

Supplemental information

Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity

Connor A. Tsuchida, Shouyue Zhang, Mohammad Saffari Doost, Yuqian Zhao, Elizabeth O'Brien, Huan Fang, Jia Wang, Cheng-Ping Li, Danyuan Li, Zhuo-Yan Hai, Jonathan Chuck, Julian Brötzmann, Araz Vartoumian, David Burstein, Xiao-Wei Chen, Eva Nogales, Jennifer A. Doudna, Jun-Jie Gogo Liu

List:

Figure S1-S8

Table S1

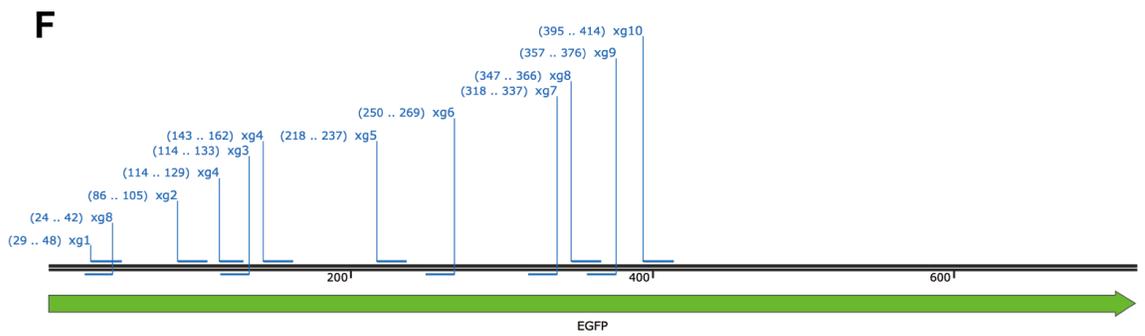
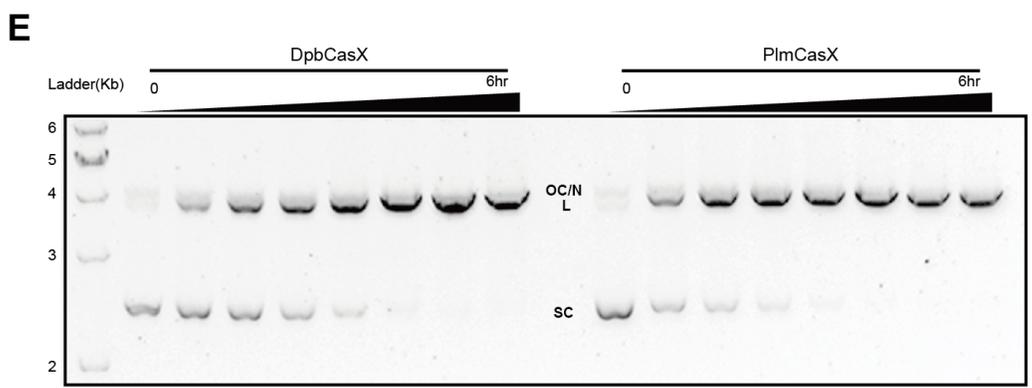
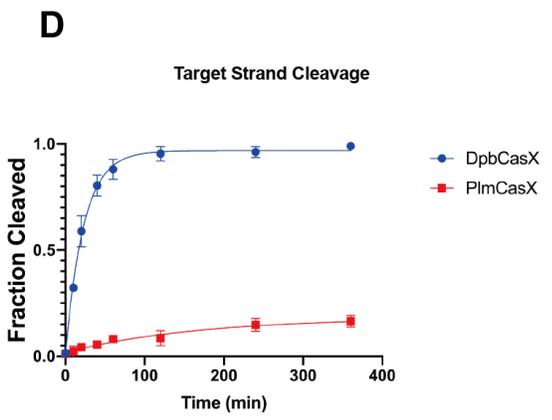
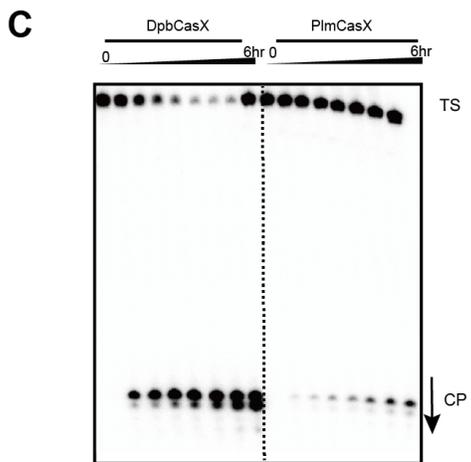
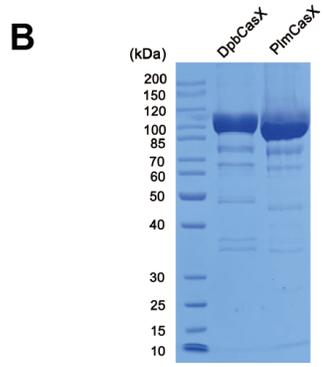
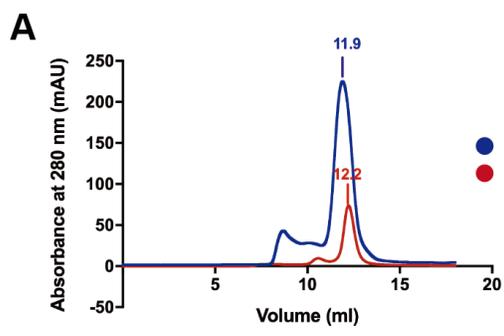


Figure S1. Targeted DNA cleavage by DpbCasX and PlmCasX, Related to Figure 1.

(A) Purification of DpbCasX and PlmCasX by size exclusion chromatography. The representative S200 size exclusion traces by UV280 absorbance are aligned and shown. DpbCasX eluted at 12.2 mL and PlmCasX eluted at 11.9 mL. **(B)** SDS-PAGE for DpbCasX and PlmCasX samples taken from the elution peak after size exclusion chromatography. **(C)** *In vitro* dsDNA cleavage activity comparison between DpbCasX and PlmCasX revealed by denaturing PAGE. TS denotes the target strand DNA which was ³²P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 5 mins, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(D)** Plot of DNA cleavage kinetics based on the TS band density from the cleavage fractions compared to the input TS band density at the reaction time of 0 min (n = 3, mean ± SD). The rate constant k values for DpbCasX and PlmCasX cleavage of the TS were 0.04351 and 0.004978 (fraction/minute), respectively. **(E)** *In vitro* plasmid cleavage activity comparison between DpbCasX and PlmCasX revealed by agarose gel. OC/N indicates the open-circle or nicked plasmid, L indicates the linearized plasmid and SC indicates the super-coiled plasmid (n = 3, representative gel shown). The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(F)** The distribution of ten sgRNA_{v1} spacers across the genomically integrated GFP gene.

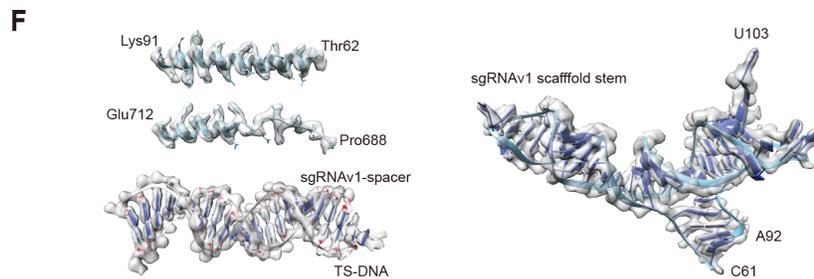
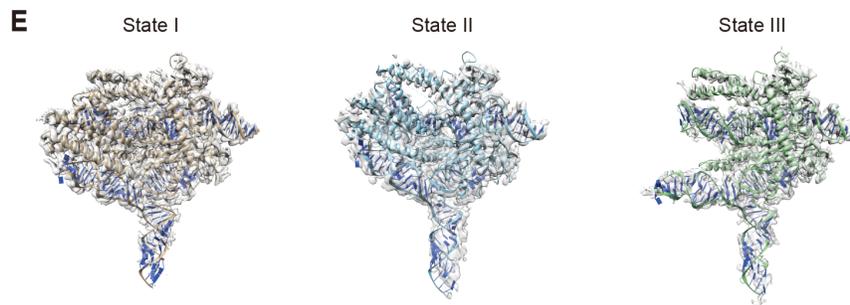
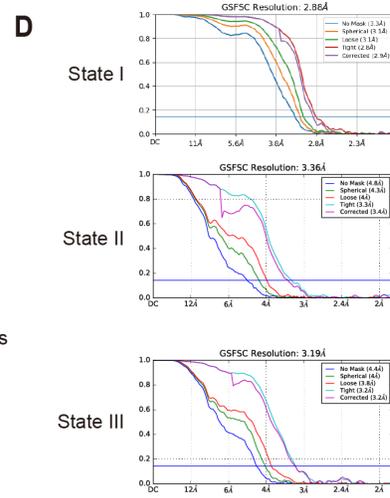
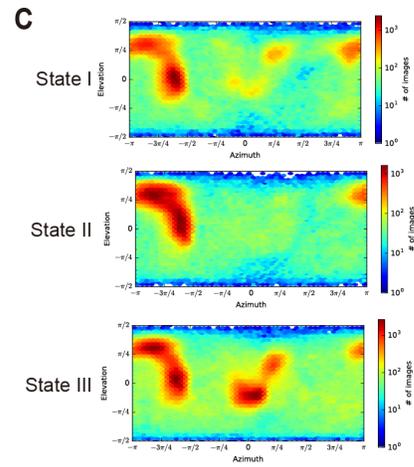
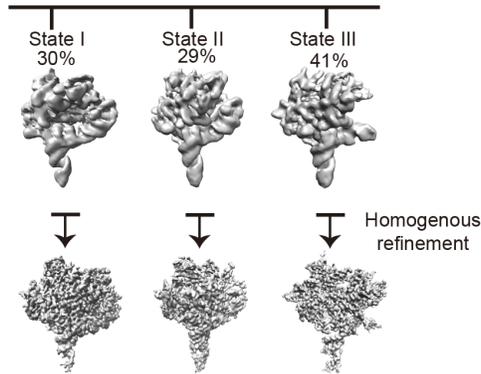
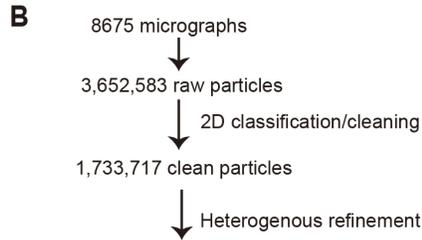
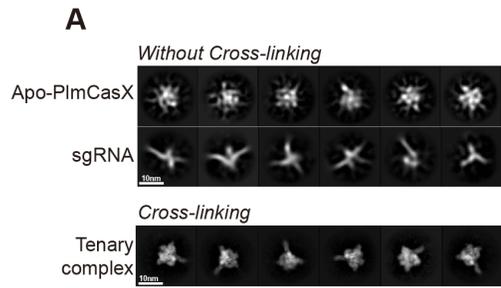


Figure S2. Single particle cryo-EM analysis of the dPlmCasX-sgRNAv1-dsDNA complex, Related to Figure 2.

(A) Representative 2D class averages of native and cross-linked (BS3 crosslinking) complexes. (B) The workflow for single particle cryo-EM analysis in CryoSparc. The particles in State I, State II and State III account for 30%, 29% and 41%, respectively, of all the particles used for 3D refinement. (C) Euler angle distribution of the refined particles belonging to the three states. (D) Fourier shell correlation (FSC) curves calculated using two independent half maps. The final resolutions for B-factor corrected maps were 2.9 Å (State I), 3.4 Å (State II) and 3.2 Å (State III). Panels C and D were directly taken from the standard output of CryoSparc. (E) Overall fitting between the atomic models and the EM maps from the three states of the dPlmCasX-sgRNAv1-dsDNA complex. The atomic models are shown by ribbon cartoon and EM maps are shown by transparent surface. (F) Details of the fitting between the atomic model and the EM map of State I. The amino acid sidechains are shown.

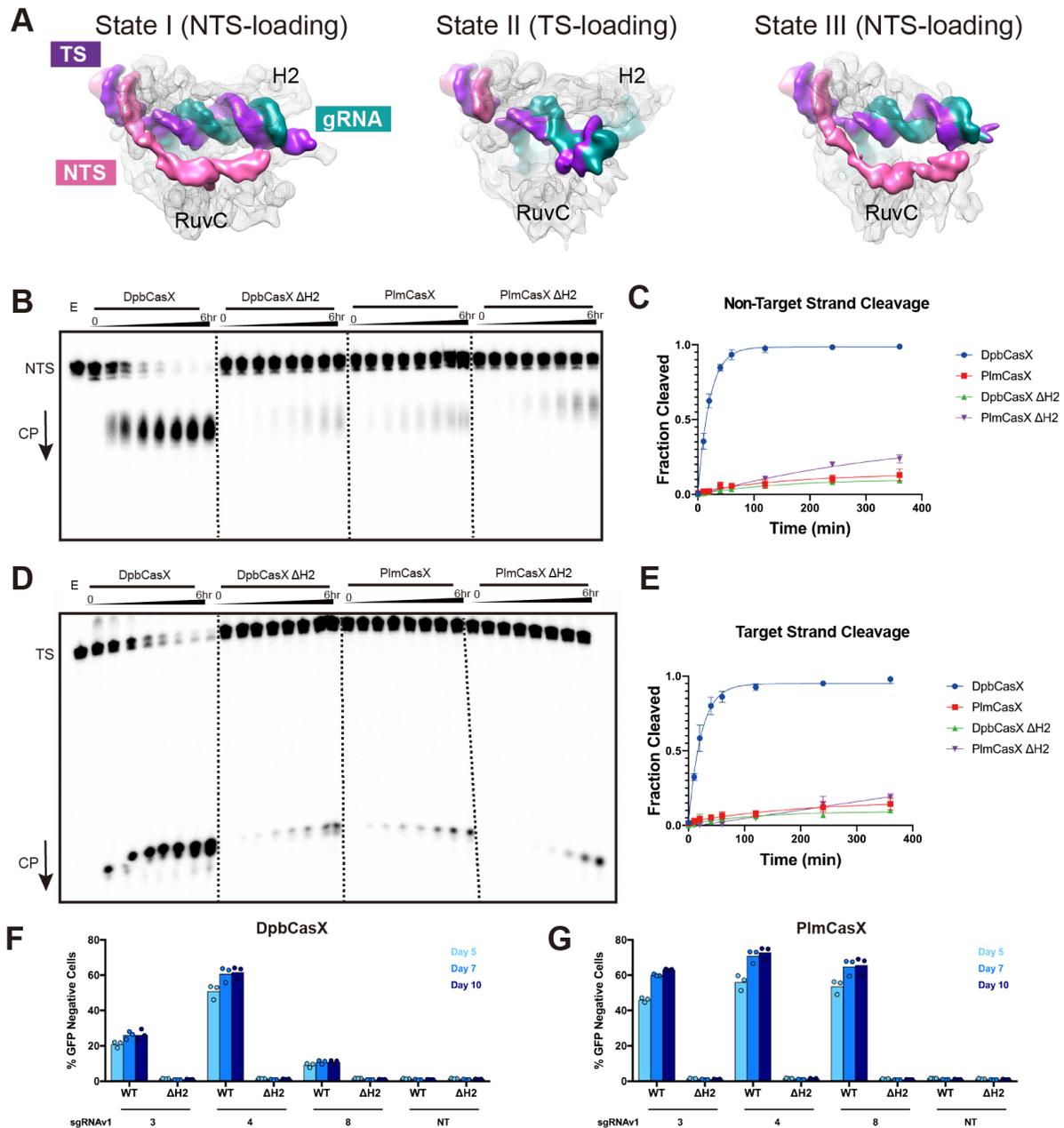


Figure S3. Effect of the Helical-II domain in DpbCasX and PlmCasX on dsDNA cleavage, Related to Figure 2.

(A) Conformational states within the dPlmCasX-sgRNAv1-dsDNA complex revealed by cryo-EM. The protein density is shown as a transparent mesh, the TS density is colored purple, the NTS is colored pink and the sgRNA density is colored teal. The RuvC and H2 domains are indicated in

each map. All maps are low-pass filtered to 4.5 Å. **(B)** *In vitro* biochemical dsDNA cleavage activity comparison between DpbCasX, DpbCasX with a Helical-II truncation (DpbCasX Δ H2), PlmCasX and PlmCasX with a Helical-II truncation (PlmCasX Δ H2). NTS denotes the non-target strand DNA which was 32 P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(C)** Plot of DNA cleavage kinetics analyzed based on the cleaved NTS band density compared to the input NTS band density at the reaction time of 0 min (n = 3, mean \pm SD). **(D)** *In vitro* biochemical dsDNA cleavage activity comparison among DpbCasX, DpbCasX Δ H2, PlmCasX and PlmCasX Δ H2. TS denotes the target strand DNA which was 32 P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(E)** The cleavage fraction analysis based on the cleaved TS band density compared to the input TS band density at the reaction time of 0 min (n = 3, mean \pm SD). **(F)** Genome editing efficacy comparison between wild type DpbCasX (WT) and DpbCasX Δ H2 (Δ H2) constructs with sgRNAv1. **(G)** Genome editing efficacy comparison between wild type PlmCasX (WT) and PlmCasX Δ H2 (Δ H2) constructs with sgRNAv1. sgRNAv1 with spacers 3, 4 and 8 targeting the GFP gene were tested. NT denotes the non-targeting sgRNAv1 control (n = 3, mean). Cells were collected and analyzed by flow cytometry at 5, 7 and 10 days after plasmid transfection.

Figure S4. Design of CasX protein chimeras, Related to Figure 3.

(A) Amino acid sequence alignment between DpbCasX and PlmCasX using Clustal. Only regions in proximity to loops R1, R2 and R3 are shown for clarity. **(B)** Details of the fitting between the atomic model and EM map at region 1, 2 and 3. The R1, R2 and R3-loops are colored in red and labeled in each model. For DpbCasX, the published model (PDB code 6NY2) with re-built R1-loop and EM map (EMDB code EMD-8996) was used. In the bottom right panel, the EM map is shown at a low contour threshold of 3.68 times sigma due to the weak density of the R3-loop in the PlmCasX reconstruction. The EM maps in other panels are shown at a contour threshold of 6 to 9 times sigma. **(C)** Cartoon models for wild type CasX proteins and chimeric designs. The R1, R2 and R3 regions are presented as red loops on the 3D structure.

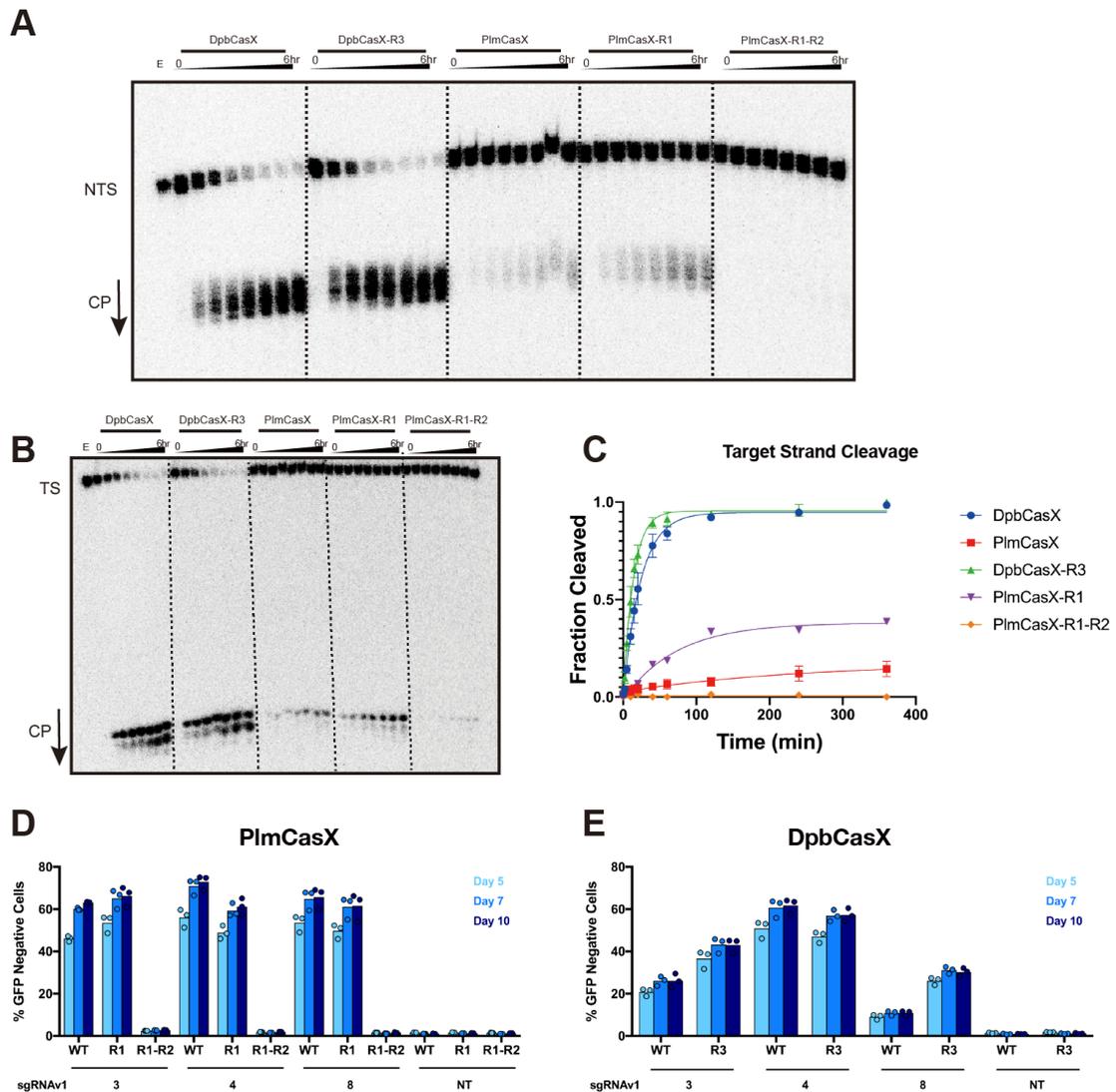


Figure S5. DNA cleavage by CasX protein chimeras, Related to Figure 3.

(A) *In vitro* dsDNA cleavage activity comparison among wild type and CasX chimeras using the sgRNAv1 scaffold revealed by denaturing PAGE. NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. CP indicates the cleavage product. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1hr, 2hrs, 4hrs and 6hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(B)** *In vitro* dsDNA cleavage activity comparison among wild type and CasX chimeras using the sgRNAv1 scaffold revealed by denaturing PAGE. TS

denotes the target strand DNA which was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1hr, 2hrs, 4hrs and 6hrs, respectively. E indicates an empty well with labeled DNA but no CasX enzyme. **(C)** Plot of DNA cleavage kinetics analyzed based on the TS band density from the cleaved fractions compared to the input TS band density at the reaction time of 0 min ($n = 3$, mean \pm SD). The rate constant k values for DpbCasX, PlmCasX, DpbCasX-R3 and PlmCasX-R1 were 0.04189, 0.004144, 0.07236 and 0.01364 (fraction/minute), respectively. **(D)** Genome editing efficacy comparison between wild type PlmCasX (WT), PlmCasX-R1 (R1), and PlmCasX-R1-R2 (R1-R2) chimeras with sgRNAv1. **(E)** Genome editing efficacy comparison between wild type DpbCasX (WT) and DpbCasX-R3 chimera (R3) with sgRNAv1. sgRNAv1 with spacers 3, 4 and 8 targeting the GFP gene were tested. NT denotes the non-targeting sgRNAv1 control ($n = 3$, mean). Cells were collected and analyzed by flow cytometry at 5, 7 and 10 days after plasmid transfection.

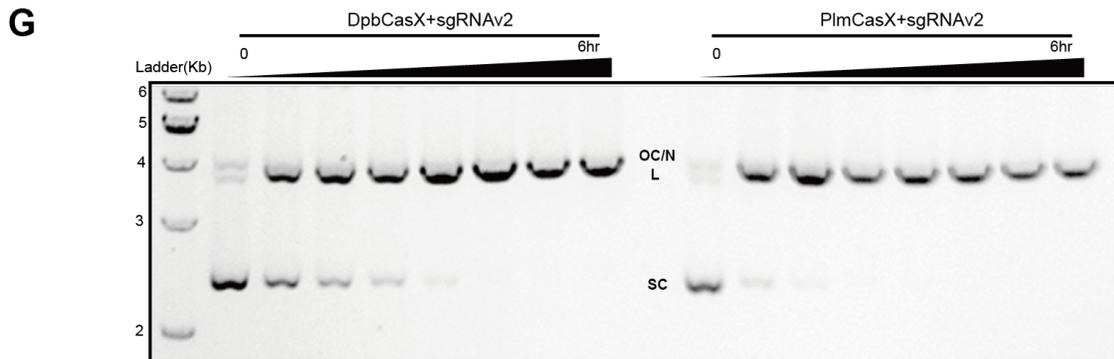
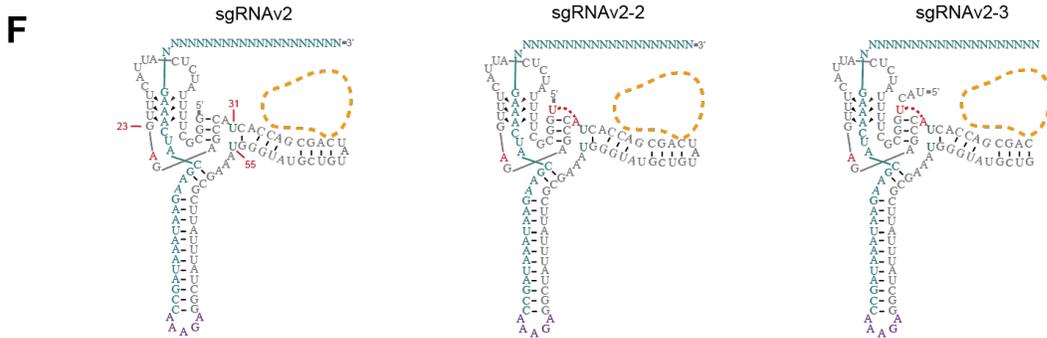
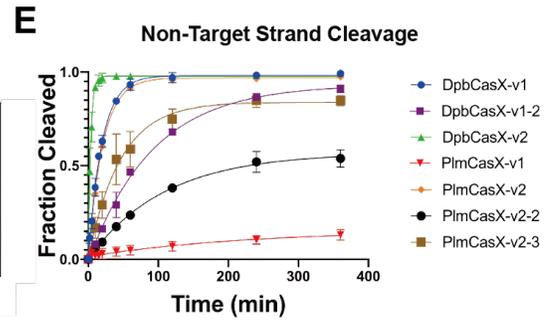
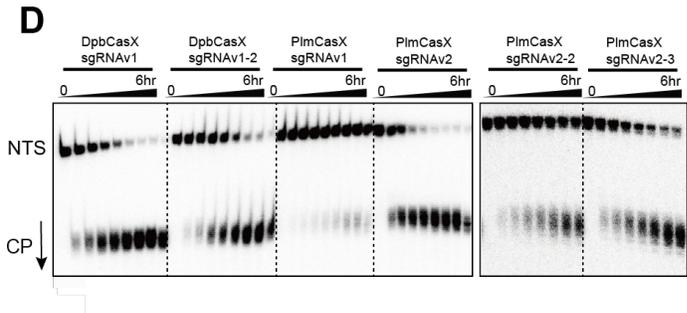
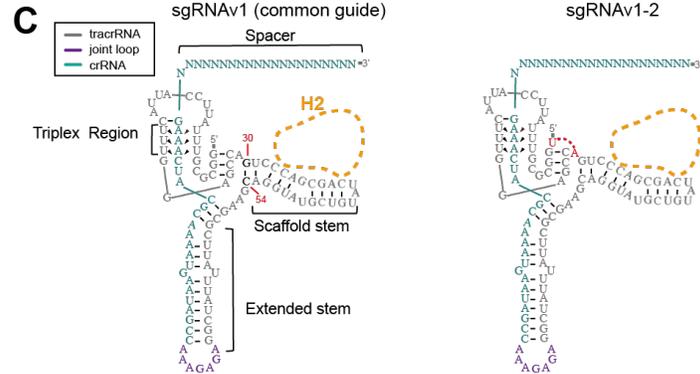
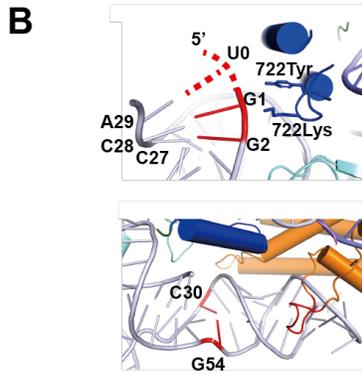
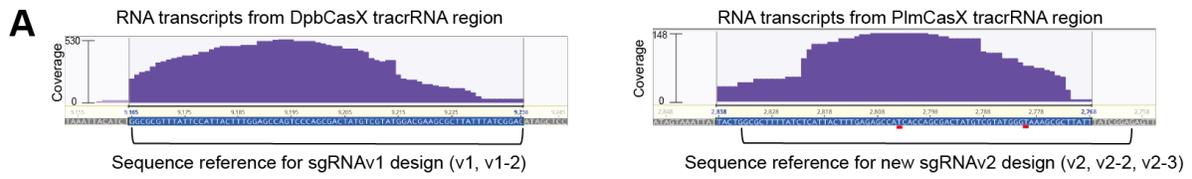


Figure S6. Updated design of the CasX sgRNA, Related to Figure 4.

(A) Transcripts from the hypothetical CasX tracrRNA loci revealed by meta-transcriptome sequencing. The coverage axis denotes the number of sequencing reads. The abscissa denotes the genome sequence of the hypothetical tracrRNA region. **(B)** The structural details for sgRNAv1. The additional U at the 5' end was hypothetically modeled in the top panel. The structural details of the G30-C54 base-pairing region in sgRNAv1 are also shown in the bottom panel. **(C)** The secondary architecture of sgRNAv1 revealed by cryo-EM. The sgRNAv1-2 architecture is modeled based on sgRNAv1. The potential base pairing between the additional 5' U and A29 is labeled with a dashed line. The Helical-II domain was modeled to indicate the interaction interface with the sgRNA scaffold stem. **(D)** *In vitro* dsDNA cleavage activity comparison among different sgRNAs by DpbCasX and PlmCasX. NTS denotes the non-target strand DNA which was ³²P labeled on the 5' end. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. **(E)** Plot of DNA cleavage kinetics analyzed based on the NTS band density from the cleaved fractions compared to the input NTS band density at the reaction time of 0 min (n = 3, mean ± SD) (DpbCasX-v2 from Figure 4A). The rate constant k values for DpbCasX-v1, DpbCasX-v1-2, DpbCasX-v2, PlmCasX-v1, PlmCasX-v2, PlmCasX-v2-2 and PlmCasX-v2-3 were 0.05065, 0.01087, 0.2817 and 0.004433, 0.04858, 0.009191 and 0.02169 (fraction/minute), respectively. **(F)** The hypothetical secondary architectures of sgRNAv2 designs. **(G)** *In vitro* plasmid cleavage activity comparison between DpbCasX and PlmCasX with sgRNAv2 revealed by agarose gel. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. OC/N indicates the open-circle or nicked plasmids, L indicates the linearized plasmids and SC indicates the super-coiled plasmids (n = 3, representative gel shown).

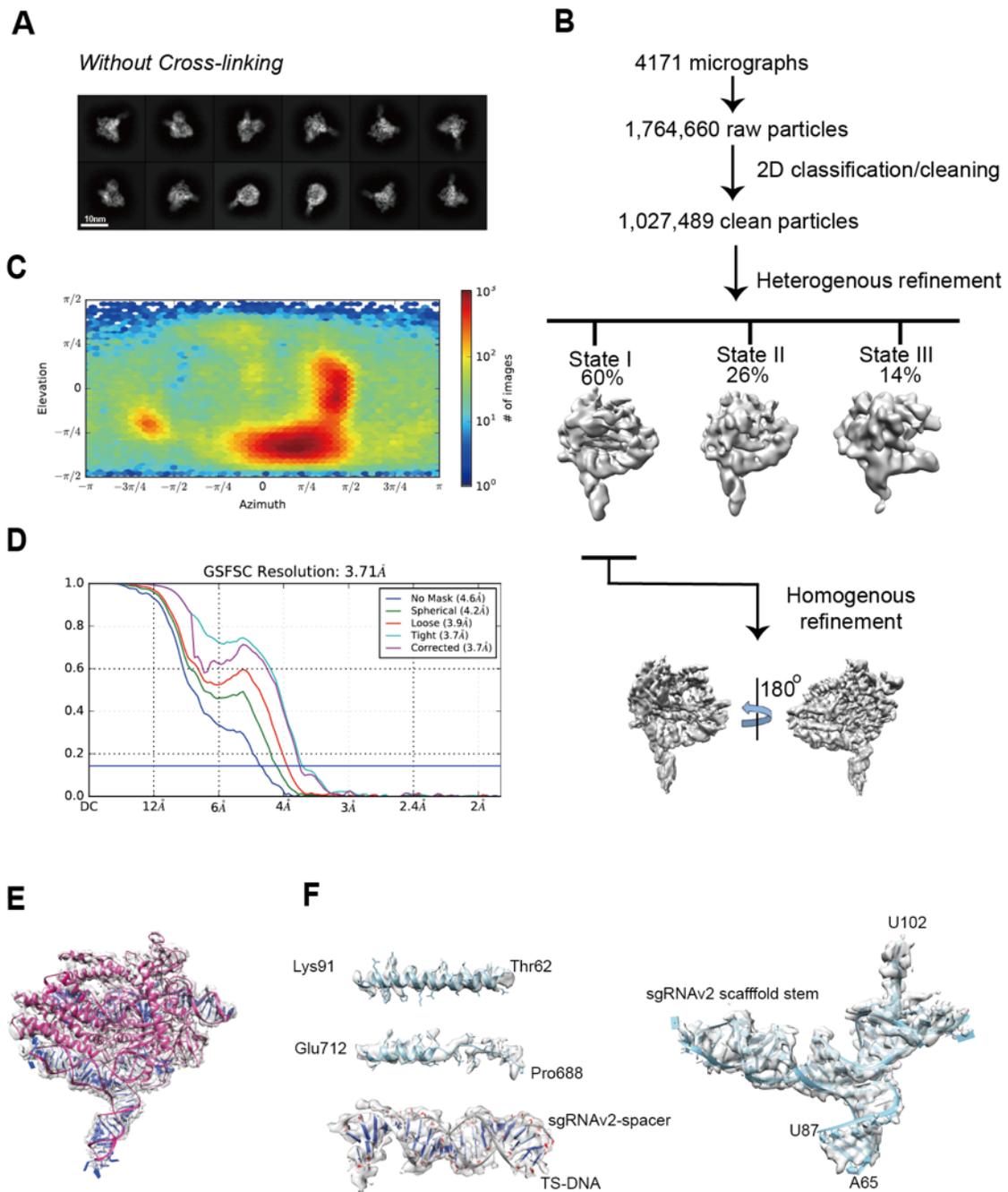


Figure S7. Single particle cryo-EM analysis of the dPlmCasX-sgRNAv2-dsDNA complex, Related to Figure 4.

(A) Representative 2D class averages of the native complex (without BS3 crosslinking); the scale bar is 10 nm. **(B)** The workflow for single particle cryo-EM analysis in CryoSparc. The particles in State I, State II and State III account for 60%, 26% and 14%, respectively, of all the particles used for 3D refinement. **(C)** Euler angle distribution of the refined particles belonging to State I. **(D)** Fourier shell correlation (FSC) curves calculated using two independent half maps. The final resolutions for the B-factor corrected State I map was 3.7 Å. Panels **C** and **D** are directly taken from the standard outputs of CryoSparc. **(E)** Overall fitting between the atomic model and the EM map of the dPlmCasX-sgRNAv2-dsDNA complex in State I. The atomic model is shown by ribbon cartoon and EM map is shown by transparent surface. **(F)** Details of the fitting between the atomic model and the EM map of State I. The amino acid sidechains are shown.

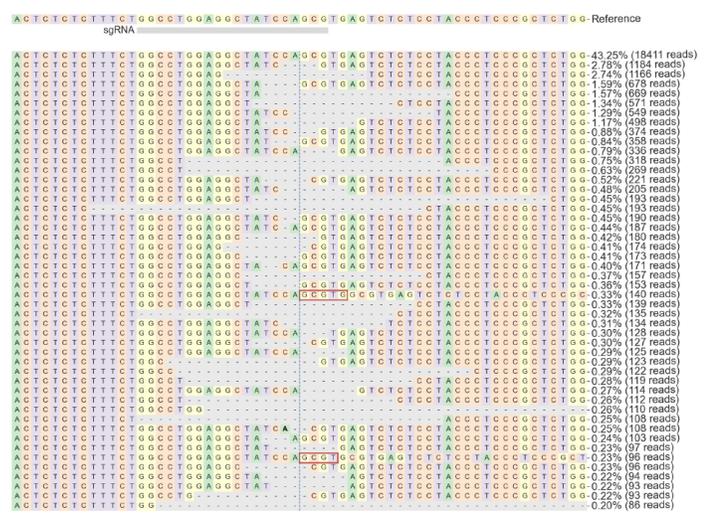
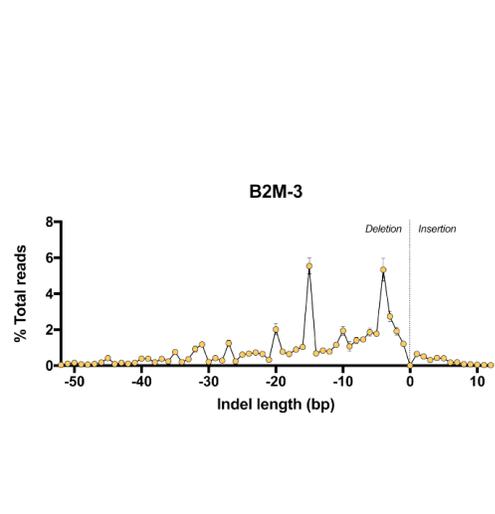
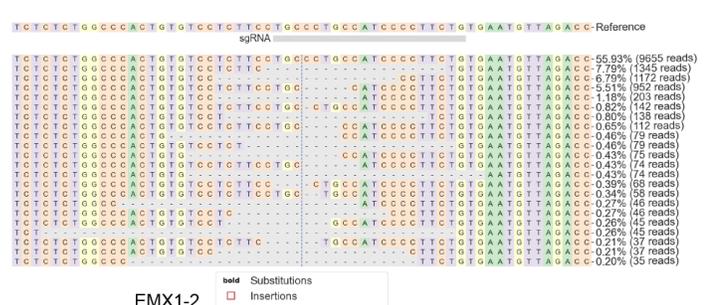
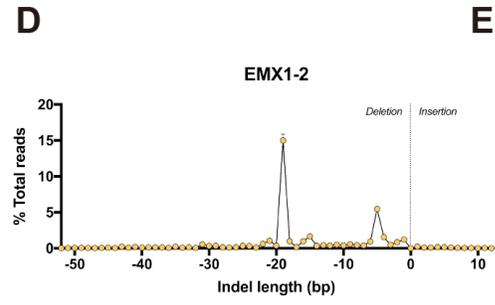
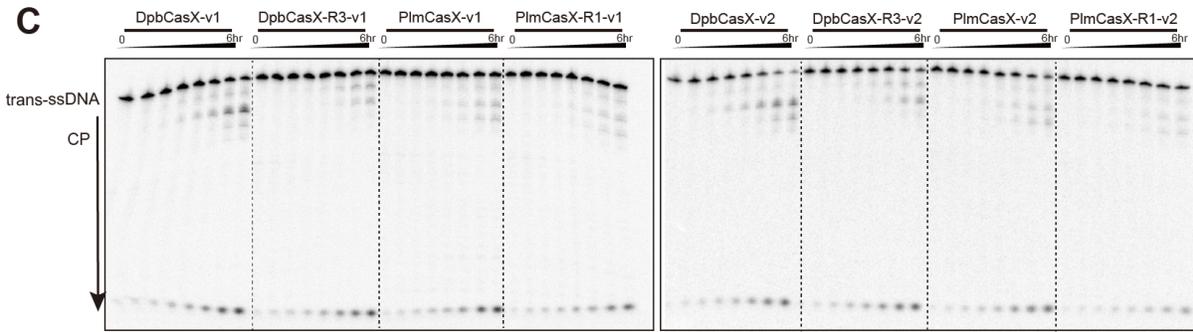
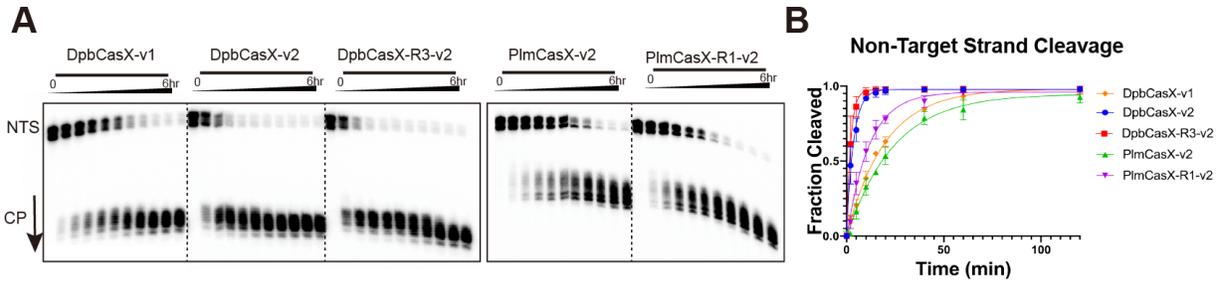


Figure S8. CasX nucleases with largely enhanced genome editing efficacy, Related to Figure 5.

(A) *In vitro* dsDNA cleavage activity comparison among different sgRNAs by wild type CasX and chimeras. NTS denotes the non-target strand DNA which was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 2 mins, 5 mins, 10 mins, 15 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively. (B) Plot of DNA cleavage kinetics analyzed based on the NTS band density from the cleaved fractions compared to the input NTS band density at the reaction time of 0 min ($n = 3$, mean \pm SD). The rate constant k values for DpbCasX-v1, DpbCasX-v2, DpbCasX-R3-v2, PlmCasX-v2 and PlmCasX-R1-v2 are 0.05065, 0.2817, 0.4730, 0.04182 and 0.08680 (fraction/minute), respectively. (C) *In vitro* trans-ssDNA cleavage activity comparison among different constructs. The trans-ssDNA was ^{32}P labeled on the 5' end. The fractions were collected at 0 min, 10 mins, 20 mins, 40 mins, 1 hr, 2 hrs, 4 hrs and 6 hrs, respectively ($n = 3$, representative gel shown). (D) Indel size distribution by PlmCasX-R1-v2 and an *EMXI*-targeting sgRNAv2 spacer 2 (top) or *B2M*-targeting sgRNAv2 spacer 3 (bottom) ($n = 3$, mean \pm SD). Deletions are plotted as negative in length and insertions are plotted as positive in length. (E) Representative summary of indels generated by PlmCasX-R1-v2 and an *EMXI*-targeting sgRNAv2 spacer 2 (top) or *B2M*-targeting sgRNAv2 spacer 3 (bottom). Only indels with frequencies $\geq 0.2\%$ of the total sequencing reads are shown for brevity.

Supplementary Table 1. DNA/RNA sequences used in this study, related to Figures 1-5.

Experiment	Name	Sequence
Preparation of the ternary complex for cryoEM	TS DNA	ATCGTTATAC TTTGATTTTCTGCTGCAGGA TGAA ATCCCG
Preparation of ternary complex for cryoEM	NTS DNA	CGGGAT TTCA TCCTGCAGCA TCCCCGACCC GTATAACGAT
sgRNAv1 used for EM complex preparations, and cleavage assays	sgRNAv1	ggCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACUA UGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGA UAAGUAAAACGCAUCAAAAGUCCUGCAGCAGAAAUCAAA
<i>In vitro</i> DNA cleavage	NTS DNA	GCCCCGCGGGAT TTCA TCCTGCAGCAGAAAATCAA GACAATG AATATTTTCGGCGC
<i>In vitro</i> DNA cleavage	TS DNA	GCGCCGAAATATTCATTGTC TTTGATTTTCTGCTGCAGGA TGA AATCCCGCGGGC
The secondary architecture of sgRNAv1	sgRNAv1	ggCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACUA UGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCGA UAAGUAAAACGCAUCAAAAG
The secondary architecture of sgRNAv1	sgRNAv1-2	UggCGCGUUUAUCCAUAUACUUUGGAGCCAGUCCAGCGACU AUGUCGUAUGGACGAAGCGCUUAUUUAUCGGAGAGAAACCG AUAAGUAAAACGCAUCAAAAG
The secondary architecture sgRNAv2	sgRNAv2	ggCGCUUUUAUCUCAUUAUACUUUGAGAGCCAUCACCAGCGACU AUGUCGUAUGGGUAAAAGCGCUUAUUUAUCGGAGAAACCGAU AAAUAAGAAGCAUCAAAAG
The secondary architecture sgRNAv2	sgRNAv2-2	UggCGCUUUUAUCUCAUUAUACUUUGAGAGCCAUCACCAGCGAC UAUGUCGUAUGGGUAAAAGCGCUUAUUUAUCGGAGAAACCGA UAAAUAAGAAGCAUCAAAAG
The secondary architecture sgRNAv2	sgRNAv2-3	UACUGGCGCUUUUAUCUCAUUAUACUUUGAGAGCCAUCACCAG CGACUAUGUCGUAUGGGUAAAAGCGCUUAUUUAUCGGAGAAA CCGAUAAAUAAGAAGCAUCAAAAG
sgRNAv2 used for cleavage assays	sgRNAv2	ggCGCUUUUAUCUCAUUAUACUUUGAGAGCCAUCACCAGCGACU AUGUCGUAUGGGUAAAAGCGCUUAUUUAUCGGAGAAACCGAU AAAUAAGAAGCAUCAAAAGUCCUGCAGCAGAAAUCAAA
trans-cleavage assay	Random ssDNA	GTTTATTTACTTTAGTCACTCCAGGATTCCAATAGATATTTACT TTGAAG
CasX human GFP targeting	sgRNA spacer 1	CCGGGGTGGTGCCCATCCTG
CasX human GFP targeting	sgRNA spacer 2	GCGTGTCCGGCGAGGGCGAG
CasX human GFP targeting	sgRNA spacer 3	GGGTCAGCTTGCCGTAGGTG
CasX human GFP targeting	sgRNA spacer 4	TCTGCACCACCGCAAGCTG
CasX human GFP targeting	sgRNA spacer 5	GCCGCTACCCGACCACATG
CasX human GFP targeting	sgRNA spacer 6	GGCATGGCGGACTTGAAGAA
CasX human GFP targeting	sgRNA spacer 7	CCTCGGCGGGTCTTGTAG
CasX human GFP targeting	sgRNA spacer 8	AGGGCGACACCCTGGTGAAC

CasX human GFP targeting	sgRNA spacer 9	GCTCGATGCGGTTACCAGG
CasX human GFP targeting	sgRNA spacer 10	AGGAGGACGGCAACATCCTG
CasX human GFP targeting	Non-targeting	CGTGATGGTCTCGATTGAGT
CasX human <i>EMX1</i> targeting	sgRNA spacer 1	CTCTGGCCCACTGTGTCCTC
CasX human <i>EMX1</i> targeting	sgRNA spacer 2	CAGAAGGGGATGGCAGGGCA
CasX human <i>EMX1</i> targeting	sgRNA spacer 3	CCTGGGCCAGGGAGGGAGGG
CasX human <i>B2M</i> targeting	sgRNA spacer 1	GGAATGCCCGCCAGCGCGAC
CasX human <i>B2M</i> targeting	sgRNA spacer 2	TGAAGCTGACAGCATTCGGG
CasX human <i>B2M</i> targeting	sgRNA spacer 3	GGCCTGGAGGCTATCCAGCG
CasX human <i>B2M</i> targeting	sgRNA spacer 4	TCTCCCGCTCTGCACCCTCT
CasX human <i>TTR</i> targeting	sgRNA spacer 1	CCGGTGCCCTGGGTGTAGAG
CasX human <i>TTR</i> targeting	sgRNA spacer 2	AGATGCTGTCCGAGGCAGTC
CasX human <i>TTR</i> targeting	sgRNA spacer 3	GAACACATGCACGGCCACAT
CasX human <i>TTR</i> targeting	sgRNA spacer 4	GAAAGGCTGCTGATGACACC
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	Non-targeting	AAGTAAAACCTCTACAAATG
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>EMX1</i> Illumina library prep fwd primer (with adaptor sequence)	GCTCTTCCGATCTCAGCTCTGTGACCCTTTGTTG
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>EMX1</i> Illumina library prep rev primer (with adaptor sequence)	GCTCTTCCGATCTCTGCCGTTTGTACTTTGTCTC
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>B2M</i> Illumina library prep fwd primer (with adaptor sequence)	GCTCTTCCGATCTCTTGGAGACAGGTGACGGTC
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>B2M</i> Illumina library prep rev primer (with adaptor sequence)	GCTCTTCCGATCTGGGCCACCAAGGAGAACTTG
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>TTR</i> Illumina library prep fwd primer (with adaptor sequence)	GCTCTTCCGATCTAGTGTGTAATTCTTGTTCGCTCCA
CasX human <i>EMX1</i> , <i>B2M</i> , and <i>TTR</i> targeting	<i>TTR</i> Illumina library prep rev primer (with adaptor sequence)	GCTCTTCCGATCTCAAGTGAGGGGCAAACGG