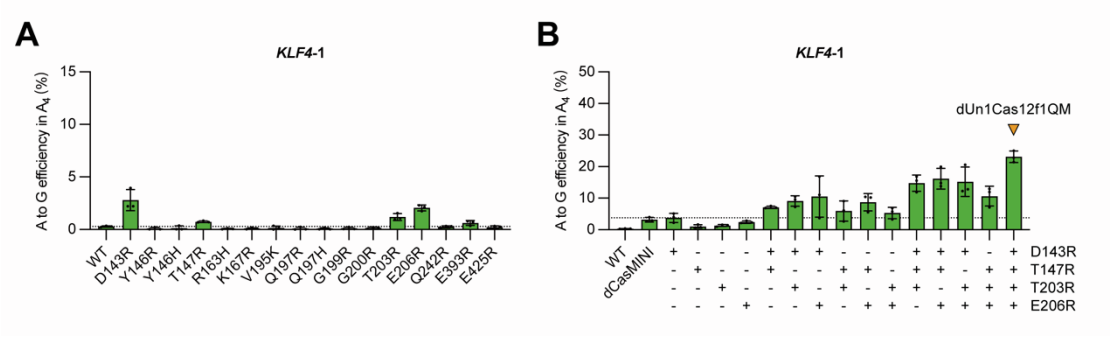

Figure S1. Comparison of editing efficiencies of dUn1Cas12f1 variants-mediated ABEs at *KLF4*-1 site

(A) Comparison of A-to-G editing efficiencies mediated by dUn1Cas12f1 variants at *KLF4*-1 site in HEK293FT cells. The most efficient position A_4 of the *KLF4*-1 site is chosen for the comparison. WT is the wild-type dUn1Cas12f1.

(B) Comparison of A-to-G editing efficiencies mediated by the combinations of dUn1Cas12f1 variants D143R, T147R, T203R or E206R at *KLF4*-1 site. dCasMINI is the dUn1Cas12f1 variant with D143R/T147R/K330R/E528R mutations. The orange arrow refers to the variant dUn1Cas12f1QM with D143R/T147R/T203R/E206R mutations.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

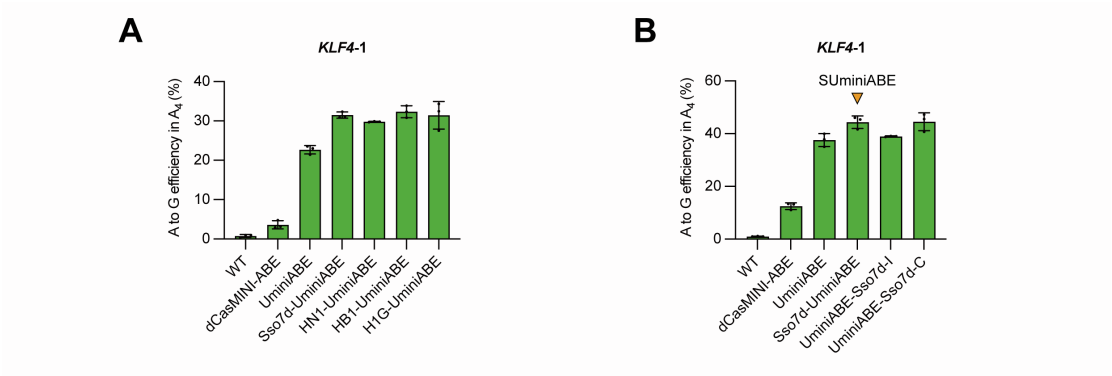

Figure S2. Comparison of editing efficiencies of different DBD-UminiABEs at *KLF4*-1 site

(A) Comparison of A-to-G editing efficiencies induced by dUn1Cas12f1-ABE, dCasMINI-ABE, UminiABE and different DBD-UminiABEs at *KLF4*-1 site. DBDs contain Sso7d, HMGN1 (HN1), HMGB1 box A (HB1) and histone H1 central globular domain (H1G). The most efficient position A_4 of the *KLF4*-1 site is chosen for the comparison.

(B) Comparison of A-to-G editing efficiencies induced by dUn1Cas12f1-ABE, dCasMINI-ABE, UminiABE, Sso7d-UminiABE, UminiABE-Sso7d-I and UminiABE-Sso7d-C at *KLF4*-1 site. The orange arrow refers to SUminiABE, which is Sso7d-UminiABE containing original sgRNA.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

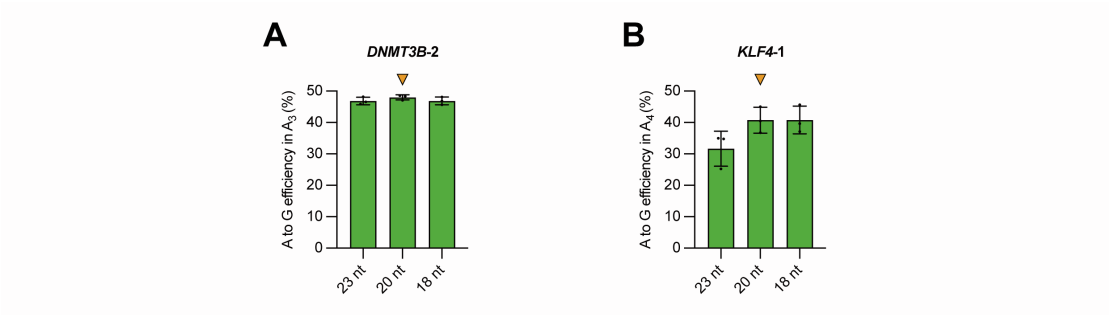

Figure S3. The effect of spacer lengths of sgRNA on the editing efficiencies of Sso7d-UminiABEs at *DNMT3B*-2 and *KLF4*-1 sites

(A) Comparison of A-to-G editing efficiencies induced by Sso7d-UminiABE containing 23 nt-, 20 nt- or 18 nt-spacer original sgRNA at *DNMT3B*-2 site. The most efficient position A_3 of the *DNMT3B*-2 site is chosen for the comparison. The orange arrow refers to 20 nt as spacer length for further study.

(B) Comparison of A-to-G editing efficiencies induced by Sso7d-UminiABE containing 23 nt-, 20 nt- or 18 nt-spacer original sgRNA at *KLF4*-1 site. The most efficient position A_4 of the *KLF4*-1 site is chosen for the comparison. The orange arrow refers to 20 nt as spacer length for further study.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

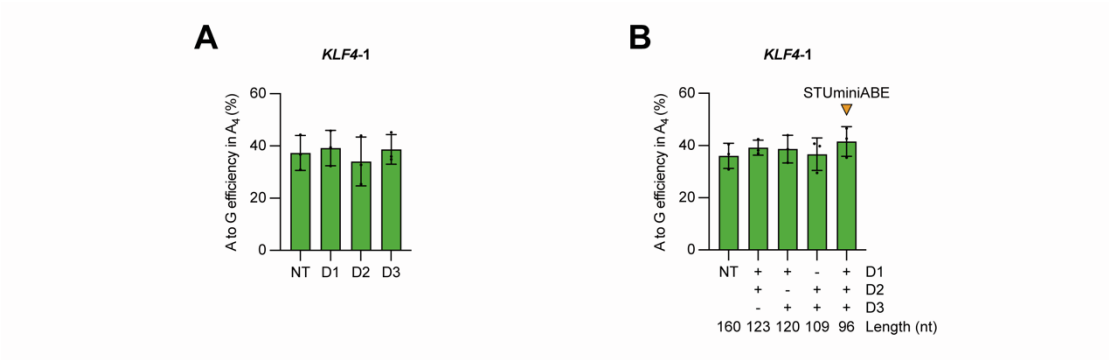

Figure S4. The effect of truncated sgRNA on the editing efficiencies of Sso7d-UminiABEs at *KLF4*-1 site

(A) Comparison of A-to-G editing efficiencies at *KLF4*-1 site induced by Sso7d-UminiABEs containing original sgRNA (NT = natural tracrRNA, spacer length = 20 nt) or truncated sgRNA variants.

(B) Comparison of A-to-G editing efficiencies at *KLF4*-1 site induced by Sso7d-UminiABEs containing original sgRNA (NT = natural tracrRNA, spacer length = 20 nt) or the combinations of truncated sgRNA variants D1, D2 or D3. The orange arrow refers to STUminiABE, which is Sso7d-UminiABE containing truncated sgRNA-D1/D2/D3.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

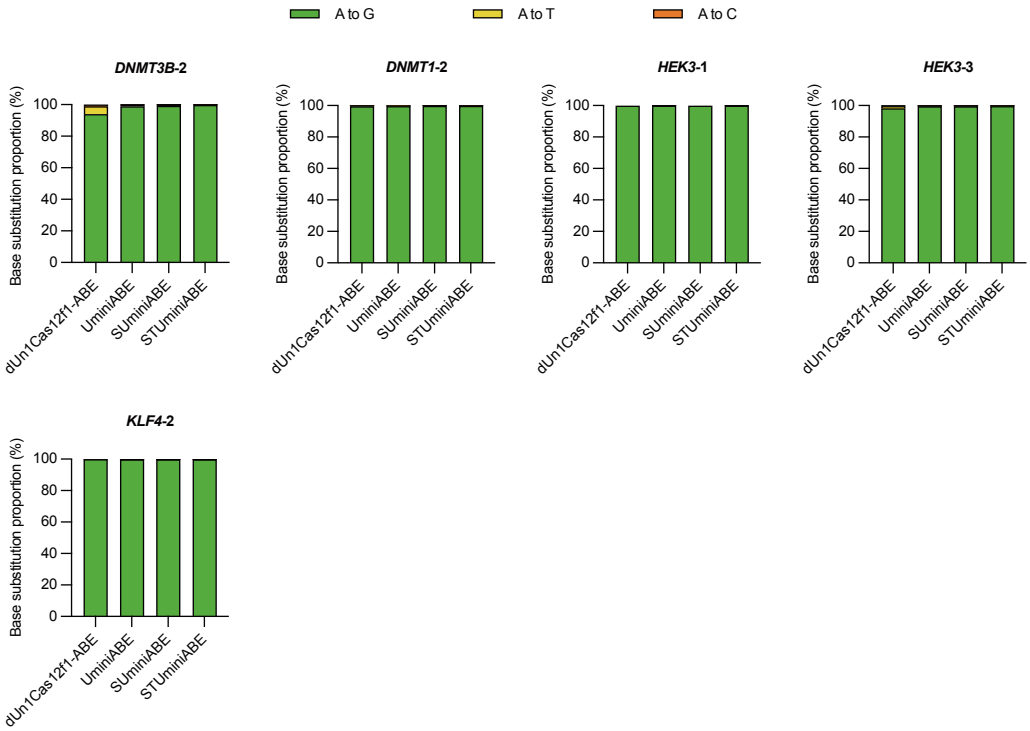

Figure S5. Comparison of the proportion of base substitution induced by dUn1Cas12f1-ABE, UminiABE, SUminiABE or STUminiABE

Comparison of the A-to-G editing purity induced by dUn1Cas12f1-ABE, UminiABE, SUminiABE or STUminiABE at *DNMT3B*-2, *DNMT1*-2, *HEK3*-1, *HEK3*-3 or *KLF4*-2 site. The most efficient position of each site is chosen for the comparison. All values represent means, n = 3 independent biological replicates.

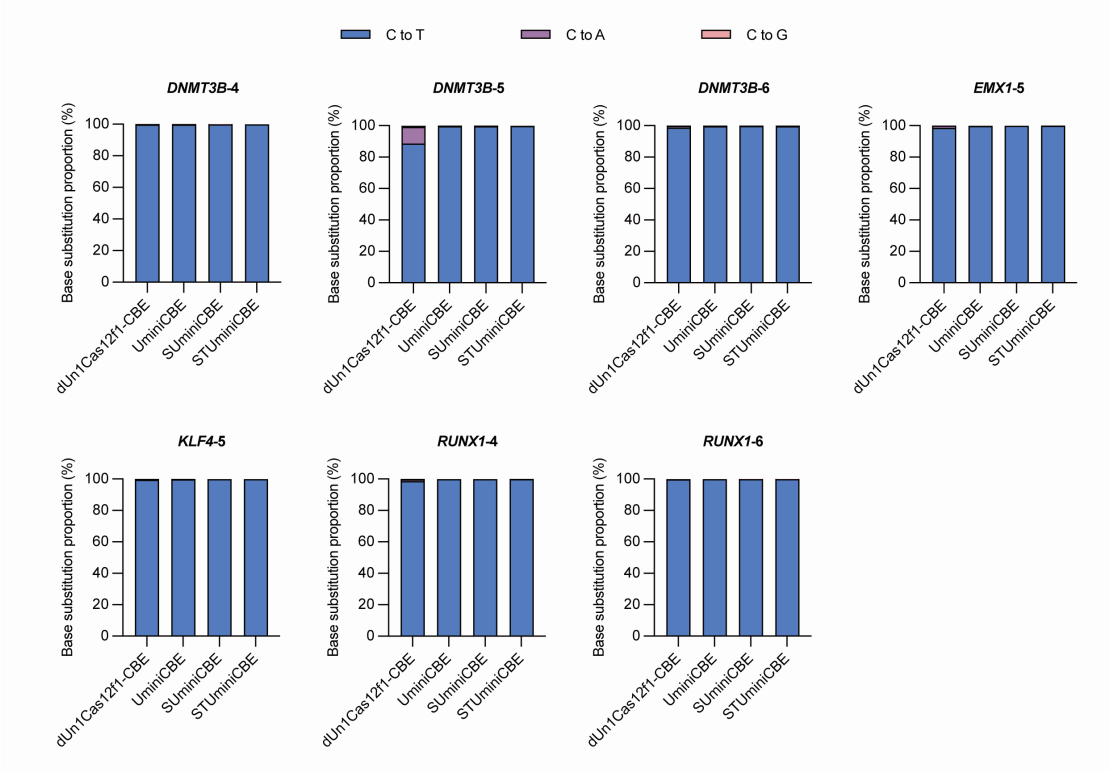

Figure S6. Comparison of the proportion of base substitution induced by dUn1Cas12f1-CBE, UminiCBE, SUminiCBE or STUminiCBE

Comparison of the C-to-T editing purity induced by dUn1Cas12f1-CBE, UminiCBE, SUminiCBE or STUminiCBE at *DNMT3B*-4, *DNMT3B*-5, *DNMT3B*-6, *EMX1*-5, *KLF4*-5, *RUNX1*-4 or *RUNX1*-6 site. The most efficient position of each site is chosen for the comparison. All values represent means, n = 3 independent biological replicates.

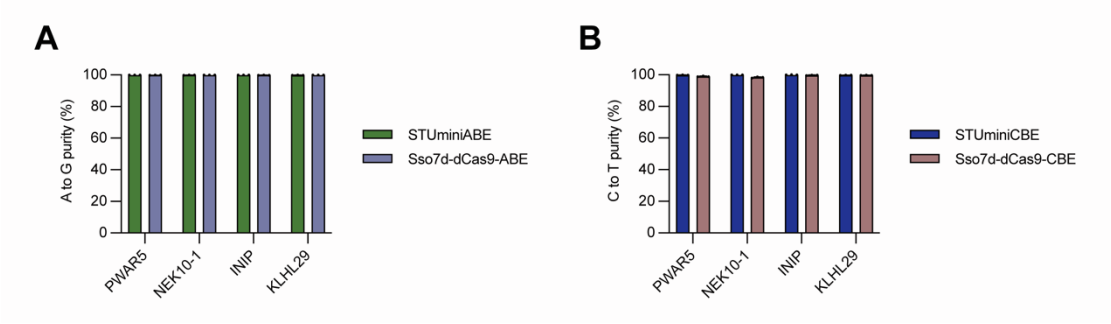

Figure S7. Comparison of the proportion of base substitution induced by STUminiBEs or dCas9-derived BEs

(A) Comparison of the A-to-G editing purity induced by STUminiABE or Sso7d-dCas9-ABE at *PWAR5*, *NEK10*-1, *INIP* or *KLHL29* site. The most efficient position of each site is chosen for the comparison.

(B) Comparison of the C-to-T editing purity induced by STUminiCBE or Sso7d-dCas9-CBE at *PWAR5*, *NEK10*-1, *INIP* or *KLHL29* site. The most efficient position of each site is chosen for the comparison.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

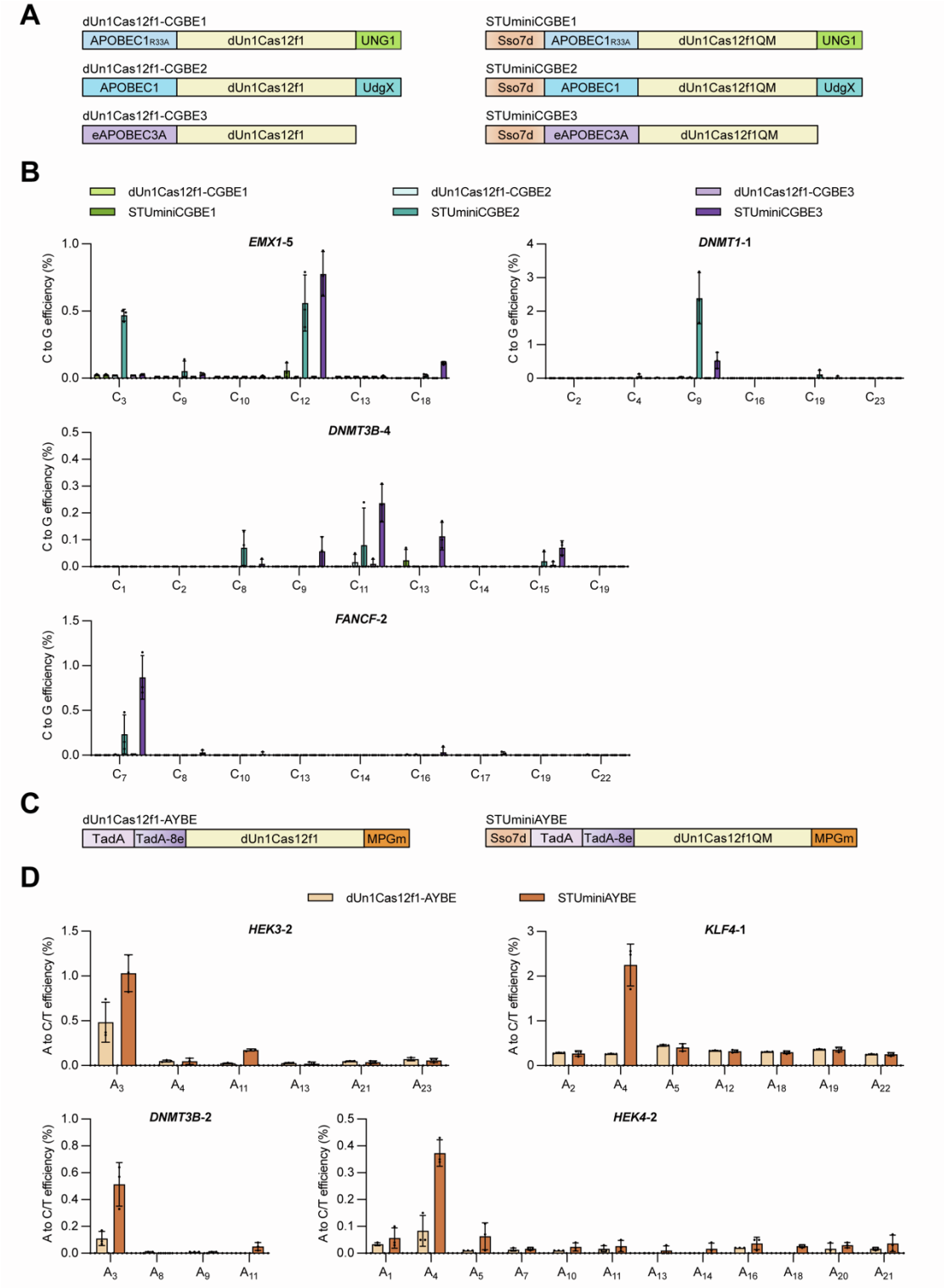

Figure S8. Development of dUn1Cas12f1-mediated base editors STUminiCGBEs and STUminiAYBE

(A) Schematic of dUn1Cas12f1-mediated C-to-G base editors. dUn1Cas12f1-CGBE1 and STUminiCGBE1 both contain APOBEC1$_{R33A}$ and UNG1; dUn1Cas12f1-CGBE2

and STUminiCGBE2 both contain APOBEC1 and UdgX; dUn1Cas12f1-CGBE3 and STUminiCGBE3 both contain eAPOBEC3A. dUn1Cas12f1-CGBEs all contain dUn1Cas12f1 while STUminiCGBEs all contain dUn1Cas12f1QM and Sso7d that is fused to the N-terminal of cytosine deaminase. APOBEC1$_{R33A}$: rat APOBEC1$_{R33A}$; UNG1: uracil DNA N-glycosylase 1 from *S. cerevisiae*; APOBEC1: rat APOBEC1; UdgX: UNG ortholog from *Mycobacterium smegmatis*; eAPOBEC3A: human APOBEC3A$_{N57G}$. Original sgRNA or truncated sgRNA-D1/D2/D3 is not shown in the diagram.

(B) Comparison of the C-to-G editing efficiencies of dUn1Cas12f1-CGBEs and STUminiCGBEs at *EMX1*-5, *DNMT1*-1, *DNMT3B*-4 or *FANCF*-2 site in HEK293FT cells.

(C) Schematic of dUn1Cas12f1-mediated adenine transversion base editor. dUn1Cas12f1-AYBE consists of the adenine deaminase heterodimer TadA-TadA-8e, dUn1Cas12f1 and MPGm. STUminiAYBE consists of Sso7d, TadA-TadA-8e, dUn1Cas12f1QM and MPGm. MPGm: N-methylpurine DNA glycosylase mutant. Original sgRNA or truncated sgRNA-D1/D2/D3 is not shown in the diagram.

(D) Comparison of the A-to-Y (C/T) editing efficiencies of dUn1Cas12f1-AYBE and STUminiAYBE at *HEK3*-2, *KLF4*-1, *DNMT3B*-2 or *HEK4*-2 site in HEK293FT cells.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

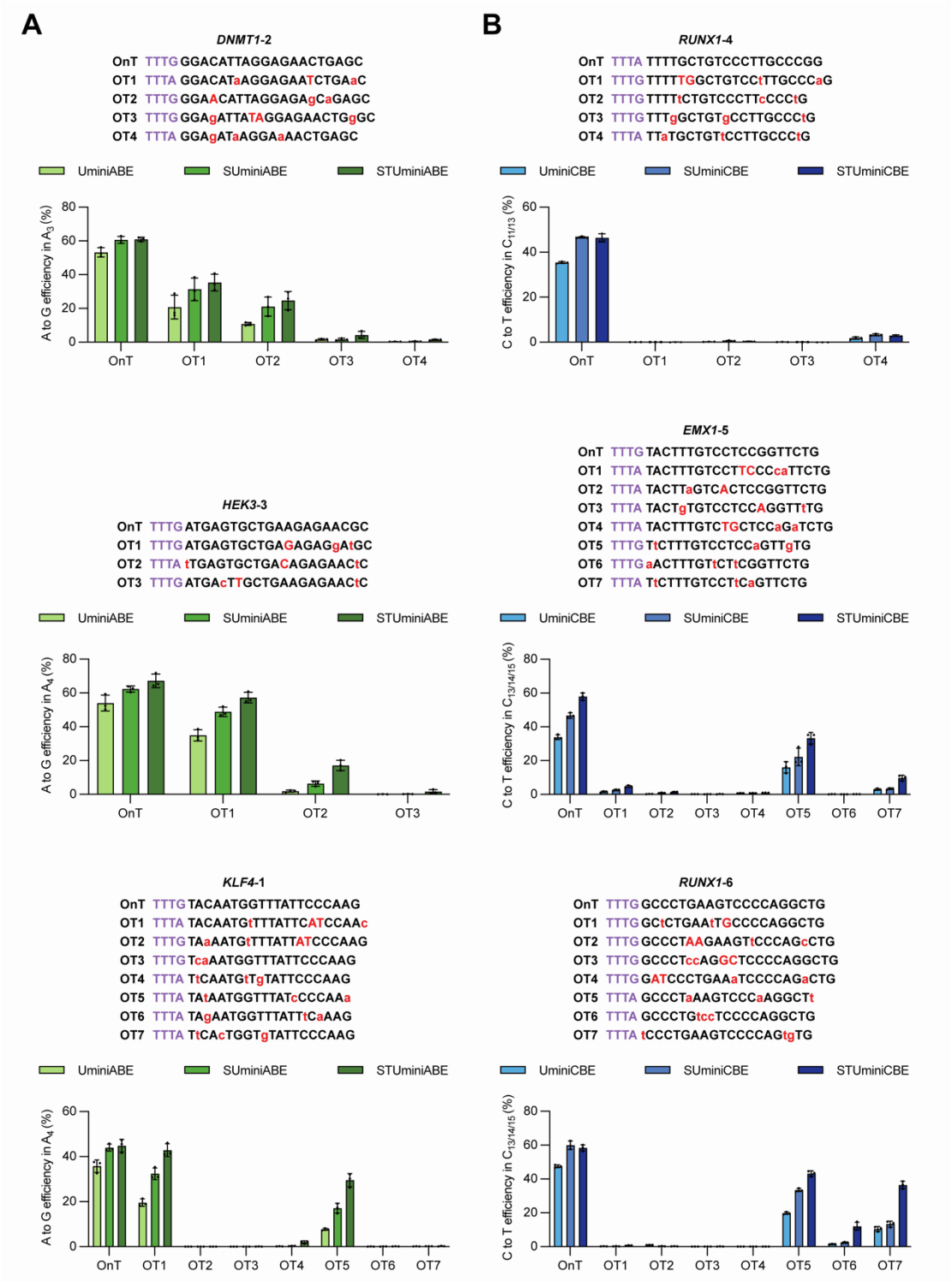

Figure S9. Off-target analysis of UminiBEs, SUminiBEs and STUminiBEs

(A) Comparison of on-target and off-target efficiency of UminiABE, SUminiABE or STUminiABE at *DNMT1*-2, *HEK3*-3 or *KLF4*-1 site in HEK293FT cells. The off-target loci are screened based on 20 nt spacer length. PAM sequences are in light

purple, the mismatches are shown as red lowercase letters while the bulges are shown as red capital letters. A_3 of the *DNMT1*-2 site, A_4 of the *HEK3*-3 site or A_4 of the *KLF4*-1 site at on-target site is chosen for the comparison.

(B) Comparison of on-target and off-target efficiency of UminiCBE, SUminiCBE or STUminiCBE at *RUNX1*-4, *EMX1*-5 or *RUNX1*-6 site in HEK293FT cells. The off-target loci are screened based on 20 nt spacer length. PAM sequences are in light purple, the mismatches are shown as red lowercase letters while the bulges are shown as red capital letters. C_{11} of the *RUNX1*-4 site, C_{13} of the *EMX1*-5 site or C_{13} of the *RUNX1*-6 site is chosen for the comparison. C_{13} of partial off-target sites corresponding to *RUNX1*-4 site, C_{14} or C_{15} of partial off-target sites corresponding to *EMX1*-5 site while C_{14} or C_{15} of partial off-target sites corresponding to *RUNX1*-6 site is chosen for the comparison due to the bulges.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

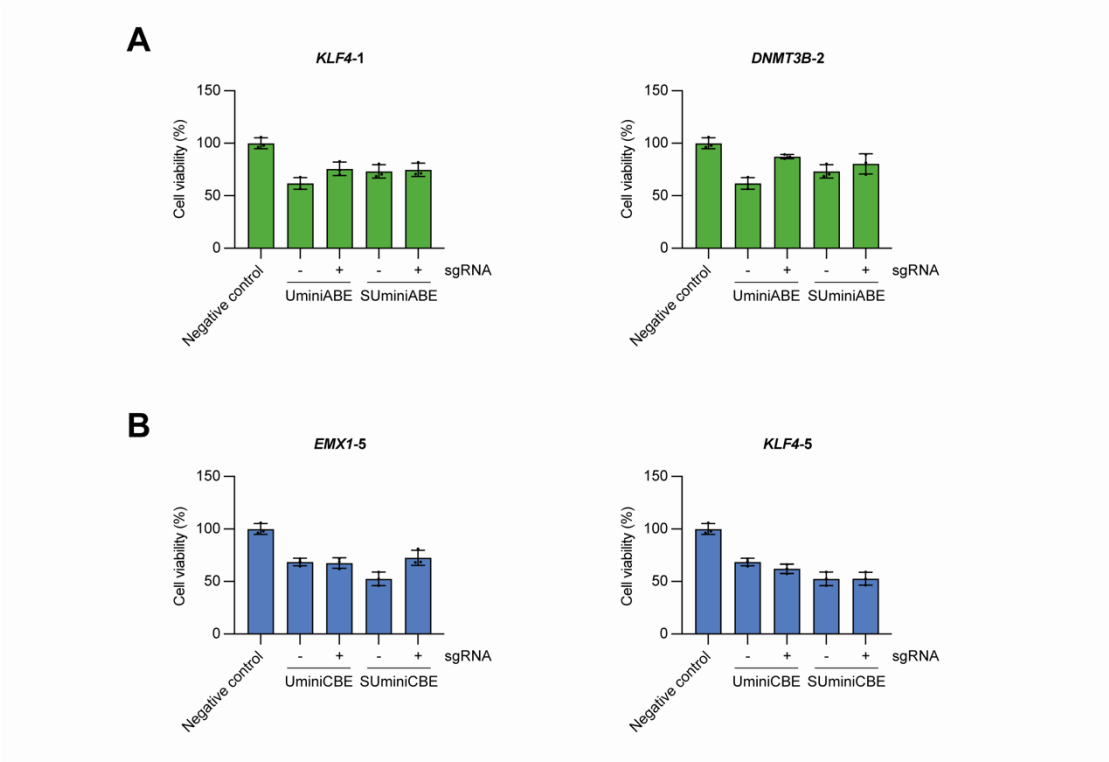

Figure S10. The effect of cell viability by fusing the non-specific dsDNA binding protein Sso7d to the N-terminal of deaminases in UminiBEs

(A) Comparison of the activity of HEK293FT cells edited by UminiABE or SUminiABE at *KLF4*-1 or *DNMT3B*-2 site. Negative control is the group of HEK293FT cells that are only transfected with the transfection reagent without plasmids.

(B) Comparison of the activity of HEK293FT cells edited by UminiCBE or SUminiCBE at *EMX1*-5 or *KLF4*-5 site. Negative control is the group of HEK293FT cells that are only transfected with the transfection reagent without plasmids.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

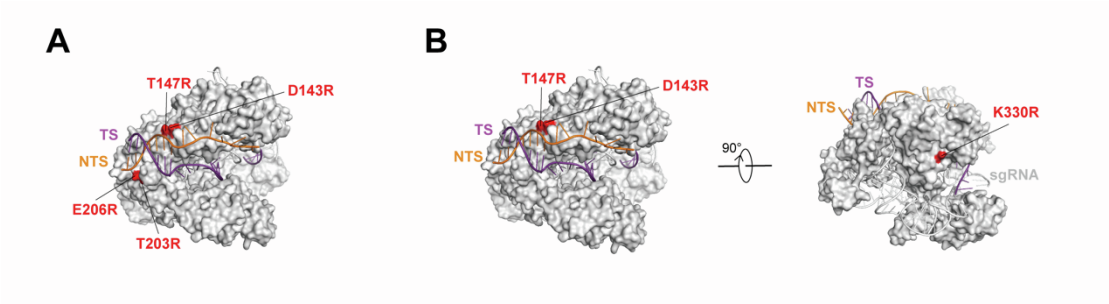

Figure S11. Comparison of point mutations of dUn1Cas12f1QM or dCasMINI fitting in Un1Cas12f1-sgRNA-dsDNA structure

(A) The point mutations (D143R/T147R/T203R/E206R in dUn1Cas12f1QM) of Un1Cas12f1 residues to enhance the interaction between Un1Cas12f1-sgRNA and target dsDNA (Un1Cas12f1 PDB: 7C7L). Un1Cas12f1 is in gray, target strand (TS) DNA is in purple, non-target strand (NTS) DNA is in orange and sgRNA is in white.

(B) The point mutations (D143R/T147R/K330R in dCasMINI) of Un1Cas12f1 residues fitting in the Un1Cas12f1-sgRNA-dsDNA structure (PDB: 7C7L) in two views with each subunit color coded as in **Figure S11A**. E528R, one of the point mutations in dCasMINI, is not shown due to the E528 residue cannot be represented in the structure.

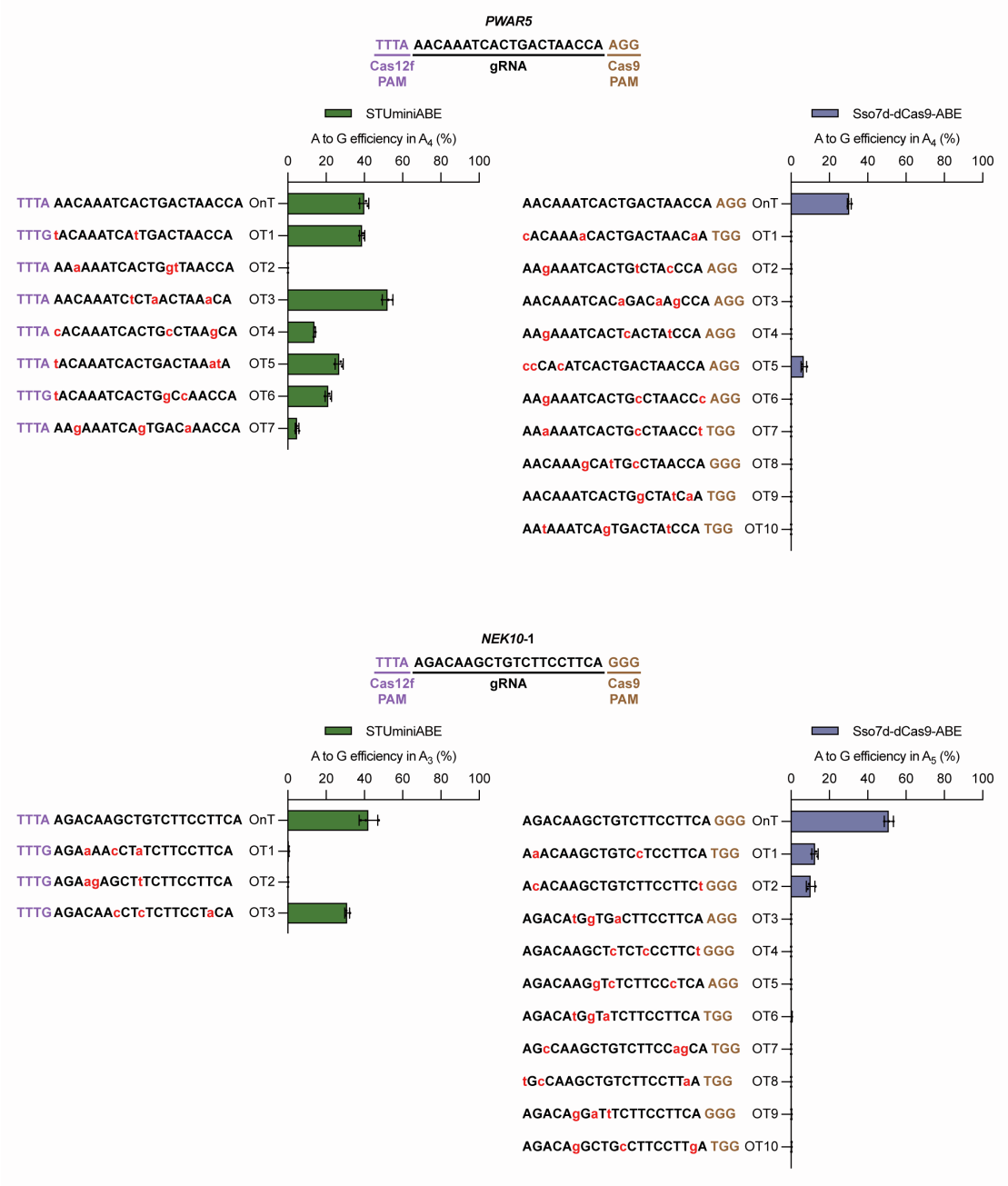

Figure S12. Off-target analysis of STUminiABE and Sso7d-dCas9-ABE

Comparison of on-target and off-target editing efficiency of STUminiABE (left panel) or Sso7d-dCas9-ABE (right panel) at *PWAR5* or *NEK10*-1 site in HEK293FT cells. 1~3 nt-mismatch with no bulge off-target sites were searched in the human genome using Cas-OFFinder. dUn1Cas12f1 PAM sequences are in light purple and dCas9 PAM sequences are in brown, the mismatches are shown as red lowercase letters. The

most efficient position of each on-target site induced by STUminiABE or Sso7d-dCas9-ABE is chosen for the comparison. All values and error bars represent means ± s.d., n = 3 independent biological replicates.

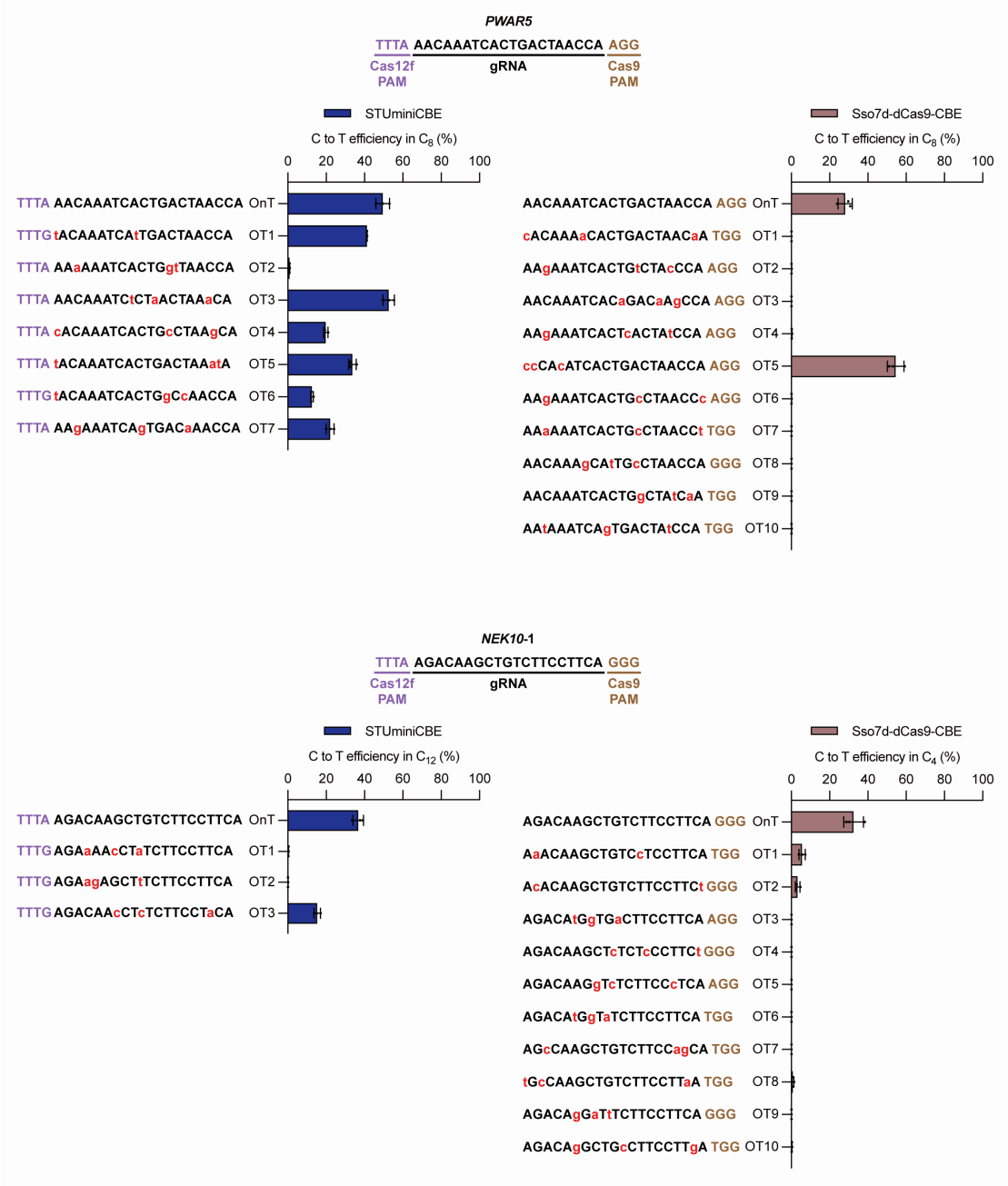

Figure S13. Off-target analysis of STUminiCBE and Sso7d-dCas9-CBE

Comparison of on-target and off-target editing efficiency of STUminiCBE (left panel) or Sso7d-dCas9-CBE (right panel) at *PWAR5* or *NEK10*-1 site in HEK293FT cells. 1~3 nt-mismatch with no bulge off-target sites were searched in the human genome using Cas-OFFinder. dUn1Cas12f1 PAM sequences are in light purple and dCas9 PAM sequences are in brown, the mismatches are shown as red lowercase letters. The

most efficient position of each on-target site induced by STUminiCBE or Sso7d-dCas9-CBE is chosen for the comparison. All values and error bars represent means ± s.d., n = 3 independent biological replicates.

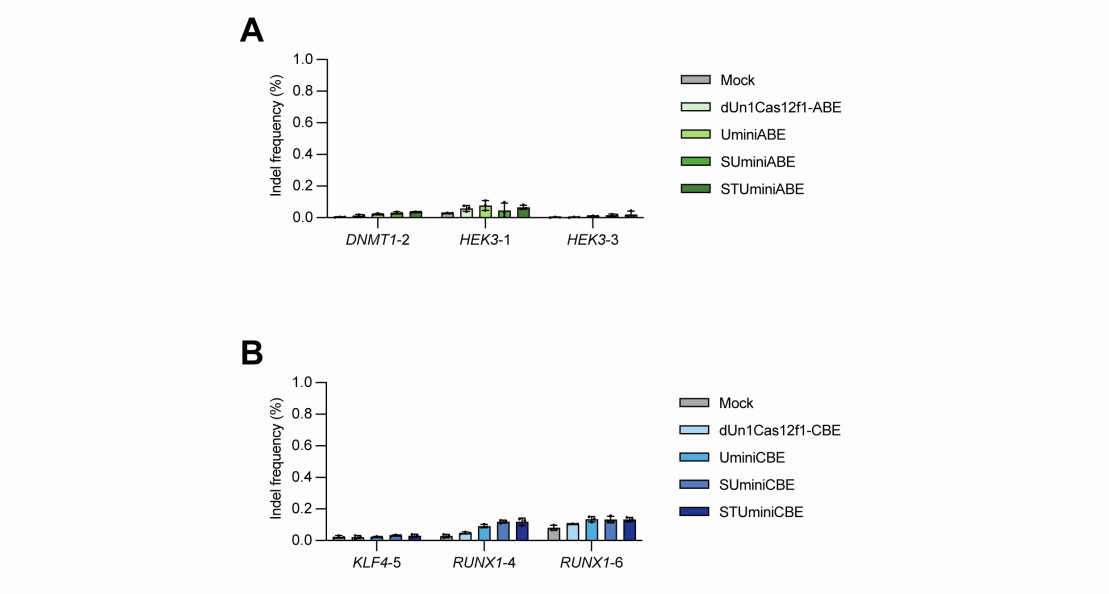

Figure S14. Indel analysis of dUn1Cas12f1-BEs, UminiBEs, SUminiBEs and STUminiBEs

(A) The indel frequencies of dUn1Cas12f1-ABE, UminiABE, SUminiABE and STUminiABE at *DNMT1*-2, *HEK3*-1 or *HEK3*-3 site in HEK293FT cells. Mock represents non-transfected control.

(B) The indel frequencies of dUn1Cas12f1-CBE, UminiCBE, SUminiCBE and STUminiCBE at *KLF4*-5, *RUNX1*-4 or *RUNX1*-6 site in HEK293FT cells. Mock represents non-transfected control.

All values and error bars represent means ± s.d., n = 3 independent biological replicates.

Table S1. Spacer and PAM sequences of Un1Cas12f1 sgRNA target sites

Targets	Un1Cas12f1 sgRNA spacer sequences	PAM	Figure
23nt-*DNMT3B*-2	CCACGTGAATACTGTGGTTTTTC	TTTA	Fig.1C,1D,1F,1G,S3A,S5,S8D,S10A
20nt-*DNMT3B*-2	CCACGTGAATACTGTGGTTT	TTTA	Fig.1I,1J,S3A,S5,S8D
18nt-*DNMT3B*-2	CCACGTGAATACTGTGGT	TTTA	Fig.S3A
23nt-*DNMT1*-2	GGACATTAGGAGAACTGAGCCTT	TTTG	Fig.1K,S5,S9A,S14A
20nt-*DNMT*1-2	GGACATTAGGAGAACTGAGC	TTTG	Fig.1K,S5,S9A,S14A
23nt-*HEK3*-1	AGTATTAGTCTCAGCGAAATGGA	TTTA	Fig.1K,S5,S14A
20nt-*HEK3*-1	AGTATTAGTCTCAGCGAAAT	TTTA	Fig.1K,S5,S14A
23nt-*HEK3*-3	ATGAGTGCTGAAGAGAACGCCAC	TTTG	Fig.1K,S5,S9A,S14A
20nt-*HEK3*-3	ATGAGTGCTGAAGAGAACGC	TTTG	Fig.1K,S5,S9A,S14A
23nt-*KLF4*-2	AGCAAACGTCTATTTTGTATATT	TTTA	Fig.1K,S5
20nt-*KLF4*-2	AGCAAACGTCTATTTTGTAT	TTTA	Fig.1K,S5
23nt-*EMX1*-5	TACTTTGTCCTCCGGTTCTGGAA	TTTG	Fig.2C,S6,S8B,S9B,S10B
20nt-*EMX1*-5	TACTTTGTCCTCCGGTTCTG	TTTG	Fig.2C,S6,S8B,S9B
23nt-*KLF4*-5	GTTTAAACACACCGGGTTAATAA	TTTG	Fig.2C,S10B,S14B
20nt-*KLF4*-5	GTTTAAACACACCGGGTTAA	TTTG	Fig.2C,S14B
23nt-*DNMT3B*-4	CCTTAATCCTCTCCCAGACATAA	TTTA	Fig.2C,S6,S8B
20nt-*DNMT3B*-4	CCTTAATCCTCTCCCAGACA	TTTA	Fig.2C,S6,S8B
23nt-*DNMT3B*-5	TGTGAGCAATAAAGCTGTTTATT	TTTG	Fig.2C,S6
20nt-*DNMT3B*-5	TGTGAGCAATAAAGCTGTTT	TTTG	Fig.2C,S6
23nt-*DNMT3B*-6	GACTCAGCCCACCTGCACTCAGG	TTTG	Fig.2C,S6
20nt-*DNMT3B*-6	GACTCAGCCCACCTGCACTC	TTTG	Fig.2C,S6
23nt-*RUNX1*-4	TTTTGCTGTCCCTTGCCCGGAAT	TTTA	Fig.2C,S6,S9B,S14B
20nt-*RUNX1*-4	TTTTGCTGTCCCTTGCCCGG	TTTA	Fig.2C,S6,S9B,S14B
23nt-*RUNX1*-6	GCCCTGAAGTCCCCAGGCTGGTG	TTTG	Fig.2C,S6,S9B,S14B
20nt-*RUNX1*-6	GCCCTGAAGTCCCCAGGCTG	TTTG	Fig.2C,S6,S9B,S14B
PCSK9-1	CCCAGAGCATCCCGTGGAAC	TTTG	Fig.4A
PCSK9-2	TTCGGAAAAGCCAGCTGGTC	TTTA	Fig.4A,4C
PCSK9-3	TGTCACAGAGTGGGACATCA	TTTG	Fig.4A
PCSK9-4	CAGGTTGGCAGCTGTTTTGC	TTTG	Fig.4A
PCSK9-5	ACTCTAAGGCCCAAGGGGGC	TTTG	Fig.4A
23nt-*KLF4*-1	TACAATGGTTTATTCCCAAGTAT	TTTG	Fig.S1,S2,S3B,S8D,S9A,S10A
20nt-*KLF4*-1	TACAATGGTTTATTCCCAAG	TTTG	Fig.S3B,S4,S8D,S9A
18nt-*KLF4*-1	TACAATGGTTTATTCCCA	TTTG	Fig.S3B
23nt-*DNMT1*-1	ACACAGTTCTAGGGGCATCAGGC	TTTG	Fig.S8B
20nt-*DNMT*1-1	ACACAGTTCTAGGGGCATCA	TTTG	Fig.S8B
23nt-*FANCF*-2	ATTAATCCACAACCACCTCATCT	TTTA	Fig.S8B
20nt-*FANCF*-2	ATTAATCCACAACCACCTCA	TTTA	Fig.S8B
23nt-*HEK3*-2	GTAAGCGGGGAGATGGGCCCACA	TTTG	Fig.S8D
20nt-*HEK3*-2	GTAAGCGGGGAGATGGGCCC	TTTG	Fig.S8D

Targets	Un1Cas12f1 sgRNA spacer sequences	PAM	Figure
23nt-*HEK4*-2	AGCAACATCAACAACAGACAATG	TTTG	Fig.S8D
20nt-*HEK4*-2	AGCAACATCAACAACAGACA	TTTG	Fig.S8D

Table S2. Sequences of spacer-overlap target sites between Cas12f and Cas9

Targets	Cas12f PAM	sgRNA spacer sequences	Cas9 PAM	Figure
PWAR5	TTTA	AACAAATCACTGACTAACCA	AGG	Fig.3B,S7,S12,S13
NEK10-1	TTTA	AGACAAGCTGTCTTCCTTCA	GGG	Fig.3B,S7,S12,S13
INIP	TTTA	AGAGCAGCGATTGTAAGGAG	AGG	Fig.3B,S7
KLHL29	TTTA	GAGAGACCGCTCAGGCTGGA	GGG	Fig.3B,S7

Table S3. Sequences of primers used in amplifying target genomic DNAs for efficiency detection