

COMPOSITIONS AND METHODS FOR GENE EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to US Application No. 63/171,698 filed April 7, 2021, US Application No. 63/114,850 filed November 17, 2020, and US Application No. 63/033,397 filed June 2, 2020, the disclosures of which are incorporated by reference herein in their entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made with government support under grant no. HR0011-17-2-0043 awarded by The Defense Advanced Research Projects Agency. The government has certain rights in the invention.

REFERENCE TO A "SEQUENCE LISTING," A TABLE, OR A COMPUTER PROGRAM LISTING APPENDIX SUBMITTED AS AN ASCII FILE

[0003] The Sequence Listing written in file 048536-682001WO_SequenceListing_ST25.txt, created May 31, 2021, 563,294 bytes, machine format IBM-PC, MS Windows operating system, is hereby incorporated by reference.

BACKGROUND

[0004] Although considered a promising therapeutic approach for treatment of disease, genome editing carries inherent risks due to the potential for genotoxicity from double strand breaks. Further, genome editing often is associated with an all-or-none effect on the target gene (i.e., it produces a full knockout). In contrast, targeted epigenome engineering does not carry the risk of DSB-induced genotoxicity; further, it affords the opportunity to create a more graded effect on gene expression and thus function from a complete silencing through a less pronounced effect. Provided herein are solutions to these and other needs in the art.

BRIEF SUMMARY

[0005] The disclosure provides fusion proteins comprising a DNA methyltransferase domain, a first XTEN linker comprising from about 5 to about 864 amino acid residues, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker comprising from about 5 to about 864 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues. In embodiments, the nuclease-deficient

RNA-guided endonuclease enzyme is a CRISPR-associated protein.

[0006] The disclosure provides methods of silencing a target nucleic acid sequence in a cell, including delivering a first polynucleotide encoding a fusion protein as described herein, including embodiments and aspects thereof, to a cell containing the target nucleic acid, and delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA. In aspects, the target nucleic acid comprises a CpG island. In aspects, the target nucleic acid does not comprise a CpG island. In embodiments, the nuclease-deficient RNA-guided endonuclease enzyme is a CRISPR-associated protein. Methods of silencing a target nucleic acid sequence can be used to treat numerous diseases, such as infectious diseases.

[0007] The disclosure provides fusion proteins comprising a DNA methyltransferase domain, a first XTEN linker comprising from about 5 to about 864 amino acid residues, a nuclease-deficient endonuclease enzyme (e.g., a zinc finger domain, a TALE), a second XTEN linker comprising from about 5 to about 864 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues.

[0008] The disclosure provides methods of silencing a target nucleic acid sequence in a cell, including delivering a first polynucleotide encoding a fusion protein as described herein, including embodiments and aspects thereof, to a cell containing the target nucleic acid. Methods of silencing a target nucleic acid sequence can be used to treat numerous diseases, such as infectious diseases.

[0009] These and other embodiments and aspects of the disclosure are described in detail herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] **FIGS. 1A-1G** show durable and multiplexed gene silencing by CRISPRoff. **FIG. 1A:** a schematic of dCas9 epigenetic editor fusion proteins that were tested for gene silencing activity. 3L denotes Dnmt3L. **FIG. 1B:** plasmids encoding dCas9 fusions and sgRNAs were co-transfected into HEK293T cells stably expressing a DNA methylation-sensitive Snrpn-GFP reporter. Transfected cells were sorted 2 days after transfection and GFP silencing was monitored over time. **FIG. 1C:** a time course comparing GFP silencing activities of CRISPRoff-V1, dCas9-3A-3L, and dCas9-KRAB. **FIG. 1D:** bisulfite PCR analysis of the *Snrpn* locus before or after CRISPRoff targeting. The white circles indicate unmethylated CpG dinucleotides and black circles represent methylated CpG dinucleotides. Each row represents one sequencing read. The red square denotes the sgRNA binding site. **FIG. 1E:** a comparison of CRISPRoff-V1

(black) and CRISPRoff-V2 (blue) editors in silencing the endogenously GFP-tagged *H2B* gene. The dotted lines represent protein expression of CRISPRoff-V1 and -V2. **FIG. 1F:** a representative flow cytometry plot of H2B-GFP expression of cells at 50 days post-transfection of CRISPRoff V2. **FIG. 1G:** bisulfite sequencing analysis of a 126 bp region of the *H2B* CpG island. The red square denotes the sgRNA binding site. **FIG. 1H:** quantification of cells with *ITGB1*, *CD81*, or *CD151* silenced 3 weeks post-transfection (p.t.) of CRISPRoff-V1 or -V2 with individual sgRNAs (a-c) or a pool of three sgRNAs (a, b, c). **FIG. 1I:** quantification of cells with *ITGB1*, *CD81*, *CD151* silenced 30 days p.t. from single or double gene targeting experiments. **FIG. 1J:** quantification of multiplexed triple gene silencing by either gating on *ITGB1*-off cells then gating for *CD81*- and *CD151*-off cells (left bar) or by first gating on *ITGB1*-off cells, then *CD151*-off cells, and finally *CD81*-off cells (right bar). The asterisks denote the population of cells with the marked gene turned off. **FIG. 1K:** a representative flow cytometry plot of cells targeted for *ITGB1*, *CD81*, and *CD151* silencing. Cells were first gated on *ITGB1* silencing and the represented population displays *CD81* and *CD151* silencing. **FIG. 1L:** a histogram plot of *CLTA* expression at 15 months p.t. showing 38 clones that retained *CLTA* repression and one clone that reactivated *CLTA* expression. The mean values in FIGS. 1C, 1E, 1H, 1I, 1J, and 1K were measured from three independent experiments. Error bars represent SD of the mean.

[0011] **FIGS. 2A-2H** show highly specific and robust transcriptional silencing by CRISPRoff. **FIGS. 2A-2D:** RNA-seq plots of HEK293T cells transfected with CRISPRoff and non-targeting (NT) sgRNAs compared to sgRNAs targeting (**FIG. 2B**) *ITGB1*, (**FIG. 2C**) *CD81*, or (**FIG. 2D**) *CD151*. A comparison of untransfected cells and CRISPRoff with NT sgRNA is shown in (**FIG. 2A**). The volcano plots (bottom) display the targeted genes as the most significantly repressed transcripts globally. The data are representative of the average of two independent replicates. **FIG. 2E:** a Manhattan plot displaying differentially methylated CpGs between cells treated with CRISPRoff and *CLTA*-targeting or NT sgRNAs (30 days post-transfection) analyzed by WGBS. Red dots represent CpGs that gained DNA methylation in targeting sgRNA cells and blue dots represent CpGs that gained DNA methylation in NT sgRNA cells. The arrow denotes the genomic position of *CLTA*. **FIG. 2F:** a comparison of CpG methylation along a 55 kb window that includes the *CLTA* locus. Tracks labelled 'Untr.' represent untransfected cells; the 'NT' tracks represent cells transfected with CRISPRoff and non-targeting sgRNA; the 'T' tracks represent cells transfected with CRISPRoff and targeting sgRNA. R1 and R2 represent two technical replicates. Red marks represent methylated (beta-value >0.5) and the blue marks represent unmethylated (<0.5) CpG dinucleotides. CpG islands are shown in green. **FIG. 2G:** a

comparison of H3K9me3 ChIP-seq signal across the *H2B* gene in cells transfected with CRISPRoff and *H2B*-targeting (purple) or NT sgRNAs (blue) taken at 5 days and 30 days p.t. The sgRNA binding site is denoted along with the CpG islands and neighboring genes. The *BOLA1* gene contains two annotated transcriptional start sites, labeled TSS1 and TSS2. **FIG. 2H:** volcano plot comparing H3K9me3 ChIP-seq data between CRISPRoff transfected with either *H2B*-targeting or NT sgRNAs. Red dots highlight the genes proximal to the *H2B* target.

[0012] **FIGS. 3A-3D** show that *Flavivirus* infections can be blocked by epigenome editing. **FIG. 3A** is a schematic showing epigenome editing of HEK293T cells for blocking viral infections. **FIG. 3B** YFV infection and detected a 50% reduction in YFV infection of CLTA-off cells (right panel) compared to WT HEK293T (left panel). **FIG. 3C** shows the infection of DENV-2 is reduced significantly in SPCS1 and STT3A targeted cells. **FIG. 3D** shows gene silencing of SPCS1 and STT3A mirrored the DENV-2 infection in edited cells.

[0013] **FIGS. 4A-4G** show genome-wide gene silencing by CRISPRoff. **FIG. 4A:** a schematic of the dual sgRNA lentiviral vector used in the CRISPRoff genome-wide screens that contains two unique sgRNAs targeting the same gene. **FIG. 4B:** a schematic of a pooled genome-wide screen to determine the targeting landscape of CRISPRoff. **FIG. 4C:** a time course of CLTA expression in HEK293T after transfection of dCas9-KRAB (gray), CRISPRoff-V2 (black), or mutant CRISPRoff-D3A^{E765A} (orange). **FIG. 4D:** a comparison of phenotype scores (γ) between CRISPRoff (y-axis) and CRISPRoff mutant (x-axis) screens. Three types of expected negative controls are highlighted as negative control pseudo-genes (blue), olfactory genes (orange), and Y chromosome genes (green). **FIG. 4E:** a violin plot of the phenotype scores (γ) for genes defined as essential or nonessential from DepMap. Each replicate screen is plotted for CRISPRoff (green) and CRISPRoff mutant (orange). **FIG. 4F:** a plot of true and false positive rates of genes defined as essential by DepMap. **FIG. 4G:** a plot illustrating the distance of an essential gene hit, defined as having a $\gamma \leq -0.2$, from the nearest essential gene hit. Each dot corresponds to a gene hit's nearest neighboring essential gene, with the x-axis showing the distance between the two genes and the y-axis as the neighboring gene's phenotype score.

[0014] **FIGS. 5A-5K** show CRISPRoff-mediated silencing of genes without promoter CpG island annotations. **FIG. 5A:** a plot comparing the phenotype score of genes between the CRISPRoff and CRISPRoff mutant screens with genes that lack a CGI annotation highlighted in red. **FIG. 5B:** histograms of mNeonGreen fluorescence of five HEK293T cell lines, each with the indicated gene endogenously tagged with split mNeonGreen. **FIG. 5C:** quantification of cells with CALD1, DYNC2LI1, LAMP2, MYL6, or VPS25 silenced after CRISPRoff or CRISPRoff mutant treatment. The data were measured at 14 days p.t., except for VPS25 which

was collected at 11 days p.t. due to a growth defect upon gene knockdown. **FIG. 5D:** quantification of percent of cells with *DYNC2LI1* or *LAMP2* reactivated after TETv4 treatment with targeting or non-targeting sgRNAs, obtained at 14 days p.t. **FIG. 5E:** CpG methylation profiling within the *LAMP2*, *DYNC2LI1*, and *MYL6* promoters after CRISPRoff treatments. White circles represent the CpG methylation status of untransfected HEK293T cells. Each dot is an average of eight independent clones. **FIGS. 5F-5H:** time course plots of *DYNC2LI1* (**FIG. 5F**), *LAMP2* (**FIG. 5G**), and *MYL6* (**FIG. 5H**) expression after transfection of either CRISPRoff or CRISPRoff mutant. Error bars represent the SD of three independent replicates. **FIG. 5I:** a histogram of *DYNC2LI1* expression in 33 clonal lines, measured at 50 days p.t. A positive control of untransfected cells is labeled. **FIG. 5J:** a Manhattan plot displaying differentially methylated CpGs between cells treated with CRISPRoff and either *DYNC2LI1*-targeting or NT sgRNA, as analyzed by WGBS. The arrow points to the genomic location of *DYNC2LI1*. **FIG. 5K:** a view of a 10 kb genomic window containing the *DYNC2LI1* locus, highlighting gain of CpG methylation (red) at the promoter in cells transfected with CRISPRoff and *DYNC2LI1*-targeting sgRNAs.

[0015] **FIGS. 6A-6L** show pooled sgRNA tiling screens that reveal a wide targetable window of CRISPRoff-mediated gene repression. **FIG. 6A:** a schematic of the sgRNA library that tiles PAM-containing sgRNAs within a +/- 1 kb window from annotated transcription start sites (TSS). **FIG. 6B:** a summary of the number of genes per indicated category that comprise the tiling sgRNA library. **FIG. 6C:** a comparison of the phenotype score (γ) for genes with annotated CGI between CRISPRoff (y-axis) and CRISPRoff mutant (x-axis). Each dot is the average of the three most active sgRNAs for each gene. The red dots highlight genes that lack a promoter CGI annotation. **FIG. 6D:** an aggregate plot comparing the normalized phenotype score for each sgRNA targeting genes with one annotated CGI. The green line represents screen data from CRISPRoff in HEK293Ts, orange from CRISPRoff mutant in HEK293Ts, and purple from CRISPRi in K562s. **FIGS. 6E-6G:** representative sgRNA activity score profiles for *DKC1*, *GPN2*, and *ZCCHC9* from the indicated screen (y-axis). The green bar depicts the annotated CGI obtained from UCSC Genome Browser. **FIG. 6H:** representative sgRNA activity score profile for *ORC5* from the indicated screen (y-axis). **FIG. 6I:** an aggregate plot comparing the normalized phenotype score for each sgRNA for genes without annotated CGIs. **FIG. 6J:** an overlay of normalized sgRNA phenotype score from the CRISPRoff screen (green) with MNase signal that represents nucleosome occupancy (gray). The plot is an aggregate of genes with one annotated CGI. **FIG. 6K:** an overlay of normalized sgRNA phenotype score from the CRISPRoff screen (green) with MNase signal that represents nucleosome occupancy (gray). The

plot is an aggregate of the 39 genes with no annotated CGI. **FIG. 6L:** a plot of sgRNA activity along with MNase signal for *H2B*, derived from the sgRNA tiling screen outlined in Figure S6F.

[0016] FIGS. 7A-7J: show CRISPRoff gene silencing in iPSCs, iPSC-derived neurons, and enhancers. **FIG. 7A:** an experimental workflow of CD81 knockdown by CRISPRoff in iPSCs, followed by NGN2-mediated differentiation of edited cells into neurons. **FIG. 7B:** quantification of cells with CD81 silenced by CRISPRi or CRISPRoff with *CD81*-targeting or NT sgRNAs, measured at 30 days p.t. The error bars represent SD from three independent experiments. **FIG. 7C:** quantification of cells with CD81 silenced at the indicated time points from (A). The gray bars indicate the percent of iPSC-edited cells with CD81 silenced that were not differentiated during the experiment. The red bars represent cells that were carried through the neuronal differentiation protocol. The error bars represent SD from three independent experiments. **FIG. 7D:** a representative histogram of CD81 expression at days 8 of neuronal differentiation of parental-unedited (gray) or CD81-edited iPSCs (red). **FIG. 7E:** bisulfite PCR of a 140 bp region of the *CD81* promoter in cells transfected with CRISPRoff and NT or *CD81*-targeting sgRNA. **FIG. 7F:** representative bright field microscopy images of differentiated neurons derived from iPSCs transfected with CRISPRoff and *MAPT*-targeting or NT sgRNA. **FIG. 7G:** quantification of cells with Tau-off in cells transfected with CRISPRoff and NT or *MAPT*-targeting sgRNA, measured at 10 days post-differentiation. **FIG. 7H:** representative flow cytometry plots of Tau protein staining in iPSC-derived neurons after CRISPRoff transfection with NT or *MAPT*-targeting sgRNA. The gates are based on unperturbed iPSC-derived neurons. **FIG. 7I:** a schematic of the *PVT1* locus with the promoter and four enhancer elements (E1-E4) labeled with the distance from the TSS. **FIG. 7J:** plots of normalized *PVT1* transcript levels from quantitative RT-qPCR of cells treated with CRISPRoff (left) or CRISPRoff D3A mutant (right) and sgRNAs targeting either the promoter (Pr.) or the four enhancer elements (E1-E4), normalized to control sgRNAs. Asterisks denote statistical significance by *t*-test and each technical replicate is shown as red dots.

[0017] FIGS. 8A-8K show optimization of CRISPRoff design for durable gene silencing, related to FIG. 1. **FIG. 8A:** a schematic of the CRISPRoff-V1 construct and various linker sequences used to generate protein variants. **FIG. 8B:** a time course of CLTA-GFP silencing after transfection of CRISPRoff-V1 variants or controls dCas9-KRAB (gray) and dCas9-D3A-D3L (orange). **FIG. 8C:** a crystal structure of DNMT3A (orange) and DNMT3L (yellow) in complex with CpG-containing DNA (PDB 5YX2). The arrows point to the dCas9 attachment positions for CRISPRoff-V1 and CRISPRoff-V2. **FIG. 8D:** a schematic of four CRISPRoff-V2 constructs that varies BFP as a linker between dCas9 and KRAB or separated from CRISPRoff

by a P2A sequence. The V2.3 and V2.4 constructs encode NLS sequences at the amino and carboxyl termini of dCas9. **FIG. 8E:** a western blot of dCas9, dCas9-KRAB, CRISPRoff construct protein expression. **FIG. 8F:** a time course of CLTA-GFP silencing after transfection of the V1 and V2.1, V2.3, V2.4 constructs, along with dCas9-KRAB and dCas9 only controls. **FIG. 8G:** a time course of Snrpn-GFP silencing after transfection of dCas9-D3A-3L (orange), dCas9-KRAB (gray), or CRISPRoff-V1 (black) and V2 (blue). **FIG. 8H:** a representative flow cytometry plot of HEK293T CLTA-GFP cells 6 days after transfection of mRNA encoding CRISPRoff. **FIG. 8I:** representative flow cytometry plots of multiplexing gene targeting of two genes simultaneously, measured at 30 days post-transfection. **FIG. 8J:** quantification of gene silencing measured at 31 days post-transfection of CRISPRoff with four simultaneous sgRNAs targeting *ITGB1*, *CLTA*, *CD81*, and *CD151*. **FIG. 8K:** quantification of cells with CLTA, CD81, and CD151 silenced in cells that have ITGB1 either silenced (blue) or unsilenced (gray) in the four gene knockdown experiment. The error bars in **FIGS. 8B, F, G, J, and K** represent SD from three independent experiments.

[0018] FIGS. 9A-9F: show that CRISPRoff is applicable in various cell lines and with orthogonal RNA-guided CRISPR proteins, related to FIG. 1. **FIG. 9A:** flow cytometry histograms of CRISPRoff expression (BFP) before and after doxycycline (dox) treatment. After 24 hours of dox administration, the media was replaced with media without doxycycline (1d and 2d post dox-wash) to turn off CRISPRoff expression. **FIG. 9B:** quantification of K562 cells with CD81 silenced 10 days after initial dox-induction of CRISPRoff expression. Doxycycline was included in the media for either 3 days (middle) or 4 days (right) prior to washing cells to remove doxycycline. **FIG. 9C:** quantification of cells with ITGB1, CD81, or CD151 silenced in HeLa and U2OS measured at 18 days post-CRISPRoff-V1 transfection with sgRNAs targeting the indicated genes. **FIG. 9D:** a representative flow cytometry plot of CD81 expression in iPS cells after transfection of CRISPRoff with either non-targeting sgRNAs or sgRNAs targeting *CD81*. **FIGS. 9E-9F:** a comparison of cells with CLTA (**FIG. 9E**) and H2B (**FIG. 9F**) silenced 10 days after transfection of CRISPRoff with dCas9 from *S. pyogenes* (dSpyCas9) or *S. aureus* (dSauCas9) or dCas12a from *Lachnospiraceae bacterium* (dLbCas12a). The error bars in **FIGS. 9B, 9C, 9E, and 9F** are SD from three independent experiments.

[0019] FIGS. 10A-10L show transcriptional specificity of CLTA, H2B, and RAB11A silencing, related to FIG 2. **FIG. 10A:** an RNA-sequencing TPM (transcripts per kilobase million) plot for HEK293T cells transfected with CRISPRoff and NT (non-targeting) sgRNAs compared to untransfected HEK293T cells. **FIGS. 10B-10D:** RNA-sequencing TPM (transcripts per kilobase million) are plotted for HEK293T cells transfected with CRISPRoff and either NT

sgRNAs compared to sgRNAs targeting *CLTA* (**FIG. 10B**), *HIST2H2BE* (**FIG. 10C**), or *RAB11A* (**FIG. 10D**). The data are representative of the average of two independent replicates. **FIGS. 10E-10F**: Representation of gene expression changes $\pm$ 1000 kb from the annotated targeted gene for *CD81*, *CD151*, *ITGB1* (**FIG. 10E**) and *RAB11*, *HIST2H2BE*, *CLTA* (**FIG. 10F**). Each box represents a gene. **FIG. 10G**: a barplot of genome-wide average CpG coverage obtained from whole genome bisulfite sequencing. WT indicates untransfected cells; NT indicates CRISPRoff delivered with non-targeting sgRNAs; T indicates CRISPRoff delivered with sgRNAs targeting *CLTA* or *DYNC2LI1*. The experiments were performed in two replicates. *CLTA* experiments are presented in FIG. 2, *DYNC2LI1* experiments are presented in FIG. 5. **FIG. 10H**: a barplot of genome-wide average CpG methylation (beta-values) in the samples described in FIG. 10G. **FIG. 10I**: heatmap plots comparing average CpG methylation (beta-values) of 20-CpG sliding windows between untransfected (WT) replicates (left), WT and NT (middle), and WT and targeting sgRNAs (right) for *CLTA* (top row) and *DYNC2LI1* (bottom row) experiments. Red color indicates highest density and blue color indicates lowest density. White areas indicate the absence of windows with the respective average methylation levels. **FIG. 10J**: a heat map showing pairwise correlations between genome-wide CpG methylation profiles of the samples described in Figure S3G, including replicates. Samples are sorted by unsupervised hierarchical clustering. Dark brown color indicates highest correlation, light yellow color indicates lowest correlation. This analysis highlights the variations in global CpG methylation between samples for *CLTA* and *DYNC2LI1* experiments. **FIG. 10K**: a close up of a 17 kb genomic region containing the *ZSCAN16* gene. DNA methylation profiles for untransfected, NT, and targeting sgRNA cells from the *CLTA* experiment are shown. The box highlights a differentially methylated region at the gene promoter, indicating a gain of CpG methylation in the targeting sgRNA cells compared to the control cells. **FIG. 10L**: a close up of a 30 kb genomic region containing the *RPS6KA6* gene. The box highlights a gain of CpG methylation at the promoter in the targeting sgRNA cells. A CpG island located in the promoter of the gene is indicated in green.

[0020] **FIGS. 11A-11E** show that genome-wide silencing by CRISPRoff is reproducible and specific (related to FIG. 4). **FIG. 11A**: a plot comparing the phenotype score (γ) of genes between technical replicates of the CRISPRoff (left) and CRISPRoff D3A methyltransferase mutant (right) genome-wide screens. The negative control sgRNAs are highlighted in blue. **FIG. 11B**: violin plots of the phenotype score (γ) of all genes from each screen. **FIG. 11C**: a histogram of the number of genes with the indicated phenotype score (γ) from the CRISPRoff and CRISPRoff mutant screens. The light green and light orange lines correspond to the

phenotype scores of negative control sgRNAs. **FIGS. 11D-11E:** gene set enrichment analysis (GSEA) for genes associated with DNA replication and ribosome, confirming enrichment of expected essential genes. Genes are ranked from lowest (red) to highest (blue) phenotype scores.

[0021] **FIGS. 12A-12I** show design and validation of tiling sgRNA screens show flexible targeting genomic window of CRISPRoff activity (related to FIG. 6). **FIG. 12A:** the genes chosen for the sgRNA tiling screens are highlighted on a volcano plot depicting gene phenotype scores from previous genome-wide CRISPRi screens in K562 cells (Horlbeck et al. 2016). The colors represent genes with one (orange), multiple (purple), or no annotated CGI (green). **FIG. 12B:** an aggregate plot comparing the normalized phenotype score for each sgRNA for genes with multiple CGIs. The green line represents screen data from CRISPRoff transfection into HEK293T cells, orange from CRISPRoff mutant into HEK293T, and purple from K562 CRISPRi. **FIG. 12C:** representative sgRNA activity score profile for *TBCD* from the three screens. The green bar depicts the annotated CGIs obtained from UCSC Genome Browser. **FIGS. 12D-12E:** plots overlaying the sgRNA phenotype scores and MNase signals for *GFER* and *IMMT*. **FIG. 12F:** an experimental workflow of tiling sgRNA screens to determine optimal sgRNAs for four endogenously GFP-tagged genes: *CLTA*, *H2B*, *RAB11A*, and *VIM*. The indicated population in the histograms represent the population of cells that have maintained gene silencing 4 weeks after CRISPRoff transfection. **FIGS. 12G-12I:** overlay plots of sgRNA activity and MNase signal for *CLTA*, *RAB11A*, and *VIM* from the sgRNA tiling screen.

[0022] **FIGS. 13A-13I:** show confinement of H3K9me3 and CpG methylation despite distal sites of epigenetic establishment (related to FIG. 6). **FIG. 13A:** a schematic of the H2B promoter with two sgRNA sites annotated: sg-A at the TSS and sg-B located about 2 kb upstream of the TSS. The CpG island spans 1.4 kb. Sites 1, 2, and 3 represent regions probed for CpG methylation by bisulfite PCR as described in FIGS. 13E-13I. **FIG. 13B:** a time course of H2B silencing after transfection of CRISPRoff with sg-A or sg-B. **FIG. 13C:** a comparison of H3K9me3 profiles at the *H2B* (*HIST2H2BE*, colored red) promoter at day 5 (green tracks) and day 30 (purple tracks) post-transfection of CRISPRoff with sg-A, sg-B, or non-targeting (NT) sgRNA. The sgRNA binding sites are labeled, along with CpG island annotations (green), and the basal unmethylated CpG region of the H2B promoter prior to transfection obtained from WGBS of WT untransfected cells in FIG. 2. **FIG. 13D:** a comparison of H3K9me3 profiles at the H2B promoter, as described in C, except in experiments using CRISPRoff with a D3A methyltransferase mutation. **FIGS. 13E-13F:** quantification of CpG methylation at site 1 and site 3 of the *H2B* promoter (labeled in FIG. 13A) in cells transfected with CRISPRoff and either sg-A (blue), sg-B (orange), or non-targeting (gray) sgRNA. The cells were harvested for

bisulfite PCR at day 5 post transfection. **FIGS. 13G-13I:** quantification of CpG methylation at sites 1, 2, and 3 of the *H2B* promoter, obtained at 30 days post transfection.

[0023] In the Figures and Examples, CRISPRoff-V1 refers to SEQ ID NO:1, and CRISPRoff-V2 refers to SEQ ID NO:107 (V2.1), unless otherwise stated.

DETAILED DESCRIPTION

[0024] Definitions

[0025] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0026] The use of a singular indefinite or definite article (e.g., “a,” “an,” “the,” etc.) in this disclosure and in the following claims follows the traditional approach in patents of meaning “at least one” unless in a particular instance it is clear from context that the term is intended in that particular instance to mean specifically one and only one. Likewise, the term “comprising” is open ended, not excluding additional items, features, components, etc.

[0027] The terms “comprise,” “include,” and “have,” and the derivatives thereof, are used herein interchangeably as comprehensive, open-ended terms. For example, use of “comprising,” “including,” or “having” means that whatever element is comprised, had, or included, is not the only element encompassed by the subject of the clause that contains the verb.

[0028] For specific proteins described herein (e.g., KRAB, dCas9, Dnmt3A, Dnmt3L), the named protein includes any of the protein's naturally occurring forms, or variants or homologs that maintain the protein activity (e.g., within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to the native protein). In aspects, variants or homologs have at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring form. In aspects, the protein is the protein as identified by its NCBI sequence reference. In aspects, the protein is the protein as identified by its NCBI sequence reference or functional fragment or homolog thereof.

[0029] The term “nuclease-deficient RNA-guided DNA endonuclease enzyme” and the like refer, in the usual and customary sense, to an RNA-guided DNA endonuclease (e.g. a mutated form of a naturally occurring RNA-guided DNA endonuclease) that targets a specific phosphodiester bond within a DNA polynucleotide, wherein the recognition of the phosphodiester bond is facilitated by a separate polynucleotide sequence (for example, a RNA sequence (e.g., single guide RNA (sgRNA))), but is incapable of cleaving the target phosphodiester bond to a significant degree (e.g. there is no measurable cleavage of the phosphodiester bond under physiological conditions). A nuclease-deficient RNA-guided DNA endonuclease thus retains DNA-binding ability (e.g. specific binding to a target sequence) when complexed with a polynucleotide (e.g., sgRNA), but lacks significant endonuclease activity (e.g. any amount of detectable endonuclease activity). In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, dCpf1, ddCpf1, Cas-phi, a nuclease-deficient Cas9 variant, a nuclease-deficient Class II CRISPR endonuclease, a zinc finger domain, a transcription activator-like effector (TALE), a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a zinc finger domain, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a leucine zipper domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a winged helix domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a helix-turn-helix motif. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a helix-loop-helix domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is an HMB-box domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a Wor3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is an OB-fold domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is an immunoglobulin domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, ddCpf1, Cas-phi, a nuclease-deficient Cas9 variant, or a nuclease-deficient Class II CRISPR endonuclease. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 from *S. pyogenes*. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 from *S. aureus*. In aspects, the nuclease-

deficient RNA-guided DNA endonuclease enzyme is dCas12a from *Lachnospiracea* (dLbCas12a). In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas12a from *Lachnospiracea bacterium*. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas12a. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas12. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is ddCas12a. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is Cas-phi.

[0030] The term “CRISPR-associated protein” or “CRISPR protein” refers to any CRISPR protein that functions as a nuclease-deficient RNA-guided DNA endonuclease enzyme, i.e., a CRISPR protein in which catalytic sites for endonuclease activity are defective or lack activity. Exemplary CRISPR-associated proteins include dCas9, dCpf1, dCas12, Cas-phi, a nuclease-deficient Cas9 variant, a nuclease-deficient Class II CRISPR endonuclease, and the like.

[0031] A “CRISPR associated protein 9,” “Cas9,” “Csn1” or “Cas9 protein” as referred to herein includes any of the recombinant or naturally-occurring forms of the Cas9 endonuclease or variants or homologs thereof that maintain Cas9 endonuclease enzyme activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Cas9). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring Cas9 protein. In aspects, the Cas9 protein is substantially identical to the protein identified by the UniProt reference number Q99ZW2 or a variant or homolog having substantial identity thereto. In aspects, the Cas9 protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q99ZW2. In aspects, the Cas9 protein has at least 80% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q99ZW2. In aspects, the Cas9 protein has at least 85% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q99ZW2. In aspects, the Cas9 protein has at least 90% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q99ZW2. In aspects, the Cas9 protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q99ZW2.

[0032] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9. The terms “dCas9” or “dCas9 protein” as referred to herein is a Cas9 protein in which both catalytic sites for endonuclease activity are defective or lack activity. In aspects, the dCas9 protein has mutations at positions corresponding to D10A and H840A of *S. pyogenes* Cas9. In

aspects, the dCas9 protein lacks endonuclease activity due to point mutations at both endonuclease catalytic sites (RuvC and HNH) of wild type Cas9. The point mutations can be D10A and H840A. In aspects, the dCas9 has substantially no detectable endonuclease (*e.g.*, endodeoxyribonuclease) activity. In aspects, dCas9 includes the amino acid sequence of SEQ ID NO:23. In aspects, dCas9 has the amino acid sequence of SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:23. In aspects, dCas9 has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:23. In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 from *S. pyogenes*. In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 from *S. aureus*.

[0033] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is “ddCpf1” or “ddCas12a”. The terms “DNAse-dead Cpf1” or “ddCpf1” refer to mutated *Acidaminococcus sp.* Cpf1 (AsCpf1) resulting in the inactivation of Cpf1 DNase activity. In aspects, ddCpf1 includes an E993A mutation in the RuvC domain of AsCpf1. In aspects, the ddCpf1 has substantially no detectable endonuclease (*e.g.*, endodeoxyribonuclease) activity. In aspects, ddCpf1 includes the amino acid sequence of SEQ ID NO:34. In aspects, ddCpf1 has the amino acid sequence of SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:34. In aspects, ddCpf1 has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:34. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas12a from *Lachnospiraceae bacterium*.

[0034] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dLbCpf1. The term “dLbCpf1” refers to mutated Cpf1 from *Lachnospiraceae bacterium* ND2006 (LbCpf1) that lacks DNase activity. In aspects, dLbCpf1 includes a D832A mutation. In

aspects, the dLbCpf1 has substantially no detectable endonuclease (e.g., endodeoxyribo-nuclease) activity. In aspects, dLbCpf1 includes the amino acid sequence of SEQ ID NO:35. In aspects, dLbCpf1 has the amino acid sequence of SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:35. In aspects, dLbCpf1 has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:35.

[0035] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dFnCpf1. The term “dFnCpf1” refers to mutated Cpf1 from *Francisella novicida* U112 (FnCpf1) that lacks DNase activity. In aspects, dFnCpf1 includes a D917A mutation. In aspects, the dFnCpf1 has substantially no detectable endonuclease (e.g., endodeoxyribo-nuclease) activity. In aspects, dFnCpf1 includes the amino acid sequence of SEQ ID NO: 36. In aspects, dFnCpf1 has the amino acid sequence of SEQ ID NO: 36. In aspects, dFnCpf1 has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:36. In aspects, dFnCpf1 has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:36. In aspects, dFnCpf1 has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:36. In aspects, dFnCpf1 has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:36. In aspects, dFnCpf1 has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:36. In aspects, dFnCpf1 has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:36.

[0036] A “Cpf1” or “Cpf1 protein” as referred to herein includes any of the recombinant or naturally-occurring forms of the Cpf1 (CRISPR from *Prevotella* and *Francisella* 1) endonuclease or variants or homologs thereof that maintain Cpf1 endonuclease enzyme activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Cpf1). In aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring Cpf1 protein. In aspects, the Cpf1 protein is substantially identical to the protein identified by the UniProt reference number U2UMQ6 or a variant or homolog having substantial identity thereto.

In aspects, the Cpf1 protein is identical to the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein has at least 80% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein is identical to the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein has at least 85% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein is identical to the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein has at least 90% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein is identical to the protein identified by the UniProt reference number U2UMQ6. In aspects, the Cpf1 protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number U2UMQ6.

[0037] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a nuclease-deficient Cas9 variant. The term “nuclease-deficient Cas9 variant” refers to a Cas9 protein having one or more mutations that increase its binding specificity to PAM compared to wild type Cas9 and further include mutations that render the protein incapable of or having severely impaired endonuclease activity. Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). The binding specificity of nuclease-deficient Cas9 variants to PAM can be determined by any method known in the art. Descriptions and uses of known Cas9 variants may be found, for example, in Shmakov et al., Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 15, 2017 and Cebrian-Serrano et al, CRISPR–Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* 7-8, 2017. Exemplary Cas9 variants are listed in the Table 1 below.

[0038] Table 1

Cas9 Variants	PAM domains	References
<i>Strep pyogenes</i> (Sp) Cas9	NGG	Hsu et al. 2014 Cell
<i>Staph aureus</i> (Sa) Cas9	NNGRRT or NNGRR NNGGGT, NNGAAT, NNGAGT (Zetsche)	Ran et al. 2015 Nature

SpCas9 VQR mutant (D1135V, R1335Q, T1337R)	NGAG>NGAT=NGAA>NGAC NGCG	Kleinstiver et al. 2015 Nature
SpCas9 VRER mutant (D1135V/G1218R/R1335E/T1337R)	NGCG	Kleinstiver et al. 2015 Nature
SpCas9 D1135E	NGG, greater fidelity, less cutting at NAG and NGA sites	Kleinstiver et al. 2015 Nature
eSpCas9 1.1 mutant (K848A/K1003A/R1060A)	NGG	Slaymaker et al. Science 2015
SpCas9 HF1 (Q695A, Q926A, N497A, R661A)	NGG	Kleinstiver et al. 2016 Nature
AsCpfI	TTTN (5' of sgRNA)	Zetsche et al. 2015 Cell
HypaCas9 (N692A, M694A, Q695A, H698A)		Chen et al., Nature volume 550, pages 407– 410 (19 October 2017)

[0039] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a nuclease-deficient Class II CRISPR endonuclease. The term “nuclease-deficient Class II CRISPR endonuclease” as used herein refers to any Class II CRISPR endonuclease having mutations resulting in reduced, impaired, or inactive endonuclease activity.

[0040] The term “nuclease-deficient DNA endonuclease enzyme” refers to a DNA endonuclease (e.g. a mutated form of a naturally occurring DNA endonuclease) that targets a specific phosphodiester bond within a DNA polynucleotide, but that does not require an RNA guide. In embodiments, the “nuclease-deficient DNA endonuclease enzyme” is a zinc finger domain or a TALE.

[0041] In embodiments, the nuclease-deficient DNA endonuclease enzyme is a “zinc finger domain.” The term “zinc finger domain” or “zinc finger binding domain” or “zinc finger DNA binding domain” are used interchangeably and refer to a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. In embodiments, the zinc finger domain is non-naturally occurring in that it is engineered to bind to a target site of choice. In aspects, the zinc finger binding domain refers to a protein, a domain within a larger protein, or a nuclease-deficient RNA-guided DNA endonuclease enzyme that is capable of binding to any zinc finger known in the art, such as the C2H2 type, the CCHC type, the PHD type, or the RING type of zinc fingers.

[0042] As used herein, a “zinc finger” is a polypeptide structural motif folded around a bound zinc cation. In embodiments, the polypeptide of a zinc finger has a sequence of the form

X₃-Cys-X₂₋₄-Cys-X₁₂-His-X₃₋₅-His-X₄, wherein X is any amino acid (e.g., X₂₋₄ indicates an oligopeptide 2-4 amino acids in length). There is generally a wide range of sequence variation in the 28-31 amino acids of the known zinc finger polypeptides. Only the two consensus histidine residues and two consensus cysteine residues bound to the central zinc atom are invariant. Of the remaining residues, three to five are highly conserved, while there may be significant variation among the other residues. Despite the wide range of sequence variation in the polypeptide, zinc fingers of this type have a similar three dimensional structure. However, there is a wide range of binding specificities among the different zinc fingers, i.e. different zinc fingers bind double stranded polynucleotides having a wide range of nucleotides sequences. In aspects, the zinc finger is the C2H2 type. In aspects, the zinc finger is the CCHC type. In aspects, the zinc finger is the PHD type. In aspects, the zinc finger is the RING type.

[0043] In embodiments, the nuclease-deficient DNA endonuclease enzyme is a TALE. “TALE” or “transcription activator-like effector” refer to artificial restriction enzymes generated by fusing the TAL effector DNA binding domain to a DNA cleavage domain. TALEs enable efficient, programmable, and specific DNA cleavage and represent powerful tools for genome editing in situ. Transcription activator-like effectors (TALEs) can be quickly engineered to bind practically any DNA sequence. The term TALE, as used herein, is broad and includes a monomeric TALE that can cleave double stranded DNA without assistance from another TALE. The term TALE is also used to refer to one or both members of a pair of TALEs that are engineered to work together to cleave DNA at the same site. TALEs that work together may be referred to as a left-TALE and a right-TALE, which references the handedness of DNA. TALE are proteins secreted by *Xanthomonas* bacteria. The DNA binding domain contains a highly conserved 33-34 amino acid sequence with the exception of the 12th and 13th amino acids. These two locations are highly variable (repeat variable diresidue (RVD)) and show a strong correlation with specific nucleotide recognition. This simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA binding domains by selecting a combination of repeat segments containing the appropriate RVDs.

[0044] The term “Krüppel associated box domain” or “KRAB domain” as provided herein refers to a category of transcriptional repression domains present in approximately 400 human zinc finger protein-based transcription factors. KRAB domains typically include about 45 to about 75 amino acid residues. A description of KRAB domains, including their function and use, may be found, for example, in Ecco, G., Imbeault, M., Trono, D., KRAB zinc finger proteins, *Development* 144, 2017; Lambert et al. The human transcription factors, *Cell* 172, 2018; Gilbert et al., *Cell* (2013); and Gilbert et al., *Cell* (2014). In aspects, the KRAB domain is a KRAB

domain of Kox 1. In aspects, the KRAB domain includes the sequence set forth by SEQ ID NO:16. In aspects, the KRAB domain is the sequence of SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:16. In embodiments, the KRAB domain is a ZIM3 KRAB domain or an amino acid sequence having 85%, 90%, or 95% sequence identity thereto. In embodiments, the KRAB domain is KOX1 or an amino acid sequence having 85%, 90%, or 95% sequence identity thereto. Gilbert et al, Cell, 159:647-661 (2014); Alerasool et al, Nature Methods, (2020).

[0045] The term “DNA methyltransferase” as provided herein refers to an enzyme that catalyzes the transfer of a methyl group to DNA. Non-limiting examples of DNA methyltransferases include Dnmt1, Dnmt3A, and Dnmt3B. In aspects, the DNA methyltransferase is mammalian DNA methyltransferase. In aspects, the DNA methyltransferase is human DNA methyltransferase. In aspects, the DNA methyltransferase is mouse DNA methyltransferase. In aspects, the DNA methyltransferase is a bacterial cytosine methyltransferase and/or a bacterial non-cytosine methyltransferase. Depending on the specific DNA methyltransferase, different regions of DNA are methylated. For example, Dnmt3A typically targets CpG dinucleotides for methylation. Through DNA methylation, DNA methyltransferases can modify the activity of a DNA segment (e.g., gene expression) without altering the DNA sequence. In aspects, DNA methylation results in repression of gene transcription and/or modulation of methylation sensitive transcription factors or CTCF. As described herein, fusion proteins may include one or more (e.g., two) DNA methyltransferases. When a DNA methyltransferase is included as part of a fusion protein, the DNA methyltransferase may be referred to as a “DNA methyltransferase domain.” In aspects, a DNA methyltransferase domain includes one or more DNA methyltransferases. In aspects, a DNA methyltransferase domain includes two DNA methyltransferases. In aspects, the DNA methyltransferase domain further comprises a catalytically inactive regulatory factor of DNA methyltransferase (e.g., Dnmt3L) that is essential for the functioning of Dnmt1, Dnmt3A, and

Dnmt3B. In aspects, the DNA methyltransferase domain comprises Dnmt1. In aspects, the DNA methyltransferase domain comprises Dnmt3B. In aspects, the DNA methyltransferase domain comprises Dnmt3A. In aspects, the DNA methyltransferase domain further comprises Dnmt3L. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain is Dnmt3L. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain includes Dnmt3A and Dnmt3L. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence

identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain further comprises the Dnmt3L regulatory factor, as described, for example, in Siddique et al, Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity, J. Mol. Biol. 425, 2013 and Stepper et al, Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase, Nucleic Acids Res. 45, 2017.

[0046] A "Dnmt3A", "Dnmt3a," "DNA (cytosine-5)-methyltransferase 3A" or "DNA methyltransferase 3a" protein as referred to herein includes any of the recombinant or naturally-occurring forms of the Dnmt3A enzyme or variants or homologs thereof that maintain Dnmt3A enzyme activity (e.g. within at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Dnmt3A). In aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g., a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring Dnmt3A protein. In aspects, the Dnmt3A protein is substantially identical to the protein identified by the UniProt reference number Q9Y6K1 or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3A polypeptide is encoded by a nucleic acid sequence identified by the NCBI reference sequence Accession number NM_022552, homologs or functional fragments thereof. In aspects, Dnmt3A includes the sequence set forth by SEQ ID NO:26. In aspects, Dnmt3A is the sequence set forth by SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:26. In aspects, Dnmt3A is s Dnmt3A transcript 1 (Dnmt3A1). In aspects, Dnmt3A is s Dnmt3A transcript 2 (Dnmt3A2). In aspects, Dnmt3A is s Dnmt3A transcript 3 (Dnmt3A3). In aspects, is s Dnmt3A transcript 4 (Dnmt3A4).

[0047] A "Dnmt3L", "DNA (cytosine-5)-methyltransferase 3L" or "DNA methyltransferase

3L" protein as referred to herein includes any of the recombinant or naturally-occurring forms of the Dnmt3L regulatory factor or variants or homologs thereof that maintain Dnmt3L regulatory activity (e.g., within at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Dnmt3L). In aspects, the variants or homologs have at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence compared to a naturally occurring Dnmt3L protein. In aspects, the Dnmt3L protein is substantially identical to the protein identified by the UniProt reference number Q9CWR8 or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3L protein is identical to the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 80% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 85% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8.

[0048] In aspects, the Dnmt3L protein is substantially identical to the protein identified by the UniProt reference number Q9UJW or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3L protein is identical to the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 50% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 55% sequence identity to the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 60% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 65% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 70% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 80% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 85% sequence identity to the amino acid sequence of the protein

identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 90% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L polypeptide is encoded by a nucleic acid sequence identified by the NCBI reference sequence Accession number NM_001081695, or homologs or functional fragments thereof. In aspects, Dnmt3L includes the sequence set forth by SEQ ID NO:28. In aspects, Dnmt3L is the sequence set forth by SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 50% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 55% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 60% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 65% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:28.

[0049] The term “RNA-guided DNA endonuclease” and the like refer, in the usual and customary sense, to an enzyme that cleave a phosphodiester bond within a DNA polynucleotide chain, wherein the recognition of the phosphodiester bond is facilitated by a separate RNA sequence (for example, a single guide RNA).

[0050] The term “Class II CRISPR endonuclease” refers to endonucleases that have similar endonuclease activity as Cas9 and participate in a Class II CRISPR system. An example Class II CRISPR system is the type II CRISPR locus from *Streptococcus pyogenes* SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30 bp each). The Cpf1 enzyme belongs to a putative type V CRISPR-Cas system. Both type II and type V systems are included in Class II of the CRISPR-Cas system.

[0051] A “nuclear localization sequence” or “nuclear localization signal” or “NLS” is a peptide that directs proteins to the nucleus. In aspects, the NLS includes five basic, positively charged amino acids. The NLS may be located anywhere on the peptide chain. In aspects, the NLS is an NLS derived from SV40. In aspects, the NLS includes the sequence set forth by SEQ ID NO:25. In aspects, the NLS is the sequence set forth by SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence of SEQ ID NO:25.

[0052] In embodiments, the DNA methyltransferase domain is a Dnmt3A-3L domain. A “Dnmt3A-3L domain” as provided herein refers to a protein including both Dnmt3A (i.e., the DNA methyltransferase) and Dnmt3L (i.e., the catalytically inactive regulatory factor of DNA methyltransferase). In aspects, Dnmt3A and Dnmt3L are covalently linked. In aspects, Dnmt3A is covalently linked to Dnmt3L through a peptide linker. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:27. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:27. In aspects, the peptide linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:27. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:27. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:27. In aspects, the Dnmt3A-3L domain includes the sequence set forth by SEQ ID NO:33. In aspects, the Dnmt3A-3L domain is the sequence set forth by SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 90%

sequence identity to SEQ ID NO:33. In aspects, the Dnmt3A-3L domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:33.

[0053] The term “Krüppel associated box domain” or “KRAB domain” as provided herein refers to a category of transcriptional repression domains present in approximately 400 human zinc finger protein-based transcription factors. KRAB domains typically include about 45 to about 75 amino acid residues. A description of KRAB domains, including their function and use, may be found, for example, in Ecco, G., Imbeault, M., Trono, D., KRAB zinc finger proteins, *Development* 144, 2017; Lambert et al. The human transcription factors, *Cell* 172, 2018; Gilbert et al., *Cell* (2013); and Gilbert et al., *Cell* (2014). In aspects, the KRAB domain is a KRAB domain of Kox 1. In aspects, the KRAB domain includes the sequence set forth by SEQ ID NO:16. In aspects, the KRAB domain is the sequence of SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:16. In aspects, the KRAB domain includes an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:16.

[0054] The term “DNA methyltransferase” as provided herein refers to an enzyme that catalyzes the transfer of a methyl group to DNA. Non-limiting examples of DNA methyltransferases include Dnmt1, Dnmt3A, and Dnmt3B. In aspects, the DNA methyltransferase is mammalian DNA methyltransferase. In aspects, the DNA methyltransferase is human DNA methyltransferase. In aspects, the DNA methyltransferase is mouse DNA methyltransferase. In aspects, the DNA methyltransferase is a bacterial cytosine methyltransferase and/or a bacterial non-cytosine methyltransferase. Depending on the specific DNA methyltransferase, different regions of DNA are methylated. For example, Dnmt3A typically targets CpG dinucleotides for methylation. Through DNA methylation, DNA methyltransferases can modify the activity of a DNA segment (e.g., gene expression) without altering the DNA sequence. In aspects, DNA methylation results in repression of gene transcription and/or modulation of methylation sensitive transcription factors or CTCF. As described herein, fusion proteins may include one or more (e.g., two) DNA methyltransferases. When a DNA methyltransferase is included as part of a fusion protein, the DNA

methyltransferase may be referred to as a “DNA methyltransferase domain.” In aspects, a DNA methyltransferase domain includes one or more DNA methyltransferases. In aspects, a DNA methyltransferase domain includes two DNA methyltransferases. In aspects, the DNA methyltransferase domain further comprises a catalytically inactive regulatory factor of DNA methyltransferase (e.g., Dnmt3L) that is essential for the functioning of Dnmt1, Dnmt3A, and Dnmt3B. In aspects, the DNA methyltransferase domain comprises Dnmt1. In aspects, the DNA methyltransferase domain comprises Dnmt3B. In aspects, the DNA methyltransferase domain comprises Dnmt3A. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain is Dnmt3L. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:28. In aspects, the DNA methyltransferase domain includes Dnmt3A and Dnmt3L. In aspects, the DNA methyltransferase domain has the amino acid sequence of SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to

SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:33. In aspects, the DNA methyltransferase domain further comprises the Dnmt3L regulatory factor, as described, for example, in Siddique et al, Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity, J. Mol. Biol. 425, 2013 and Stepper et al, Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase, Nucleic Acids Res. 45, 2017.

[0055] A "Dnmt3A", "Dnmt3a," "DNA (cytosine-5)-methyltransferase 3A" or "DNA methyltransferase 3a" protein as referred to herein includes any of the recombinant or naturally-occurring forms of the Dnmt3A enzyme or variants or homologs thereof that maintain Dnmt3A enzyme activity (e.g. within at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Dnmt3A). In aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g., a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring Dnmt3A protein. In aspects, the Dnmt3A protein is substantially identical to the protein identified by the UniProt reference number Q9Y6K1 or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3A polypeptide is encoded by a nucleic acid sequence identified by the NCBI reference sequence Accession number NM_022552, homologs or functional fragments thereof. In aspects, Dnmt3A includes the sequence set forth by SEQ ID NO:26. In aspects, Dnmt3A is the sequence set forth by SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:26. In aspects, the DNA methyltransferase domain has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:26. In aspects, Dnmt3A has an amino acid sequence that has at least 95% sequence

identity to SEQ ID NO:26.

[0056] A "Dnmt3L", "DNA (cytosine-5)-methyltransferase 3L" or "DNA methyltransferase 3L" protein as referred to herein includes any of the recombinant or naturally-occurring forms of the Dnmt3L regulatory factor or variants or homologs thereof that maintain Dnmt3L regulatory activity (e.g., within at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Dnmt3L). In aspects, the variants or homologs have at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence compared to a naturally occurring Dnmt3L protein. In aspects, the Dnmt3L protein is substantially identical to the protein identified by the UniProt reference number Q9CWR8 or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3L protein is identical to the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 80% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 85% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8. In aspects, the Dnmt3L protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9CWR8.

[0057] In aspects, the Dnmt3L protein is substantially identical to the protein identified by the UniProt reference number Q9UJW or a variant or homolog having substantial identity thereto. In aspects, the Dnmt3L protein is identical to the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 50% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 55% sequence identity to the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 60% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 65% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 70% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 75% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 80% sequence identity to the amino acid

sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 85% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 90% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L protein has at least 95% sequence identity to the amino acid sequence of the protein identified by the UniProt reference number Q9UJW. In aspects, the Dnmt3L polypeptide is encoded by a nucleic acid sequence identified by the NCBI reference sequence Accession number NM_001081695, or homologs or functional fragments thereof. In aspects, Dnmt3L includes the sequence set forth by SEQ ID NO:28. In aspects, Dnmt3L is the sequence set forth by SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 50% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 55% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 60% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 65% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 70% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:28. In aspects, Dnmt3L has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:28.

[0058] The term “RNA-guided DNA endonuclease” and the like refer, in the usual and customary sense, to an enzyme that cleave a phosphodiester bond within a DNA polynucleotide chain, wherein the recognition of the phosphodiester bond is facilitated by a separate RNA sequence (for example, a single guide RNA).

[0059] A “nuclear localization sequence” or “nuclear localization signal” or “NLS” is a peptide that directs proteins to the nucleus. In aspects, the NLS includes five basic, positively charged amino acids. The NLS may be located anywhere on the peptide chain. In aspects, the NLS is an NLS derived from SV40. In aspects, the NLS includes the sequence set forth by SEQ ID NO:25. In aspects, the NLS is the sequence set forth by SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%,

91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO: 25. In aspects, NLS has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:25. In aspects, NLS has an amino acid sequence of SEQ ID NO:25.

[0060] A "guide RNA" or "gRNA" as provided herein refers to any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In aspects, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more.

[0061] In embodiments, the polynucleotide (e.g., gRNA) is a single-stranded ribonucleic acid. In aspects, the polynucleotide (e.g., gRNA) is from about 10 to about 200 nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is from about 50 to about 150 nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is from about 80 to about 140 nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is from about 90 to about 130 nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is from about 100 to about 120 nucleic acid residues in length. In aspects, the length of the polynucleotide (e.g., gRNA) is about 113 nucleic acid residues in length.

[0062] In general, a guide sequence (i.e., a DNA-targeting sequence) is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence (e.g., a genomic or mitochondrial DNA target sequence) and direct sequence-specific binding of a complex (e.g., CRISPR complex) to the target sequence. In aspects, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In aspects, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is at least about 80%, 85%, 90%, 95%, or 100%. In aspects, the degree of complementarity is at least 90%. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms

based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In aspects, a guide sequence is about or more than about 10, 20, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In aspects, a guide sequence is about 10 to about 50, about 15 to about 30, or about 20 to about 25 nucleotides in length. In aspects, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. In aspects, the guide sequence is about or more than about 20 nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a complex (e.g., CRISPR complex) to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a complex (e.g., CRISPR complex), including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay known in the art. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a complex (e.g., CRISPR complex), including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0063] The terms “sgRNA,” “single guide RNA,” and “single guide RNA sequence” are used interchangeably and refer to the polynucleotide sequence including the crRNA sequence and the tracrRNA sequence. The crRNA sequence includes a guide sequence (i.e., “guide” or “spacer”) and a tracr mate sequence (i.e., direct repeat(s)). The term “guide sequence” refers to the sequence that specifies the target site. In aspects, the two RNA can be encoded separately by a crRNA and tracrRNA as 2 RNA molecules which then form an RNA/RNA complex due to complementary base pairing between the crRNA and tracrRNA (i.e., before being competent to bind to nuclease-deficient RNA-guided DNA endonuclease enzyme). In aspects, a first nucleic acid includes a tracrRNA sequence, and a separate second nucleic acid includes a gRNA sequence lacking a tracrRNA sequence. In aspects, the first nucleic acid including the tracrRNA sequence and the second nucleic acid including the gRNA sequence interact with one another, and optionally are included in a complex (e.g., CRISPR complex). Exemplary sgRNA, and their targeted sequences, are shown in Tables 2, 3, and 4. Exemplary sgRNA described in the examples are shown in Table 5.

[0064] Table 2

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
A (JKNg156)	ACTGCGGAAATTGAGCGT (SEQ ID NO: 37)	ACGCUCAAUUUCCGCAGU (SEQ ID NO: 38)
B (JKNg158)	AGGCAATGGCTGCACATGC (SEQ ID NO: 39)	GCAUGUGCAGCCAUUGCCU (SEQ ID NO: 40)
C (JKNg160)	GACGCTTGTTCTGAGGAG (SEQ ID NO: 41)	CUCCUCAGAACCAAGCGUC (SEQ ID NO: 42)

[0065] Table 3

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
CD29, sgRNA-A	TCCGGAAACGCATTCCTCT (SEQ ID NO: 43)	AGAGGAAUGCGUUUCCGGA (SEQ ID NO: 44)
CD29, sgRNA-B	CCGCGTCAGCCCGGCCCGG (SEQ ID NO: 45)	CCGGGCCGGGCGACGCGG (SEQ ID NO: 46)
CD29, sgRNA-C	CGACTCCCGCTGGGCCTCT (SEQ ID NO: 47)	AGAGGCCCGAGCGGGAGUCG (SEQ ID NO: 48)
CD81, sgRNA-A	CCGTTGCGCGCTCGCTCTC (SEQ ID NO: 49)	GAGAGCGAGCGCGCAACGG (SEQ ID NO: 50)
CD81, sgRNA-B	CCGCGCATCCTGCCAGGCC (SEQ ID NO: 51)	GGCCUGGCAGGAUGCGCGG (SEQ ID NO: 52)
CD81, sgRNA-C	CCAACTTGCGCGTTTCGG (SEQ ID NO: 53)	CCGAAACGCGCCAAGUUGG (SEQ ID NO: 54)
CD151, sgRNA-A	ACCACGCGTCCGAGTCCGG (SEQ ID NO: 55)	CCGGACUCGGACGCGUGGU (SEQ ID NO: 56)
CD151, sgRNA-B	TGCTCATTGTCCCTGGACA (SEQ ID NO: 57)	UGUCCAGGGACAAUGAGCA (SEQ ID NO: 58)
CD151, sgRNA-C	GGACACCCTGCTCATTGTC (SEQ ID NO: 59)	GACAAUGAGCAGGGUGUCC (SEQ ID NO: 60)

[0066] Table 4

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
Pcsk9 sgRNA-1	TCCGGAAACGCATTCCTCT (SEQ ID NO: 43)	AGAGGAAUGCGUUUCCGGA (SEQ ID NO: 44)
Pcsk9 sgRNA-2	ACCGGCAGCCTGCGCGTCC (SEQ ID NO: 61)	GGACGCGCAGGCUGCCGGU (SEQ ID NO: 62)
Pcsk9 sgRNA-3	CGATGGGCACCCACTGCTC (SEQ ID NO: 63)	GAGCAGUGGGUGCCCAUCG (SEQ ID NO: 64)
Pcsk9 sgRNA-4	CCTTCACGTGGACGCGCAG (SEQ ID NO: 65)	CUGCGCGUCCACGUGAAGG (SEQ ID NO: 66)
Pcsk9 sgRNA-5	CGTGAAGGTGGAAGCCTTC (SEQ ID NO: 67)	GAAGGCUUCCACCUUCACG (SEQ ID NO: 68)
Npc1 sgRNA-1	CTCCTTGCTCAGGCGCCGG (SEQ ID NO: 69)	CCGGCGCCUGACCAAGGAG (SEQ ID NO: 70)
Npc1 sgRNA-2	TGGTCAGGCGCCGTTCCG (SEQ ID NO: 71)	CGGAACCGGCGCCUGACCA (SEQ ID NO: 72)
Npc1 sgRNA-3	TAGAGGTCGCCTTCTCCTC (SEQ ID NO: 73)	GAGGAGAAGGCGACCUCUA (SEQ ID NO: 74)

Npc1 sgRNA-4	CGACGCTCGGGTCGCGGTG (SEQ ID NO:75)	CACCGCGACCCGAGCGUCG (SEQ ID NO:76)
Npc1 sgRNA-5	ATGCTGTCGCCGCGCGGGG (SEQ ID NO:77)	CCCCGCGCGGCGACAGCAU (SEQ ID NO:78)
Spes1 sgRNA-1	CTCACCTCACCGGAGCCA (SEQ ID NO:79)	UGGCUCCGGUGAGGGUGAG (SEQ ID NO:80)
Spes1 sgRNA-2	CCGCAAACCTTTACTCCTTA (SEQ ID NO:81)	UAAGGAGUAAAGUUUGCGG (SEQ ID NO:82)
Spes1 sgRNA-3	CTCGGAGACATCCGCTTCC (SEQ ID NO: 60)	GGAAGCGGAUGUCUCCGAG (SEQ ID NO: 60)
Spes1 sgRNA-4	CTCCTAAGATTGGCTTCAC (SEQ ID NO:83)	GUGAAGCCAAUCUUAGGAG (SEQ ID NO:84)
Spes1 sgRNA-5	CCGGAGCCACTCCTAAGAT (SEQ ID NO:85)	AUCUUAGGAGUGGCUCCGG (SEQ ID NO:86)
Cd81 sgRNA-1	TTCTCTACCCTACGTCTCA (SEQ ID NO:87)	UGAGACGUAGGUAGAGAA (SEQ ID NO:88)
Cd81 sgRNA-2	TACGTCTCATTCTCCGCAA (SEQ ID NO:89)	UUGCGGAGAAUGAGACGUA (SEQ ID NO:90)
Cd81 sgRNA-3	GCTAGGCCTCCAGCCCTTC (SEQ ID NO:91)	GAAGGGCUGGAGGCCUAGC (SEQ ID NO:92)
Cd81 sgRNA-4	ACAGGTGGCGCCGCAACTT (SEQ ID NO:93)	AAGUUGCGGCGCCACCUGU (SEQ ID NO:94)
Cd81 sgRNA-5	AGCCGGAGGCGCGAGAGTC (SEQ ID NO:95)	GACUCUCGCGCCUCCGGCU (SEQ ID NO:96)

[0067] Table 5

Name	Protospacer (5' to 3')	SEQ ID NO:
Snrpn promoter	CTCCTCAGAACCAAGCGTC	SEQ ID NO:127
H2B sgRNA-a	GTAAGACACAGTACAAACG	SEQ ID NO:128
H2B sgRNA-b	GAACCGGCAAAATCCGCTC	SEQ ID NO:129
H2B sgRNA-c	GGGCCGGAGCGGATTTTGC	SEQ ID NO:130
CLTA sgRNA-a	GAAGCCCTACCCGTGTATC	SEQ ID NO:131
CLTA sgRNA-b	CTTGTCCTCCTCTCCCAGT	SEQ ID NO:132
CLTA sgRNA-c	CTCCCAGTCGGCACCACAG	SEQ ID NO:133
ITGB1 sgRNA-a	AGAGGAATGCGTTTCCGGA	SEQ ID NO:134
ITGB1 sgRNA-b	CCGGGCCGGGCTGACGCGG	SEQ ID NO:135
ITGB1 sgRNA-c	AGAGGCCAGCGGGAGTCG	SEQ ID NO:136
CD81 sgRNA-a	GAGAGCGAGCGCGCAACGG	SEQ ID NO:137
CD81 sgRNA-b	GGCCTGGCAGGATGCGCGG	SEQ ID NO:138
CD81 sgRNA-c	CCGAAACGCGCCAAGTTGG	SEQ ID NO:139
CD151 sgRNA-a	CCGGACTCGGACGCGTGGT	SEQ ID NO:140
CD151 sgRNA-b	TGTCCAGGGACAATGAGCA	SEQ ID NO:141
CD151 sgRNA-c	GACAATGAGCAGGGTGTCC	SEQ ID NO:142
CALD1 sgRNA-a	TTGGCTGGGGTGTCTCAT	SEQ ID NO:143
CALD1 sgRNA-b	TCTGGGAGGGGCCAAGGAA	SEQ ID NO:144
CALD1 sgRNA-c	GAGATGCCTGTCATTCCCT	SEQ ID NO:145
CALD1 sgRNA-d	CCTCCCCACTGTAAACATA	SEQ ID NO:146
CALD1 sgRNA-e	AGGCAGGCTGCATCCACCT	SEQ ID NO:147

DYNC2LI1 sgRNA-a	GCCGAGAATGAGATGTAAA	SEQ ID NO:148
DYNC2LI1 sgRNA-b	CGCGGTTGCGTGGGGAGAC	SEQ ID NO:149
DYNC2LI1 sgRNA-c	GTGGGGAGACGGGCATCAT	SEQ ID NO:150
DYNC2LI1sgRNA-d	GTCGCGGCCGCGGTTGCGT	SEQ ID NO:151
DYNC2LI1 sgRNA-e	GGGAGCGACGTCGCGGCCG	SEQ ID NO:152
LAMP2 sgRNA-a	TCTCTCAGGAGCATAGGAA	SEQ ID NO:153
LAMP2 sgRNA-b	GAGAGGGGTGGTGATGTAG	SEQ ID NO:154
LAMP2 sgRNA-c	CCGTCCGCGCATATCTCTC	SEQ ID NO:155
LAMP2 sgRNA-d	CGACCACGCCCTGGCTTTT	SEQ ID NO:156
LAMP2 sgRNA-e	CGTTGAAAGCGGCGAGAG	SEQ ID NO:157
MYL6 sgRNA-a	CGGAGGTGGGGGGGGTCGT	SEQ ID NO:158
MYL6 sgRNA-b	GGGCACAGAACCCGCTCGG	SEQ ID NO:159
MYL6 sgRNA-c	ACCCCCCCCACCTCCGAGC	SEQ ID NO:160
MYL6 sgRNA-d	AACCCGCTCGGAGGTGGGG	SEQ ID NO:161
MYL6 sgRNA-e	CACAGAACCCGCTCGGAGG	SEQ ID NO:162
VPS25 sgRNA-a	TAGGATAACAGCGAGGGAC	SEQ ID NO:163
VPS25 sgRNA-b	TGGAAAGAGAGAGACCGGC	SEQ ID NO:164
VPS25 sgRNA-c	AAGGCCGGAAGTCCTTAGC	SEQ ID NO:165
VPS25 sgRNA-d	CCAAAAGATAGGGCTTTTT	SEQ ID NO:166
VPS25 sgRNA-e	AGGCAATGCAGAAAGGTTC	SEQ ID NO:167
PVT1 E1-a	GAGGTCATGATTCATCCAC	SEQ ID NO:168
PVT1 E1-b	GGCTACTGGTCTCCACCGGC	SEQ ID NO:169
PVT1 E1-c	GAAGCTCGCTTACAGGAGGG	SEQ ID NO:170
PVT1 E2-a	GCTGCTGAGCGATGACTCAG	SEQ ID NO:171
PVT1 E2-b	GTCCCTGCGCCAGGGTAAAC	SEQ ID NO:172
PVT1 E2-c	GCAATCTCGGTCCGCCTTAC	SEQ ID NO:173
PVT1 E3-a	GCCGGCTGTTGAGTCAGCTC	SEQ ID NO:174
PVT1 E3-b	GCTGGATGACACCACTCTGT	SEQ ID NO:175
PVT1 E3-c	GCGGTGGAGACCTCATGATC	SEQ ID NO:176
PVT1 E4-a	GCAGCTCTTGTTGCTGCT	SEQ ID NO:177
PVT1 E4-b	GTTAAGATAGAGAATGCCAT	SEQ ID NO:178
PVT1 E4-c	GTTCCGTACACAGTCCTTTC	SEQ ID NO:179
PVT1 promoter	GCCTCCGGGCAGAGCGCGTG	SEQ ID NO:180
Lambda negative control-1	GATCGTCGGTCGGGTCATACG	SEQ ID NO:181
Lambda negative control-2	GTTGTCGCGATAGATGATCG	SEQ ID NO:182
Lambda negative control-3	GCGCGTATCGTTTCATCGGCG	SEQ ID NO:183

[0068] The sequences in the Tables are the targeting crRNA sequences. As an example, the full single guide RNA (sgRNA) for SEQ ID NO:38 is:
GACGCUAAAUUUCCGAGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGU
UUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU
UUUUU (SEQ ID NO:100). A common tracr sequence of each single guide for Sp Cas9 is
GUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUU
AUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU (SEQ ID NO:101). The skilled artisan will appreciate that the sgRNA sequences in the Tables are 19 base pairs and do

not reflect that each sgRNA starts with a G which is required if expressed from a pol-III promoter for initiation of transcription. Thus, for SEQ ID NO:38, the sequence would be GACGCUAAAUUCCGCAGU (SEQ ID NO:102) rather than ACGCUCAAAUUCCGCAGU (SEQ ID NO:38). In embodiments, SEQ ID NOS:38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, and 96 each contain a G as the first nucleotide. In embodiments, SEQ ID NOS:127-183 each contain a G as the first nucleotide.

[0069] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracrRNA sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a complex (e.g., CRISPR complex) at a target sequence, wherein the complex (e.g., CRISPR complex) comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracrRNA sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracrRNA sequence or tracr mate sequence. In aspects, the degree of complementarity between the tracrRNA sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In aspects, the degree of complementarity is about or at least about 80%, 90%, 95%, or 100%. In aspects, the tracrRNA sequence is about or more than about 5, 10, 15, 20, 30, 40, 50, or more nucleotides in length. In aspects, the tracrRNA sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin.

[0070] "Nucleic acid" refers to nucleotides (e.g., deoxyribonucleotides or ribonucleotides) and polymers thereof in either single-, double- or multiple-stranded form, or complements thereof. The terms "polynucleotide," "oligonucleotide," "oligo" or the like refer, in the usual and customary sense, to a linear sequence of nucleotides. The term "nucleotide" refers, in the usual and customary sense, to a single unit of a polynucleotide, *i.e.*, a monomer. Nucleotides can be ribonucleotides, deoxyribonucleotides, or modified versions thereof. Examples of polynucleotides contemplated herein include single and double stranded DNA, single and double stranded RNA, and hybrid molecules having mixtures of single and double stranded DNA and RNA. Examples of nucleic acids, e.g. polynucleotides, contemplated herein include, but are not limited to, any type of RNA, e.g., mRNA, siRNA, miRNA, sgRNA, and guide RNA

and any type of DNA, genomic DNA, plasmid DNA, and minicircle DNA, and any fragments thereof. In aspects, the nucleic acid is messenger RNA. In aspects, the messenger RNA is messenger ribonucleoprotein (RNP). The term “duplex” in the context of polynucleotides refers, in the usual and customary sense, to double strandedness. Nucleic acids can be linear or branched. For example, nucleic acids can be a linear chain of nucleotides or the nucleic acids can be branched, e.g., such that the nucleic acids comprise one or more arms or branches of nucleotides. Optionally, the branched nucleic acids are repetitively branched to form higher ordered structures such as dendrimers and the like.

[0071] As may be used herein, the terms “nucleic acid,” “nucleic acid molecule,” “nucleic acid oligomer,” “oligonucleotide,” “nucleic acid sequence,” “nucleic acid fragment” and “polynucleotide” are used interchangeably and are intended to include, but are not limited to, a polymeric form of nucleotides covalently linked together that may have various lengths, either deoxyribonucleotides or ribonucleotides, or analogs, derivatives or modifications thereof. Different polynucleotides may have different three-dimensional structures, and may perform various functions, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, an exon, an intron, intergenic DNA (including, without limitation, heterochromatic DNA), messenger RNA (mRNA), transfer RNA, ribosomal RNA, a ribozyme, cDNA, a recombinant polynucleotide, a branched polynucleotide, a plasmid, a vector, isolated DNA of a sequence, isolated RNA of a sequence, sgRNA, guide RNA, a nucleic acid probe, and a primer. Polynucleotides useful in the methods of the disclosure may comprise natural nucleic acid sequences and variants thereof, artificial nucleic acid sequences, or a combination of such sequences.

[0072] A polynucleotide is typically composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). Thus, the term “polynucleotide sequence” is the alphabetical representation of a polynucleotide molecule; alternatively, the term may be applied to the polynucleotide molecule itself. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching. Polynucleotides may optionally include one or more non-standard nucleotide(s), nucleotide analog(s) and/or modified nucleotides.

[0073] Nucleic acids, including e.g., nucleic acids with a phosphothioate backbone, can include one or more reactive moieties. As used herein, the term reactive moiety includes any group capable of reacting with another molecule, e.g., a nucleic acid or polypeptide through covalent, non-covalent or other interactions. By way of example, the nucleic acid can include an

amino acid reactive moiety that reacts with an amino acid on a protein or polypeptide through a covalent, non-covalent or other interaction.

[0074] The terms also encompass nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, include, without limitation, phosphodiester derivatives including, *e.g.*, phosphoramidate, phosphorodiamidate, phosphorothioate (also known as phosphorothioate having double bonded sulfur replacing oxygen in the phosphate), phosphorodithioate, phosphonocarboxylic acids, phosphonocarboxylates, phosphonoacetic acid, phosphonoformic acid, methyl phosphonate, boron phosphonate, or O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press) as well as modifications to the nucleotide bases such as in 5-methyl cytidine or pseudouridine; and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, modified sugars, and non-ribose backbones (*e.g.*, phosphorodiamidate morpholino oligos or locked nucleic acids (LNA) as known in the art), including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, *Carbohydrate Modifications in Antisense Research*, Sanghui & Cook, eds. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, *e.g.*, to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. In aspects, the internucleotide linkages in DNA are phosphodiester, phosphodiester derivatives, or a combination of both.

[0075] Nucleic acids can include nonspecific sequences. As used herein, the term "nonspecific sequence" refers to a nucleic acid sequence that contains a series of residues that are not designed to be complementary to or are only partially complementary to any other nucleic acid sequence. By way of example, a nonspecific nucleic acid sequence is a sequence of nucleic acid residues that does not function as an inhibitory nucleic acid when contacted with a cell or organism.

[0076] The term "complementary" or "complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. For example, the sequence A-G-T is complementary to the

sequence T-C-A. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary, respectively). “Perfectly complementary” means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. “Substantially complementary” as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions (i.e., stringent hybridization conditions).

[0077] The phrase “stringent hybridization conditions” refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acids, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, “Overview of principles of hybridization and the strategy of nucleic acid assays” (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C.

[0078] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary “moderately stringent hybridization conditions” include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice

background. One of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency. Additional guidelines for determining hybridization parameters are provided in numerous references, e.g., Current Protocols in Molecular Biology, ed. Ausubel, et al., supra.

[0079] The term "gene" means the segment of DNA involved in producing a protein; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons). The leader, the trailer as well as the introns include regulatory elements that are necessary during the transcription and the translation of a gene. Further, a "protein gene product" is a protein expressed from a particular gene.

[0080] The word "expression" or "expressed" as used herein in reference to a gene means the transcriptional and/or translational product of that gene. The level of expression of a DNA molecule in a cell may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of protein encoded by that DNA produced by the cell. The level of expression of non-coding nucleic acid molecules (e.g., sgRNA) may be detected by standard PCR or Northern blot methods well known in the art. See, Sambrook et al., 1989 Molecular Cloning: A Laboratory Manual, 18.1-18.88.

[0081] The term "transcriptional regulatory sequence" as provided herein refers to a segment of DNA that is capable of increasing or decreasing transcription (e.g., expression) of a specific gene within an organism. Non-limiting examples of transcriptional regulatory sequences include promoters, enhancers, and silencers.

[0082] The terms "transcription start site" and transcription initiation site" may be used interchangeably to refer herein to the 5' end of a gene sequence (e.g., DNA sequence) where RNA polymerase (e.g., DNA-directed RNA polymerase) begins synthesizing the RNA transcript. The transcription start site may be the first nucleotide of a transcribed DNA sequence where RNA polymerase begins synthesizing the RNA transcript. A skilled artisan can determine a transcription start site via routine experimentation and analysis, for example, by performing a run-off transcription assay or by definitions according to FANTOM5 database.

[0083] The term "promoter" as used herein refers to a region of DNA that initiates transcription of a particular gene. Promoters are typically located near the transcription start site of a gene, upstream of the gene and on the same strand (i.e., 5' on the sense strand) on the DNA. Promoters may be about 100 to about 1000 base pairs in length.

[0084] The term "enhancer" as used herein refers to a region of DNA that may be bound by

proteins (e.g., transcription factors) to increase the likelihood that transcription of a gene will occur. Enhancers may be about 50 to about 1500 base pairs in length. Enhancers may be located downstream or upstream of the transcription initiation site that it regulates and may be several hundreds of base pairs, several thousands of base pairs, or several millions of base pairs away from the transcription initiation site.

[0085] The term “silencer” as used herein refers to a DNA sequence capable of binding transcription regulation factors known as repressors, thereby negatively effecting transcription of a gene. Silencer DNA sequences may be found at many different positions throughout the DNA, including, but not limited to, upstream of a target gene for which it acts to repress transcription of the gene (e.g., silence gene expression).

[0086] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid. The terms “non-naturally occurring amino acid” and “unnatural amino acid” refer to amino acid analogs, synthetic amino acids, and amino acid mimetics which are not found in nature.

[0087] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides may also be referred to by their commonly accepted single-letter codes.

[0088] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues, wherein the polymer may, in aspects, be conjugated to a moiety that does not consist of amino acids. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally

occurring amino acid polymers. A "fusion protein" refers to a chimeric protein encoding two or more separate protein sequences that are recombinantly expressed as a single moiety.

[0089] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those nucleic acids that encode identical or essentially identical amino acid sequences. Because of the degeneracy of the genetic code, a number of nucleic acid sequences will encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0090] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure. The following eight groups each contain amino acids that are conservative substitutions for one another: (1) Alanine (A), Glycine (G); (2) Aspartic acid (D), Glutamic acid (E); (3) Asparagine (N), Glutamine (Q); (4) Arginine (R), Lysine (K); (5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); (6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (7) Serine (S), Threonine (T); and (8) Cysteine (C), Methionine (M) (*see, e.g.,* Creighton, *Proteins* (1984)).

[0091] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.,* gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of

positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0092] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (*see, e.g.*, NCBI web site ncbi.nlm.nih.gov/BLAST/ or the like). Such sequences are then said to be "substantially identical." This definition also refers to, or may be applied to, the complement of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

[0093] An amino acid or nucleotide base "position" is denoted by a number that sequentially identifies each amino acid (or nucleotide base) in the reference sequence based on its position relative to the N-terminus (or 5'-end). Due to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence determined by simply counting from the N-terminus will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where a variant has a deletion relative to an aligned reference sequence, there will be no amino acid in the variant that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to a numbered amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0094] The terms "numbered with reference to" or "corresponding to," when used in the context of the numbering of a given amino acid or polynucleotide sequence, refers to the numbering of the residues of a specified reference sequence when the given amino acid or

polynucleotide sequence is compared to the reference sequence.

[0095] A “cell” as used herein, refers to a cell carrying out metabolic or other function sufficient to preserve or replicate its genomic DNA. A cell can be identified by well-known methods in the art including, for example, presence of an intact membrane, staining by a particular dye, ability to produce progeny or, in the case of a gamete, ability to combine with a second gamete to produce a viable offspring. Cells may include prokaryotic and eukaryotic cells. Prokaryotic cells include but are not limited to bacteria. Eukaryotic cells include but are not limited to yeast cells and cells derived from plants and animals, for example mammalian, insect (*e.g.*, spodoptera) and human cells. Cells may be useful when they are naturally nonadherent or have been treated not to adhere to surfaces, for example by trypsinization.

[0096] As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a linear or circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “expression vectors.” In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, “plasmid” and “vector” can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. Additionally, some viral vectors are capable of targeting a particular cells type either specifically or non-specifically. Replication-incompetent viral vectors or replication-defective viral vectors refer to viral vectors that are capable of infecting their target cells and delivering their viral payload, but then fail to continue the typical lytic pathway that leads to cell lysis and death.

[0097] The terms “transfection”, “transduction”, “transfecting” or “transducing” can be used interchangeably and are defined as a process of introducing a nucleic acid molecule (*e.g.*, mRNA, DNA, RNP) and/or a protein to a cell. Nucleic acids may be introduced to a cell using non-viral or viral-based methods. The nucleic acid molecule can be a sequence encoding

complete proteins or functional portions thereof. Typically, a nucleic acid vector, comprising the elements necessary for protein expression (e.g., a promoter, transcription start site, etc.). Non-viral methods of transfection include any appropriate method that does not use viral DNA or viral particles as a delivery system to introduce the nucleic acid molecule into the cell.

Exemplary non-viral transfection methods include nanoparticle encapsulation of the nucleic acids that encode the fusion protein (e.g., lipid nanoparticles, gold nanoparticles, and the like), calcium phosphate transfection, liposomal transfection, nucleofection, sonoporation, transfection through heat shock, magnetofection and electroporation. For viral-based methods, any useful viral vector can be used in the methods described herein. Examples of viral vectors include, but are not limited to retroviral, adenoviral, lentiviral and adeno-associated viral vectors. In aspects, the nucleic acid molecules are introduced into a cell using a retroviral vector following standard procedures well known in the art. The terms "transfection" or "transduction" also refer to introducing proteins into a cell from the external environment. Typically, transduction or transfection of a protein relies on attachment of a peptide or protein capable of crossing the cell membrane to the protein of interest. See, e.g., Ford et al. (2001) *Gene Therapy* 8:1-4 and Prochiantz (2007) *Nat. Methods* 4:119-20.

[0098] A "peptide linker" as provided herein is a linker including a peptide moiety. In embodiments, the peptide linker is a divalent peptide, such as an amino acid sequence attached at the N-terminus and the C-terminus to the remainder of the compound (e.g., fusion protein provided herein. The peptide linker may be a peptide moiety (a divalent peptide moiety) capable of being cleaved (e.g., a P2A cleavable polypeptide). A peptide linker as provided herein may also be referred to interchangeably as an amino acid linker. In aspects, the peptide linker includes 1 to about 80 amino acid residues. In aspects, the peptide linker includes 1 to about 70 amino acid residues. In aspects, the peptide linker includes 1 to about 60 amino acid residues. In aspects, the peptide linker includes 1 to about 50 amino acid residues. In aspects, the peptide linker includes 1 to about 40 amino acid residues. In aspects, the peptide linker includes 1 to about 30 amino acid residues. In aspects, the peptide linker includes 1 to about 25 amino acid residues. In aspects, the peptide linker includes 1 to about 20 amino acid residues. In aspects, the peptide linker includes about 2 to about 20 amino acid residues. In aspects, the peptide linker includes about 2 to about 19 amino acid residues. In aspects, the peptide linker includes about 2 to about 18 amino acid residues. In aspects, the peptide linker includes about 2 to about 17 amino acid residues. In aspects, the peptide linker includes about 2 to about 16 amino acid residues. In aspects, the peptide linker includes about 2 to about 15 amino acid residues. In aspects, the peptide linker includes about 2 to about 14 amino acid residues. In aspects, the

[illegible]

about 16 amino acid residues. In aspects, the peptide linker includes about 17 amino acid residues. In aspects, the peptide linker includes about 18 amino acid residues. In aspects, the peptide linker includes about 19 amino acid residues. In aspects, the peptide linker includes about 20 amino acid residues. In aspects, the peptide linker includes about 21 amino acid residues. In aspects, the peptide linker includes about 22 amino acid residues. In aspects, the peptide linker includes about 23 amino acid residues. In aspects, the peptide linker includes about 24 amino acid residues. In aspects, the peptide linker includes about 25 amino acid residues.

[0099] In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:17. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:17. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:18. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:18. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:19. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:19. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:20. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:20. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:21. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:21. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:22. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:22. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:27. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:27. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:24. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:24. In aspects, the peptide linker includes the sequence set forth by SEQ ID NO:29. In aspects, the peptide linker is the sequence set forth by SEQ ID NO:29. In aspects, the peptide linker is an XTEN polypeptide. In aspects, the peptide linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:17, 18, 19, 20, 21, 22, 24, 27, or 29. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:17, 18, 19, 20, 21, 22, 24, 27, or 29.

[0100] In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:17. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:18. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:19. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID

NO:20. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:21. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:22. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:24. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:27. In aspects, the peptide linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:29. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:17, 18, 19, 20, 21, 22, 24, 27, or 29. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:17. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:18. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:19. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:20. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:21. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:22. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:24. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:27. In aspects, the peptide linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:29.

[0101] The terms “XTEN,” “XTEN linker,” or “XTEN polypeptide” as used herein refer to an recombinant polypeptide (e.g. unstructured recombinant peptide) lacking hydrophobic amino acid residues. The development and use of XTEN can be found in, for example, Schellenberger et al., Nature Biotechnology 27, 1186-1190 (2009). In aspects, the XTEN linker includes the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker includes the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID

NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:32.

[0102] “Epitope tag” refers to a biological moiety, such as a peptide, that is genetically engineered into a recombinant protein and that functions as a universal epitope that is easily detected by commercially available assays or antibodies and that generally does not compromise the native structure or function of the protein.

[0103] A “detectable agent” or “detectable moiety” is a composition detectable by appropriate means such as spectroscopic, photochemical, biochemical, immunochemical, chemical, magnetic resonance imaging, or other physical means. For example, useful detectable agents include ^{18}F , ^{32}P , ^{33}P , ^{45}Ti , ^{47}Sc , ^{52}Fe , ^{59}Fe , ^{62}Cu , ^{64}Cu , ^{67}Cu , ^{67}Ga , ^{68}Ga , ^{77}As , ^{86}Y , ^{90}Y , ^{89}Sr , ^{89}Zr , ^{94}Tc , ^{94}Tc , $^{99\text{m}}\text{Tc}$, ^{99}Mo , ^{105}Pd , ^{105}Rh , ^{111}Ag , ^{111}In , ^{123}I , ^{124}I , ^{125}I , ^{131}I , ^{142}Pr , ^{143}Pr , ^{149}Pm , ^{153}Sm , $^{154-158}\text{Gd}$, ^{161}Tb , ^{166}Dy , ^{166}Ho , ^{169}Er , ^{175}Lu , ^{177}Lu , ^{186}Re , ^{188}Re , ^{189}Re , ^{194}Ir , ^{198}Au , ^{199}Au , ^{211}At , ^{211}Pb , ^{212}Bi , ^{212}Pb , ^{213}Bi , ^{223}Ra , ^{225}Ac , Cr, V, Mn, Fe, Co, Ni, Cu, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, ^{32}P , fluorophore (e.g. fluorescent dyes), electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, paramagnetic molecules, paramagnetic nanoparticles, ultrasmall superparamagnetic iron oxide ("USPIO") nanoparticles, USPIO nanoparticle aggregates, superparamagnetic iron oxide ("SPIO") nanoparticles, SPIO nanoparticle aggregates, monocrystalline iron oxide nanoparticles, monocrystalline iron oxide, nanoparticle contrast agents, liposomes or other delivery vehicles containing Gadolinium chelate ("Gd-chelate") molecules, Gadolinium, radioisotopes, radionuclides (e.g., carbon-11, nitrogen-13, oxygen-15, fluorine-18, rubidium-82), fluorodeoxyglucose (e.g., fluorine-18 labeled), any gamma ray emitting radionuclides, positron-emitting radionuclide, radiolabeled glucose, radiolabeled water, radiolabeled ammonia, biocolloids, microbubbles (e.g., including microbubble shells including albumin, galactose, lipid, and/or polymers; microbubble gas core including air, heavy gases, perfluorocarbon, nitrogen, octafluoropropane, perflubron lipid microsphere, perflutren), iodinated contrast agents (e.g., iohexol, iodixanol, ioversol, iopamidol, ioxilan, iopromide, diatrizoate, metrizoate, ioxaglate), barium sulfate, thorium dioxide, gold, gold nanoparticles, gold nanoparticle aggregates, fluorophores, two-photon fluorophores, or haptens and proteins or other entities which can be made detectable, e.g., by incorporating a radiolabel into a peptide or antibody specifically reactive with a target peptide.

[0104] A detectable moiety is a monovalent detectable agent or a detectable agent capable of forming a bond with another composition. In aspects, the detectable agent is an epitope tag. In aspects, the epitope tag is an HA tag. In aspects, the HA tag includes the sequence set forth by

SEQ ID NO:24. In aspects, the HA tag is the sequence set forth by SEQ ID NO:24. In aspects, the HA tag has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:24. In aspects, the HA tag has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:24. In aspects, the HA tag has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:24. In aspects, the HA tag has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:24.

[0105] In aspects, the detectable agent is a fluorescent protein. In aspects, the fluorescent protein is blue fluorescent protein (BFP). In aspects, the BFP includes the sequence set forth by SEQ ID NO:30. In aspects, the BFP is the sequence set forth by SEQ ID NO:30. In aspects, the BFP has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:30. In aspects, the BFP has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:30. In aspects, the BFP has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:30. In aspects, the BFP has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:30.

[0106] Radioactive substances (e.g., radioisotopes) that may be used as imaging and/or labeling agents in accordance with the aspects of the disclosure include, but are not limited to, ^{18}F , ^{32}P , ^{33}P , ^{45}Ti , ^{47}Sc , ^{52}Fe , ^{59}Fe , ^{62}Cu , ^{64}Cu , ^{67}Cu , ^{67}Ga , ^{68}Ga , ^{77}As , ^{86}Y , ^{90}Y , ^{89}Sr , ^{89}Zr , ^{94}Tc , ^{94}Tc , $^{99\text{m}}\text{Tc}$, ^{99}Mo , ^{105}Pd , ^{105}Rh , ^{111}Ag , ^{111}In , ^{123}I , ^{124}I , ^{125}I , ^{131}I , ^{142}Pr , ^{143}Pr , ^{149}Pm , ^{153}Sm , ^{154}Sm , ^{158}Gd , ^{161}Tb , ^{166}Dy , ^{166}Ho , ^{169}Er , ^{175}Lu , ^{177}Lu , ^{186}Re , ^{188}Re , ^{189}Re , ^{194}Ir , ^{198}Au , ^{199}Au , ^{211}At , ^{211}Pb , ^{212}Bi , ^{212}Pb , ^{213}Bi , ^{223}Ra and ^{225}Ac . Paramagnetic ions that may be used as additional imaging agents in accordance with the aspects of the disclosure include, but are not limited to, ions of transition and lanthanide metals (e.g., metals having atomic numbers of 21-29, 42, 43, 44, or 57-71). These metals include ions of Cr, V, Mn, Fe, Co, Ni, Cu, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb and Lu.

[0107] "Contacting" is used in accordance with its plain ordinary meaning and refers to the process of allowing at least two distinct species to become sufficiently proximal to react, interact or physically touch. It should be appreciated, however, the resulting reaction product can be produced directly from a reaction between the added reagents or from an intermediate from one or more of the added reagents which can be produced in the reaction mixture.

[0108] The term "contacting" may include allowing two species to react, interact, or physically touch, wherein the two species may be, for example, a fusion protein as provided herein and a nucleic acid sequence (e.g., target DNA sequence).

[0109] As defined herein, the term "inhibition", "inhibit", "inhibiting", "repression,"

repressing,” “silencing,” “silence” and the like when used in reference to a composition as provided herein (e.g., fusion protein, complex, nucleic acid, vector) refer to negatively affecting (e.g., decreasing) the activity (e.g., transcription) of a nucleic acid sequence (e.g., decreasing transcription of a gene) relative to the activity of the nucleic acid sequence (e.g., transcription of a gene) in the absence of the composition (e.g., fusion protein, complex, nucleic acid, vector). In aspects, inhibition refers to reduction of a disease or symptoms of disease (e.g., cancer). Thus, inhibition includes, at least in part, partially or totally blocking activation (e.g., transcription), or decreasing, preventing, or delaying activation (e.g., transcription) of the nucleic acid sequence. The inhibited activity (e.g., transcription) may be 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, or less than that in a control. In aspects, the inhibition is 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, or more in comparison to a control.

[0110] A “control” sample or value refers to a sample that serves as a reference, usually a known reference, for comparison to a test sample. For example, a test sample can be taken from a test condition, e.g., in the presence of a test compound, and compared to samples from known conditions, e.g., in the absence of the test compound (negative control), or in the presence of a known compound (positive control). A control can also represent an average value gathered from a number of tests or results. One of skill in the art will recognize that controls can be designed for assessment of any number of parameters. For example, a control can be devised to compare therapeutic benefit based on pharmacological data (e.g., half-life) or therapeutic measures (e.g., comparison of side effects). One of skill in the art will understand which controls are valuable in a given situation and be able to analyze data based on comparisons to control values. Controls are also valuable for determining the significance of data. For example, if values for a given parameter are widely variant in controls, variation in test samples will not be considered as significant.

[0111] Fusion Proteins

[0112] Provided herein are, inter alia, fusion proteins that can turn off genes permanently (e.g., irreversibly) and reversibly in mammalian cells using CRISPR-based epigenome editing. In embodiments, the fusion protein includes a single polypeptide fusion of proteins (e.g., catalytically inactive Cas9 (e.g., dCas9), a KRAB domain, Dnmt3A and Dnmt3L) which can be transiently delivered as mRNA, DNA or RNP and expressed transiently in cells. The fusion protein can be directed to a specific site in a mammalian genome using a sgRNA or cr:tracrRNA. Once properly positioned and without intending to be bound by a theory, the fusion protein adds DNA methylation and/or repressive chromatin marks to the target nucleic acid, resulting in gene silencing that is inheritable across subsequent cell divisions. In this way,

the fusion protein can perform epigenome editing that bypasses the need to generate DNA double-strand breaks in the host genome, making it a safe and reversible way of manipulating the genome of a living organism.

[0113] In embodiments, the fusion protein comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme; a KRAB domain, and a DNA methyltransferase domain. In aspects, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a nuclease-deficient RNA-guided DNA endonuclease enzyme, and KRAB domain. In aspects, the fusion protein comprises, from N-terminus to C-terminus, a KRAB domain a nuclease-deficient RNA-guided DNA endonuclease enzyme, and a DNA methyltransferase domain. In embodiments, the nuclease-deficient RNA-guided endonuclease enzyme is a CRISPR-associated protein. In embodiments, the fusion protein further comprises one or more peptide linkers. In aspects, the fusion protein further comprises one or more detectable tags. In aspects, the fusion protein further comprises one or more nuclear localization sequences. In aspects, the fusion protein further comprises one or more peptide linkers, one or more detectable tags, one or more nuclear localization sequences, or a combination of two or more of the foregoing. When the fusion protein comprises one or more peptide linkers, each peptide linker can be the same or different. When the fusion protein comprises one or more detectable tags, each detectable tag can be the same or different. In aspects, the fusion protein comprises from 1 to 10 detectable tags. In aspects, the fusion protein comprises from 1 to 9 detectable tags. In aspects, the fusion protein comprises from 1 to 8 detectable tags. In aspects, the fusion protein comprises from 1 to 7 detectable tags. In aspects, the fusion protein comprises from 1 to 6 detectable tags. In aspects, the fusion protein comprises from 1 to 5 detectable tags. In aspects, the fusion protein comprises from 1 to 4 detectable tags. In aspects, the fusion protein comprises from 1 to 3 detectable tags. In aspects, the fusion protein comprises from 1 to 2 detectable tags. In aspects, the fusion protein comprises 1 detectable tag. In aspects, the fusion protein comprises 2 detectable tags. In aspects, the fusion protein comprises 3 detectable tags. In aspects, the fusion protein comprises 4 detectable tags. In aspects, the fusion protein comprises 5 detectable tags.

[0114] In embodiments, the fusion protein comprises a nuclease-deficient DNA endonuclease enzyme; a KRAB domain, and a DNA methyltransferase domain. In aspects, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a nuclease-deficient DNA endonuclease enzyme, and KRAB domain. In aspects, the fusion protein comprises, from N-terminus to C-terminus, a KRAB domain a nuclease-deficient DNA endonuclease enzyme, and a DNA methyltransferase domain. In embodiments, the fusion protein further comprises one or more peptide linkers. In aspects, the fusion protein further

comprises one or more detectable tags. In aspects, the fusion protein further comprises one or more nuclear localization sequences. In aspects, the fusion protein further comprises one or more peptide linkers, one or more detectable tags, one or more nuclear localization sequences, or a combination of two or more of the foregoing. When the fusion protein comprises one or more peptide linkers, each peptide linker can be the same or different. In embodiments, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or TALE. In embodiments, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain. In embodiments, the nuclease-deficient DNA endonuclease enzyme is a TALE.

[0115] In embodiments, the disclosure provides a fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker. In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 5 to about 864 amino acid residues, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker comprising from about 5 to about 864 amino acid residues, and a Krüppel-associated box domain. In aspects, the first and second XTEN linkers comprise from about 20 to about 100 amino acid residues. In embodiments, the nuclease-deficient RNA-guided endonuclease enzyme is a CRISPR-associated protein. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, dCpf1, ddCpf1, a nuclease-deficient Cas9 variant, a nuclease-deficient Class II CRISPR endonuclease, a zinc finger domain, a leucine zipper domain, a winged helix domain, a TALE, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is Cas-phi. In aspects, the DNA methyltransferase domain comprises a Dnmt3A. In aspects, the DNA methyltransferase domain (Dnmt3A) further comprises a Dnmt3L regulatory factor (referred to herein as a Dnmt3A-3L domain or a Dnmt3B-3L domain). In aspects, the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0116] In embodiments, the disclosure provides a fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker.

In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 5 to about 864 amino acid residues, a nuclease-deficient endonuclease enzyme, a second XTEN linker comprising from about 5 to about 864 amino acid residues, and a Krüppel-associated box domain. In aspects, the first and second XTEN linkers comprise from about 20 to about 100 amino acid residues. In aspects, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or a TALE. In aspects, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain. In aspects, the nuclease-deficient DNA endonuclease enzyme is a TALE. In aspects, the DNA methyltransferase domain comprises a Dnmt3A. In aspects, the DNA methyltransferase domain (Dnmt3A) further comprises a Dnmt3L regulatory factor (referred to herein as a Dnmt3A-3L domain or a Dnmt3B-3L domain). In aspects, the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0117] In embodiments, the disclosure provides a fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker. In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from greater than 50 to about 864 amino acid residues, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker comprising from about 5 to 50 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues. In aspects, the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, dCpf1, ddCpf1, a nuclease-deficient Cas9 variant, a nuclease-deficient Class II CRISPR endonuclease, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is Cas-phi. In aspects, the DNA methyltransferase domain comprises a Dnmt3A. In aspects, the DNA methyltransferase domain (Dnmt3A) further comprises a Dnmt3L

regulatory factor (referred to herein as a Dnmt3A-3L domain or a Dnmt3B-3L domain). In aspects, the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0118] In embodiments, the disclosure provides a fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker. In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from greater than 50 to about 864 amino acid residues, a nuclease-deficient endonuclease enzyme, a second XTEN linker comprising from about 5 to 50 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues. In aspects, the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues. In aspects, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or a TALE. In aspects, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain. In aspects, the nuclease-deficient DNA endonuclease enzyme is TALE. In aspects, the DNA methyltransferase domain comprises a Dnmt3A. In aspects, the DNA methyltransferase domain (Dnmt3A) further comprises a Dnmt3L regulatory factor (referred to herein as a Dnmt3A-3L domain or a Dnmt3B-3L domain). In aspects, the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0119] In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient RNA-guided endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, a second XTEN linker, a Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker. In aspects, the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues. In aspects, the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues. In aspects, the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues. In aspects, the nuclease-

deficient RNA-guided DNA endonuclease enzyme is dCas9, dCpf1, ddCpf1, a nuclease-deficient Cas9 variant, a nuclease-deficient Class II CRISPR endonuclease, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9. In aspects, the nuclease-deficient RNA-guided DNA endonuclease enzyme is Cas-phi. In aspects, the DNA methyltransferase domain comprises a Dnmt3A domain. In aspects, the Dnmt3A domain is linked to a Dnmt3L regulatory factor (referred to herein as a Dnmt3A-3L domain). In aspects, the DNA methyltransferase domain comprises a Dnmt3B domain. In aspects, the Dnmt3B domain is linked to a Dnmt3L regulatory factor (referred to herein as a Dnmt3B-3L domain).

[0120] In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, a second XTEN linker, a Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag. In aspects, the first XTEN linker comprises one or more amino acid residues than the second XTEN linker. In aspects, the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues. In aspects, the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues. In aspects, the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues. In aspects, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or a TALE. In aspects, the nuclease-deficient DNA endonuclease enzyme is zinc finger domain. In aspects, the nuclease-deficient DNA endonuclease enzyme is TALE. In aspects, the DNA methyltransferase domain comprises a Dnmt3A domain. In aspects, the Dnmt3A domain is linked to a Dnmt3L regulatory factor (referred to herein as a Dnmt3A-3L domain). In aspects, the DNA methyltransferase domain comprises a Dnmt3B domain. In aspects, the Dnmt3B domain is linked to a Dnmt3L regulatory factor (referred to herein as a Dnmt3B-3L domain).

[0121] In embodiments, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid

sequence having at least 75% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:110.

[0122] In embodiments, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:110.

[0123] In embodiments, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:110.

[0124] In embodiments, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:99. In aspects, the fusion protein

includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 88% sequence identity to SEQ ID NO:110.

[0125] In embodiments, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:110.

[0126] In embodiments, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 92% sequence identity to SEQ ID NO:110.

[0127] In embodiments, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:107. In

aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 94% sequence identity to SEQ ID NO:110.

[0128] In embodiments, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:110.

[0129] In embodiments, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 96% sequence identity to SEQ ID NO:110.

[0130] In embodiments, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:97. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:98. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:99. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:107. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence

identity to SEQ ID NO:108. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:109. In aspects, the fusion protein includes an amino acid sequence having at least 98% sequence identity to SEQ ID NO:110.

[0131] In embodiments, the fusion protein comprises the structure: A-B-C, or B-A-C or C-A-B, or C-B-A, or B-C-A, or A-C-B; where A comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain; and wherein the component on the left is the N-terminus and the component on the right is the C-terminus. In aspects, the fusion protein further comprises one or more peptide linkers and one or more detectable tags. In aspects, A-B, B-A, B-C, C-B, A-C, and C-A are each independently linked together via a covalent bond, a peptide linker, a detectable tag, a nuclear localization sequence, or a combination of two or more thereof. The peptide linker can be any known in the art (e.g., P2A cleavable peptide, XTEN linker, and the like). In aspects, the fusion protein comprises other components, such as detectable tags (e.g., HA tag, blue fluorescent protein, and the like). In embodiments, “A” is a nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, “A” is a CRISPR-associated protein. In embodiments, “A” is a nuclease-deficient DNA endonuclease enzyme. In embodiments, “A” is a zinc finger domain. In embodiments, “A” is TALE.

[0132] In embodiments, the fusion protein comprises the structure: A-L₁-B-L₂-C or C-L₂-B-L₁-A or C-L₂-A-L₁-B, where A comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, L₁ is absent, a covalent bond, or a peptide linker, and L₂ is absent, a covalent bond, or a peptide linker; and where the component at the left is at the N-terminus and the component on the right is at the C-terminus. In aspects, A is covalently linked to B via a peptide linker. In aspects, A is covalently linked to B via a covalent bond. In aspects, B is covalently linked to C via a peptide linker. In aspects, B is covalently linked to C via a covalent bond. The peptide linker can be any known in the art (e.g., P2A cleavable peptide, XTEN linker, and the like). In aspects, the fusion protein comprises other components, such as detectable tags, nuclear localization sequences, and the like. In aspects, L₁ is a covalent bond, a peptide linker, a detectable tag, a nuclear localization sequence, or a combination thereof. In aspects, L₂ is a covalent bond, a peptide linker, a detectable tag, a nuclear localization sequence, or a combination thereof. In embodiments, “A” is a nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, “A” is a CRISPR-associated protein. In embodiments, “A” is a nuclease-deficient DNA endonuclease enzyme. In embodiments, “A” is a zinc finger

domain. In embodiments, “A” is TALE.

[0133] In embodiments, the fusion protein has at least 80% sequence identity to the fusion protein having the amino acid sequence of Formula (A); where the amino acid sequence of Formula (A) is, from N-terminus to C-terminus:



wherein C_1 comprises SEQ ID NO:26 or SEQ ID NO:106; R^3 is absent or R^3 comprises SEQ ID NO:27; C_2 comprises SEQ ID NO:28; R^2 is absent or R^2 comprises SEQ ID NO:32; A comprises SEQ ID NO:23; R^1 is absent or R^1 comprises SEQ ID NO:25; R^4 is absent or R^4 comprises SEQ ID NO:31; and B comprises SEQ ID NO:16, SEQ ID NO:103, SEQ ID NO:104, or SEQ ID NO:105. In embodiments, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 is absent. In embodiments, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 is absent. In embodiments, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 comprises SEQ ID NO:31. In embodiments, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 comprises SEQ ID NO:31. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:16, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 is absent. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:16, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 is absent. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:16, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 comprises SEQ ID NO:31. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:16, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 comprises SEQ ID NO:31. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:103, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 is absent. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:103, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 is absent. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:103, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is absent, and R^4 comprises SEQ ID NO:31. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:103, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 comprises SEQ ID NO:31. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:104, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 comprises SEQ ID NO:25, and R^4 is absent. In embodiments, C_1 comprises SEQ ID NO:26, B comprises SEQ ID NO:104, R^2 comprises SEQ ID NO:32, R^3 comprises SEQ ID NO:27, R^1 is

[illegible]

ID NO:104, R² comprises SEQ ID NO:32, R³ comprises SEQ ID NO:27, R¹ comprises SEQ ID NO:25, and R⁴ comprises SEQ ID NO:31. In embodiments, C₁ comprises SEQ ID NO:106, B comprises SEQ ID NO:105, R² comprises SEQ ID NO:32, R³ comprises SEQ ID NO:27, R¹ comprises SEQ ID NO:25, and R⁴ is absent. In embodiments, C₁ comprises SEQ ID NO:106, B comprises SEQ ID NO:105, R² comprises SEQ ID NO:32, R³ comprises SEQ ID NO:27, R¹ is absent, and R⁴ is absent. In embodiments, C₁ comprises SEQ ID NO:106, B comprises SEQ ID NO:105, R² comprises SEQ ID NO:32, R³ comprises SEQ ID NO:27, R¹ is absent, and R⁴ comprises SEQ ID NO:31. In embodiments, C₁ comprises SEQ ID NO:106, B comprises SEQ ID NO:105, R² comprises SEQ ID NO:32, R³ comprises SEQ ID NO:27, R¹ comprises SEQ ID NO:25, and R⁴ comprises SEQ ID NO:31. In embodiments, R² comprises SEQ ID NO:32. In embodiments, R² is absent. In embodiments, R¹ and R⁴ are absent. In embodiments, R¹ is absent and R⁴ comprises SEQ ID NO:31. In embodiments, R¹ comprises SEQ ID NO:25 and R⁴ is absent. In embodiments, R¹ comprises SEQ ID NO:25 and R⁴ comprises SEQ ID NO:31. In embodiments, R³ is absent. In embodiments, R³ comprises SEQ ID NO:27. In embodiments, C₁ comprises SEQ ID NO:26. In embodiments, C₁ comprises SEQ ID NO:105. In embodiments, B comprises SEQ ID NO:16. In embodiments, B comprises SEQ ID NO:103. In embodiments, B comprises SEQ ID NO:104. In embodiments, the fusion protein has at least 85% sequence identity to the fusion protein having the structure of Formula (A). In embodiments, the fusion protein has at least 88% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 90% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 91% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 92% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 93% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 94% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 95% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 96% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 97% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 98% sequence identity to the amino acid sequence of Formula (A). In embodiments, the fusion protein has at least 99% sequence identity to the amino acid sequence of Formula (A).

[0134] In embodiments, the fusion protein comprises the structure: B-L₁-A-L₂-C or C-L₁-A-L₂-B where A comprise a nuclease-deficient RNA-guided DNA endonuclease enzyme or a

nuclease-deficient DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, L₁ is a covalent bond or a peptide linker, and L₂ is a covalent bond or a peptide linker. In embodiments the fusion protein comprises the structure: B-L₁-A-L₂-C. In embodiments the fusion protein comprises the structure: C-L₁-A-L₂-B. In aspects, L₁ is a peptide linker. In aspects, L₁ is a covalent bond. In aspects, L₂ is a peptide linker. In aspects, L₂ is a covalent bond. The peptide linker can be any known in the art or described herein (e.g., P2A cleavable peptide, XTEN linker, and the like). In aspects, the fusion protein comprises other components, such as detectable tags. In aspects, L₁ is a covalent bond, a peptide linker, a detectable tag, or a combination thereof. In aspects, L₂ is a covalent bond, a peptide linker, a detectable tag, or a combination thereof. In aspects, the fusion protein further comprises a nuclear localization sequence. In embodiments, “A” is a nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, “A” is a CRISPR-associated protein. In embodiments, “A” is a nuclease-deficient DNA endonuclease enzyme. In embodiments, “A” is a zinc finger domain. In embodiments, “A” is TALE.

[0135] In embodiments, the fusion protein comprises the structure: B-L₃-A-L₄-C-L₅-D or C-L₃-A-L₄-B-L₅-D, where A comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, D is absent or D comprises one or more detectable tags, L₃ comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L₄ comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L₅ is absent or L₅ comprises a covalent bond or a peptide linker. In embodiments, the fusion protein comprises the structure: B-L₃-A-L₄-C-L₅-D. In embodiments, the fusion protein comprises the structure: C-L₃-A-L₄-B-L₅-D. In aspects, L₃ is a peptide linker. In aspects, L₃ is a covalent bond. In aspects, L₃ comprises a peptide linker and a detectable tag. In aspects, L₃ comprises a detectable tag. In aspects, L₄ is a peptide linker. In aspects, L₄ comprises a peptide linker and a detectable tag. In aspects, L₄ is a covalent bond. In aspects, L₄ comprises a detectable tag. In aspects, L₅ is a peptide linker. In aspects, L₅ is a covalent bond. In aspects, D comprises one or a plurality of detectable tags. In aspects, D comprises one detectable tag. In aspects, D comprises two detectable tags. In aspects, D comprises three detectable tags. In aspects, D comprises a plurality of detectable tags. D can be any detectable tag known in the art and/or described herein (e.g., HA tag, blue fluorescent protein, and the like). In aspects L₅ and D are absent. When L₃, L₄, L₅, and D comprise two or more detectable tags, each detectable tag is the same or different. The peptide linker can be any known in the art and/or described herein (e.g., P2A cleavable peptide, XTEN linker, and the like). In aspects, the fusion protein

further comprises a nuclear localization sequence. In embodiments, “A” is a nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, “A” is a CRISPR-associated protein. In embodiments, “A” is a nuclease-deficient DNA endonuclease enzyme. In embodiments, “A” is a zinc finger domain. In embodiments, “A” is TALE.

[0136] In embodiments, the fusion protein comprises the structure: C-L₃-A-L₄-B-L₅-D, where A comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, D is absent or D comprises one or more detectable tags, L₃ comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L₄ comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L₅ is absent or L₅ comprises a covalent bond or a peptide linker; and where C is at the N-terminus and D is at the C-terminus. In aspects, L₃ is a peptide linker. In aspects, L₃ is a covalent bond. In aspects, L₃ comprises a detectable tag. In aspects, L₃ comprises a peptide linker and a detectable tag. In aspects, L₄ is a peptide linker. In aspects, L₄ is a covalent bond. In aspects, L₄ comprises a detectable tag. In aspects, L₄ comprises a peptide linker and a detectable tag. In aspects, L₅ is a peptide linker. In aspects, L₅ is a covalent bond. In aspects, D comprises one or a plurality of detectable tags. In aspects, D comprises one detectable tag. In aspects, D comprises two detectable tags. In aspects, D comprises three detectable tags. In aspects, D comprises a plurality of detectable tags. D can be any detectable tag known in the art and/or described herein (e.g., HA tag, blue fluorescent protein, and the like). In aspects L₅ and D are absent. When L₃, L₄, L₅, and D comprise two or more detectable tags, each detectable tag is the same or different. The peptide linker can be any known in the art and/or described herein (e.g., P2A cleavable peptide, XTEN linker, and the like). In aspects, the fusion protein further comprises a nuclear localization sequence. In embodiments, “A” is a nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, “A” is a CRISPR-associated protein. In embodiments, “A” is a nuclease-deficient DNA endonuclease enzyme. In embodiments, “A” is a zinc finger domain. In embodiments, “A” is TALE.

[0137] In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 60 to about 150 amino acid residues, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker comprising from about 5 to about 50 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from about 70 to about 90 amino acid residues, and the second XTEN linker comprises from about 10 to about 20 amino acid residues. In aspects, the first XTEN linker comprises about 80 amino acid residues, and the second XTEN linker

comprises from about 16 amino acid residues. In aspects, the fusion protein further comprises a detectable tag (e.g., an epitope tag, a fluorescent protein tag), a 2A peptide (e.g., a P2A peptide), a nuclear localization signal peptide, or a combination of two or more thereof. In aspects, the fusion protein comprises from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 60 to about 150 amino acid residues, a nuclease-deficient RNA-guided endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, a second XTEN linker comprising from about 5 to about 50 amino acid residues, a Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag. In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein.

[0138] In embodiments, the fusion protein comprises, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 60 to about 150 amino acid residues, a nuclease-deficient endonuclease enzyme, a second XTEN linker comprising from about 5 to about 50 amino acid residues, and a Krüppel-associated box domain. In aspects, the first XTEN linker comprises from about 70 to about 90 amino acid residues, and the second XTEN linker comprises from about 10 to about 20 amino acid residues. In aspects, the first XTEN linker comprises about 80 amino acid residues, and the second XTEN linker comprises from about 16 amino acid residues. In aspects, the fusion protein further comprises a detectable tag (e.g., an epitope tag, a fluorescent protein tag), a 2A peptide (e.g., a P2A peptide), a nuclear localization signal peptide, or a combination of two or more thereof. In aspects, the fusion protein comprises from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker comprising from about 60 to about 150 amino acid residues, a nuclease-deficient endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, a second XTEN linker comprising from about 5 to about 50 amino acid residues, a Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag. In embodiments, the nuclease-deficient endonuclease enzyme is a zinc finger domain or a TALE.

[0139] In embodiments, the peptide linker is a XTEN linker. In aspects, the XTEN linker includes about 16 to about 80 amino acid residues. In aspects, the XTEN linker includes about 17 to about 80 amino acid residues. In aspects, the XTEN linker includes about 18 to about 80 amino acid residues. In aspects, the XTEN linker includes about 19 to about 80 amino acid residues. In aspects, the XTEN linker includes about 20 to about 80 amino acid residues. In aspects, the XTEN linker includes about 30 to about 80 amino acid residues. In aspects, the XTEN linker includes about 40 to about 80 amino acid residues. In aspects, the XTEN linker includes about 50 to about 80 amino acid residues. In aspects, the XTEN linker includes about

60 to about 80 amino acid residues. In aspects, the XTEN linker includes about 70 to about 80 amino acid residues. In aspects, the XTEN linker includes about 16 to about 70 amino acid residues. In aspects, the XTEN linker includes about 16 to about 60 amino acid residues. In aspects, the XTEN linker includes about 16 to about 50 amino acid residues. In aspects, the XTEN linker includes about 16 to about 40 amino acid residues. In aspects, the XTEN linker includes about 16 to about 35 amino acid residues. In aspects, the XTEN linker includes about 16 to about 30 amino acid residues. In aspects, the XTEN linker includes about 16 to about 25 amino acid residues. In aspects, the XTEN linker includes about 16 to about 20 amino acid residues. In aspects, the XTEN linker includes about 16 amino acid residues. In aspects, the XTEN linker includes about 17 amino acid residues. In aspects, the XTEN linker includes about 18 amino acid residues. In aspects, the XTEN linker includes about 19 amino acid residues. In aspects, the XTEN linker includes about 20 amino acid residues.

[0140] In aspects, the fusion protein comprises at least two XTEN linkers that are the same or different. In aspects, the fusion protein comprises a first XTEN linker having more amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 10 to 150 amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 20 to 120 amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 30 to 110 amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 40 to 110 amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 50 to 100 amino acid residues than a second XTEN linker. In aspects, the fusion protein comprises a first XTEN linker having 60 to 100 amino acid residues than a second XTEN linker.

[0141] In embodiments, the XTEN linker comprises from about 50 to about 864 amino acid residues. In aspects, the XTEN linker comprises from about 50 to about 200 amino acid residues. In aspects, the XTEN linker comprises from about 55 to about 180 amino acid residues. In aspects, the XTEN linker comprises from about 60 to about 150 amino acid residues. In aspects, the XTEN linker comprises from about 60 to about 120 amino acid residues. In aspects, the XTEN linker comprises from about 60 to about 110 amino acid residues. In aspects, the XTEN linker comprises from about 60 to about 100 amino acid residues. In aspects, the XTEN linker comprises from about 70 to about 90 amino acid residues. In aspects, the XTEN linker comprises from about 75 to about 85 amino acid residues. In aspects, the XTEN linker comprises about 80 amino acid residues. In aspects, when a fusion protein comprises at least two XTEN peptide linkers, then the XTEN linker that comprise from

about 50 to about 200 amino acid residues is referred to as a first XTEN peptide linker.

[0142] In embodiments, the XTEN linker comprises from about 5 to about 55 amino acid residues. In aspects, the XTEN linker comprises from about 5 to about 50 amino acid residues. In aspects, the XTEN linker comprises from about 5 to about 40 amino acid residues. In aspects, the XTEN linker comprises from about 10 to about 30 amino acid residues. In aspects, the XTEN linker comprises from about 10 to about 25 amino acid residues. In aspects, the XTEN linker comprises from about 10 to about 20 amino acid residues. In aspects, the XTEN linker comprises from about 14 to about 18 amino acid residues. In aspects, the XTEN linker comprises about 16 amino acid residues. In aspects, when a fusion protein comprises at least two XTEN peptide linkers, then the XTEN linker that comprise from about 5 to about 55 amino acid residues is referred to as a second XTEN peptide linker.

[0143] In embodiments, the XTEN linker includes the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker includes the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:32.

[0144] The fusion protein may include amino acid sequences useful for targeting the fusion protein to specific regions of a cell (e.g., cytoplasm, nucleus). Thus, in aspects, the fusion protein further includes a nuclear localization signal (NLS) peptide. In aspects, the NLS includes the sequence set forth by SEQ ID NO:25. In aspects, the NLS is the sequence set forth by SEQ ID NO:25. In aspects, the NLS has an amino acid sequence that has at least 50%, 55%, 60%,

65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to SEQ ID NO: 25. In aspects, the NLS has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:25. In aspects, the NLS has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:25. In aspects, the NLS has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:25. In aspects, the NLS has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:25. In aspects, the NLS has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:25.

[0145] In embodiments, the fusion protein includes, from N-terminus to C-terminus, a KRAB domain, a nuclease-deficient RNA-guided DNA endonuclease enzyme, and a DNA methyltransferase domain. In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is CRISPR-associated protein and the DNA methyltransferase domain is a Dnmt3A-3L domain. In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain. In embodiments, the dCas9 is covalently linked to the KRAB domain via a peptide linker and wherein the dCas9 is covalently linked to the Dnmt3A-3L domain via a peptide linker.

[0146] In embodiments, the fusion protein includes, from N-terminus to C-terminus, a KRAB domain, a nuclease-deficient DNA endonuclease enzyme, and a DNA methyltransferase domain. In embodiments, the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain and the DNA methyltransferase domain is a Dnmt3A-3L domain. In embodiments, the nuclease-deficient DNA endonuclease enzyme is a TALE and the DNA methyltransferase domain is a Dnmt3A-3L domain. In embodiments, the nuclease-deficient DNA endonuclease enzyme is covalently linked to the KRAB domain via a peptide linker and wherein the nuclease-deficient DNA endonuclease enzyme is covalently linked to the Dnmt3A-3L domain via a peptide linker.

[0147] In embodiments, peptide linker is an XTEN linker. In aspects, the XTEN linker includes the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 75% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:31. In aspects, the XTEN linker includes the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker is the sequence set forth by SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 75% sequence identity to

SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 80% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 85% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 90% sequence identity to SEQ ID NO:32. In aspects, the XTEN linker has an amino acid sequence that has at least 95% sequence identity to SEQ ID NO:32.

[0148] In embodiments, the fusion protein includes the amino acid sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:1. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:1. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:2. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:2. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:3. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:3. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:4. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:4. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:5. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:5. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:6. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:6. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:7. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:7. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:8. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:8. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:9. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:9. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:10. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:10. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:11. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:11. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:12. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:12. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:13. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:13. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:14. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:14. In aspects, the fusion protein includes the amino acid sequence of SEQ ID NO:15. In aspects, the fusion protein is the amino acid sequence of SEQ ID NO:15.

[0149] In embodiments, the fusion protein includes an amino acid sequence having at least

75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80% 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:13. In aspects, the

fusion protein includes an amino acid sequence having at least 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:15.

[0150] Complexes

[0151] In order for the fusion protein to carry out epigenome editing, the fusion protein interacts with (e.g. is non-covalently bound to) a polynucleotide (e.g., sgRNA) that is complementary to a target polynucleotide sequence (e.g., a target DNA sequence to be edited) and further includes a sequence (i.e., a binding sequence) to which the nuclease-deficient RNA-guided DNA endonuclease enzyme of the fusion protein as described herein can bind. In aspects, the polynucleotide that is complementary to a target polynucleotide sequence (e.g., a target DNA sequence to be edited) and further includes a binding sequence to which the nuclease-deficient RNA-guided DNA endonuclease enzyme of the fusion protein as described herein can bind is sgRNA. In aspects, the polynucleotide that is complementary to a target polynucleotide sequence (e.g., a target DNA sequence to be edited) and further includes a binding sequence to which the nuclease-deficient RNA-guided DNA endonuclease enzyme of the fusion protein as described herein can bind is cr:tracrRNA. By forming this complex, the fusion protein is appropriately positioned to perform epigenome editing. The term “complex” refers to a composition that includes two or more components, where the components bind together to make a functional unit. In aspects, a complex described herein includes a fusion protein described herein and a polynucleotide described herein. Thus, in an aspect is provided a fusion protein as described herein, including embodiments and aspects thereof, and sgRNA or cr:tracrRNA (i.e., a polynucleotide including: (1) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (2) a binding sequence for the nuclease-deficient RNA-guided DNA endonuclease enzyme, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is bound to the polynucleotide via the binding sequence (e.g., an amino acid sequence capable of binding to the DNA-targeting sequence)).

[0152] A DNA-targeting sequence refers to a polynucleotide that includes a nucleotide sequence complementary to the target polynucleotide sequence (DNA or RNA). In aspects, a DNA-targeting sequence can be a single RNA molecule (single RNA polynucleotide), which may include a “single-guide RNA,” or “sgRNA.” In aspects, a DNA-targeting sequence can comprise two RNA molecules (two RNA polynucleotides), referred to as a guide RNA (gRNA). In aspects, the DNA-targeting sequence includes two RNA molecules (e.g., joined together via

hybridization at the binding sequence (e.g., dCas9-binding sequence). In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% complementary to the target polynucleotide sequence. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) binds a cellular gene sequence. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 75% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 80% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) binds a cellular gene sequence. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 85% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) binds a cellular gene sequence. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 90% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) binds a cellular gene sequence. In aspects, the DNA-targeting sequence (e.g., sgRNA) is at least 95% complementary to the sequence of a cellular gene. In aspects, the DNA-targeting sequence (e.g., sgRNA) binds a cellular gene sequence.

[0153] A “target nucleic acid” or “target nucleic acid sequence” as provided herein is a nucleic acid sequence present in, or expressed by, a cell, to which a guide sequence (or a DNA-targeting sequence) is designed to have complementarity, where hybridization between a target sequence and a guide sequence (or a DNA-targeting sequence) promotes the formation of a complex (e.g., CRISPR complex). Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a complex (e.g., CRISPR complex). In aspects, the target polynucleotide sequence is an exogenous nucleic acid sequence. In aspects, the target polynucleotide sequence is an endogenous nucleic acid sequence.

[0154] The target polynucleotide sequence may be any region of the polynucleotide (e.g., DNA sequence) suitable for epigenome editing. In aspects, the target polynucleotide sequence is part of a gene. In aspects, the target polynucleotide sequence is part of a transcriptional regulatory sequence. In aspects, the target polynucleotide sequence is part of a promoter, enhancer or silencer. In aspects, the target polynucleotide sequence is part of a promoter. In aspects, the target polynucleotide sequence is part of an enhancer. In aspects, the target polynucleotide sequence is part of a silencer.

[0155] In embodiments, the target polynucleotide sequence is a hypomethylated nucleic acid sequence. A “hypomethylated nucleic acid sequence” is used herein according to the standard meaning in the art and refers to a loss or lack of methyl groups on the 5-methylcytosine

nucleotide (e.g., in CpG). The loss or lack of methyl groups may be relative to a standard control. Hypomethylation may occur, for example, in aging cells or in cancer (e.g., early stages of neoplasia) relative to the younger cell or non-cancer cell, respectively. Thus, the complex may be useful for reestablishing normal (e.g. non-aged or non-diseased) methylation levels.

[0156] In embodiments, the target polynucleotide sequence is within or adjacent to a transcription start site. In aspects, the target polynucleotide sequence is within about 3000, 2500, 2000, 1500, 500, 100, 80, 70, 60, 50, 40, 30, 20, 10, or fewer base pairs (bp) flanking a transcription start site.

[0157] In embodiments, the target polynucleotide sequence is at, near, or within a promoter sequence. In aspects, the target polynucleotide sequence is within a CpG island. In aspects, the target polynucleotide sequence is known to be associated with a disease or condition characterized by DNA hypomethylation or hypermethylation.

[0158] In embodiments, the target polynucleotide sequence include the sequence of SEQ ID NO:37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, or 95. In aspects, the target polynucleotide sequence include an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, or 95. In aspects, the target polynucleotide sequence is SEQ ID NO:37. In aspects, the target polynucleotide sequence is SEQ ID NO:39. In aspects, the target polynucleotide sequence is SEQ ID NO:41. In aspects, the target polynucleotide sequence is SEQ ID NO:43. In aspects, the target polynucleotide sequence is SEQ ID NO:45. In aspects, the target polynucleotide sequence is SEQ ID NO:47. In aspects, the target polynucleotide sequence is SEQ ID NO:49. In aspects, the target polynucleotide sequence is SEQ ID NO:51. In aspects, the target polynucleotide sequence is SEQ ID NO:53. In aspects, the target polynucleotide sequence is SEQ ID NO:55. In aspects, the target polynucleotide sequence is SEQ ID NO:57. In aspects, the target polynucleotide sequence is SEQ ID NO:59. In aspects, the target polynucleotide sequence is SEQ ID NO:61. In aspects, the target polynucleotide sequence is SEQ ID NO:63. In aspects, the target polynucleotide sequence is SEQ ID NO:65. In aspects, the target polynucleotide sequence is SEQ ID NO:67. In aspects, the target polynucleotide sequence is SEQ ID NO:69. In aspects, the target polynucleotide sequence is SEQ ID NO:71. In aspects, the target polynucleotide sequence is SEQ ID NO:73. In aspects, the target polynucleotide sequence is SEQ ID NO:75. In aspects, the target polynucleotide sequence is SEQ ID NO:77. In aspects, the target polynucleotide sequence is SEQ ID NO:79. In aspects, the target polynucleotide sequence is SEQ ID NO:81. In aspects, the target

polynucleotide sequence is SEQ ID NO:83. In aspects, the target polynucleotide sequence is SEQ ID NO:85. In aspects, the target polynucleotide sequence is SEQ ID NO:87. In aspects, the target polynucleotide sequence is SEQ ID NO:89. In aspects, the target polynucleotide sequence is SEQ ID NO:91. In aspects, the target polynucleotide sequence is SEQ ID NO:93. In aspects, the target polynucleotide sequence is SEQ ID NO:95.

[illegible]

85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:69. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:71. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:73. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:75. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:77. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:79. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:81. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:83. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:85. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:87. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:89. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:91. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:93. In aspects, the target polynucleotide sequence has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:95.

[0160] In embodiments, the complex includes dCas9 bound to the polynucleotide through binding a binding sequence of the polynucleotide and thereby forming a ribonucleoprotein complex. In aspects, the binding sequence forms a hairpin structure. In aspects, the binding sequence is 10-200 nt, 15-150 nt, 20-140 nt, 30-100 nt, 35-50 nt, 37-47 nt, or 42 nt in length.

[0161] In embodiments, the binding sequence (e.g., Cas9-binding sequence) interacts with or binds to a Cas9 protein (e.g., dCas9 protein), and together they bind to the target polynucleotide sequence recognized by the DNA-targeting sequence. The binding sequence (e.g., Cas9-binding sequence) includes two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (a dsRNA duplex). These two complementary stretches of

nucleotides may be covalently linked by intervening nucleotides known as linkers or linker nucleotides (*e.g.*, in the case of a single-molecule polynucleotide), and hybridize to form the double stranded RNA duplex (dsRNA duplex, or “Cas9-binding hairpin”) of the binding sequence (*e.g.*, Cas9-binding sequence), thus resulting in a stem-loop structure. Alternatively, in some aspects, the two complementary stretches of nucleotides may not be covalently linked, but instead are held together by hybridization between complementary sequences (*e.g.*, a two-molecule polynucleotide).

[0162] The binding sequence (*e.g.*, Cas9-binding sequence) can have a length of from 10 nucleotides to 200 nucleotides, *e.g.*, from 20 nucleotides (nt) to 150 nt. In aspects, the binding sequence has a length of from 80 nucleotides (nt) to 100 nt. The dsRNA duplex of the binding sequence (*e.g.*, Cas9-binding sequence) can have a length from 6 base pairs (bp) to 200 bp. For example, the dsRNA duplex of the binding sequence (*e.g.*, Cas9-binding sequence) can have a length from 6 bp to 200 bp, from 10 bp to 180 bp, from 10 bp to 150 bp, from 80 bp to 100 bp, and the like.

[0163] In embodiments, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO: 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, or 94 or their corresponding RNA sequence. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, or 94 or their corresponding RNA sequence. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:38. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:40. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:42. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:44. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:46. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:48. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:50. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:52. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID

NO:54. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:56. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:58. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:60. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:62. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:64. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:66. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:68. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:70. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:72. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:74. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:76. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:78. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:80. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:82. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:84. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:86. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:88. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:90. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:92. In aspects, the polynucleotide that forms a complex with a fusion protein described herein includes the sequence of SEQ ID NO:94.

[0164] Nucleic Acids and Vectors

[0165] The fusion protein described herein, including embodiments and aspects thereof, may be provided as a nucleic acid sequence that encodes for the fusion protein. Thus, in an aspect is provided a nucleic acid sequence encoding the fusion protein described herein, including embodiments and aspects thereof. In an aspect is provided a nucleic acid sequence encoding the

fusion protein described herein (including the DNA-targeting sequence), including embodiments and aspects thereof. In aspects, the nucleic acid sequence encodes for a fusion protein described herein, including fusion proteins having amino acid sequences with certain % sequence identities described herein. In aspects, the nucleic acid is RNA. In aspects, the nucleic acid is messenger RNA. In aspects, fusion protein is delivered as DNA, mRNA, protein or an RNP. For RNP the protein would be dCas9 and the RNA would encode an sgRNA. Similarly the sgRNA could be delivered as DNA encoding a promoter and an sgRNA, RNA encoding a promoter and an sgRNA. In aspects, the nucleic acid sequence encodes for the fusion proteins described herein, including embodiments and aspects thereof. In aspects, the nucleic acid sequence encodes for the fusion protein of any one of SEQ ID NOS:1-15. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:97. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:98. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:99.

[0166] It is further contemplated that the nucleic acid sequence encoding the fusion protein as described herein, including embodiments and aspects thereof, may be included in a vector. Therefore, in an aspect is provided a vector including a nucleic acid sequence as described herein, including embodiments and aspects thereof. In aspects, the vector comprises a nucleic acid sequence that encodes for a fusion protein described herein, including fusion proteins having amino acid sequences with certain % sequence identities described herein. In aspects, the nucleic acid is messenger RNA. In aspects, the messenger RNA is messenger RNP. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of any one of SEQ ID NOS:1-15. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:97. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:98. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:99.

[0167] In embodiments, the vector further includes a polynucleotide, wherein the polynucleotide includes: (1) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (2) a binding sequence for the nuclease-deficient RNA-guided DNA endonuclease enzyme. In aspects, the vector further includes a polynucleotide, wherein the polynucleotide includes sgRNA. In aspects, the vector further includes a polynucleotide, wherein the polynucleotide includes cr:tracrRNA. Thus, one or more vectors may include all necessary components for performing epigenome editing.

[0168] Cells

[0169] The compositions described herein may be incorporated into a cell. Inside the cell, the compositions as described herein, including embodiments and aspects thereof, may perform epigenome editing. Accordingly, in an aspect is provided a cell including a fusion protein as described herein, including embodiments and aspects thereof, a nucleic acid as described herein, including embodiments and aspects thereof, a complex as described herein, including embodiments and aspects thereof, or a vector as described herein, including embodiments and aspects thereof. In aspects is provided a cell including a fusion protein as described herein, including embodiments and aspects thereof. In aspects is provided a cell including a nucleic acid as described herein, including embodiments and aspects thereof. In aspects is provided a cell including a complex as described herein, including embodiments and aspects thereof. In aspects is provided a cell including a vector as described herein, including embodiments and aspects thereof. In aspects, the cell is a eukaryotic cell. In aspects, the cell is a mammalian cell.

[0170] Methods

[0171] The fusion proteins described herein program a durable memory of gene silencing over time. As shown in the examples, the fusion proteins (e.g., SEQ ID NOS:15, 97, 98, 99, 107, 108) program a durable memory of gene silencing with over 80% of transfected cells silencing the *Snrpn*-GFP reporter and over 90% the endogenously GFP-tagged gene *HIST2H2BE* (H2B) at 50 days post-transfection (**FIGS. 1E, 1F, 8F**). Notably, starting at 10 days post-transfection, no fusion protein (i.e., CRISPR-off protein) was detected, indicating that the observed gene silencing was independent of constitutive expression of the fusion protein (**FIG. 1E**). The data in **FIGS. 1E** and **1F** is transfection of DNA encoding CRISPRoff. This skilled artisan could alternatively transfect a CRISPRoff RNP consisting of CRISPRoff protein and an sgRNA that form a complex in vitro. **FIG. 1L** also demonstrates that transient expression of CRISPRoff for less than 10 days results in stable gene silencing of the endogenous *CLTA* gene at 15 months post transfection. 15 months of time is more than 450 cell divisions which is more than most cells in the adult human body divide over the course of a lifetime. Gene silencing is achieved by transfection of mRNA encoding the fusion proteins described herein. Thus, transient expression of the fusion protein leads to effective gene silencing (**FIG. 8G**). CRISPRoff epigenetic memory using the fusion proteins described herein is propagated by the cell rather than by sustained transgene expression.

[0172] In embodiments, the disclosure provides methods of silencing a target nucleic acid sequence in a cell, the method comprising: (i) delivering a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme), to a cell containing the target

nucleic acid; and (ii) delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of silencing a target nucleic acid sequence in a cell, the method comprising delivering a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme), to a cell containing the target nucleic acid; thereby silencing the target nucleic acid sequence. In aspects, the target nucleic acid comprises a CpG island. In aspects, the target nucleic acid includes a non-CpG island. In aspects, the target nucleic acid comprises a CpG island and a non-CpG island. “Comprises a CpG island” or “comprises a non-CpG island” refers to one or more CpG islands or non-CpG islands, respectively. In aspects, the target nucleic acid sequence comprises a plurality of CpG islands (e.g., 2, 3, 4, 5, or more CpG islands). In aspects, the target nucleic acid sequence comprises a plurality of non-CpG islands (e.g., 2, 3, 4, 5, or more non-CpG islands). In aspects, the target nucleic acid sequence does not comprise a CpG island and does not comprises a non-CpG island. In aspects, the method of silencing the target nucleic acid sequence is a method of treating Angelman syndrome in a patient in need thereof. In aspects, the method of silencing the target nucleic acid sequence is a method of treating a viral infection in a patient in need thereof. In aspects, the method of silencing the target nucleic acid sequence is a method of treating an infectious disease in a patient in need thereof.

[0173] In embodiments, the disclosure provides methods of silencing a target nucleic acid sequence in a cell, the method comprising: delivering a polynucleotide: (a) encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme), and (b) comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence. In embodiments, the polynucleotide comprises sgRNA. In embodiments, the polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of silencing a target nucleic acid sequence in a cell, the method comprising delivering a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme), to a cell containing the target nucleic acid; thereby silencing the target nucleic acid sequence. In aspects, the target nucleic acid comprises a CpG island. In aspects, the target nucleic acid comprises a non-CpG island. In aspects, the target nucleic acid comprises a CpG island and a non-CpG island. “Comprises a CpG island” or “comprises a non-CpG island” refers to one or more CpG islands or non-CpG

islands, respectively. In aspects, the target nucleic acid sequence comprises a plurality of CpG islands (e.g., 2, 3, 4, 5, or more CpG islands). In aspects, the target nucleic acid sequence comprises a plurality of non-CpG islands (e.g., 2, 3, 4, 5, or more non-CpG islands). In aspects, the target nucleic acid sequence does not comprise a CpG island and does not comprises a non-CpG island. In aspects, the method of silencing the target nucleic acid sequence is a method of treating Angelman syndrome in a patient in need thereof. In aspects, the method of silencing the target nucleic acid sequence is a method of treating a viral infection in a patient in need thereof. In aspects, the method of silencing the target nucleic acid sequence is a method of treating an infectious disease in a patient in need thereof. In aspects, the method of silencing the target nucleic acid sequence is a method of treating a neurodegenerative disease in a patient in need thereof.

[0174] In embodiments, the disclosure provides methods of treating an infectious disease in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating the infectious disease in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of treating an infectious disease in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating the infectious disease.

[0175] The term “infection” or “infectious disease” refers to a disease caused by organisms such as a bacterium, virus, fungi, or any other pathogenic microbial agents. In embodiments, the infectious disease is caused by bacteria. In embodiments, the infectious disease is a bacteria associated disease (e.g., tuberculosis, which is caused by *Mycobacterium tuberculosis*). Non-limiting infectious diseases caused by bacteria include pneumonia (e.g., *Streptococcus*, *Pseudomonas*); or foodborne illnesses (e.g., *Shigella*, *Campylobacter*, *Salmonella*). Infectious diseases caused by bacteria also includes tetanus, typhoid fever, diphtheria, syphilis, and leprosy. In embodiments, the infectious disease is bacterial vaginosis (i.e. bacteria that change the vaginal microbiota caused by an overgrowth of bacteria that crowd out the *Lactobacilli* species that maintain healthy vaginal microbial populations) (e.g., yeast infection, or

Trichomonas vaginalis); bacterial meningitis (i.e., a bacterial inflammation of the meninges); bacterial pneumonia (i.e., a bacterial infection of the lungs); a urinary tract infection; bacterial gastroenteritis; or bacterial skin infections (e.g. impetigo, or cellulitis). In embodiments, the infectious disease is a *Campylobacter jejuni*, *Enterococcus faecalis*, *Haemophilus influenzae*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Legionella pneumophila*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, or *Vibrio cholera* infection. In embodiments, the infectious disease is caused by fungi. In embodiments, the infectious disease is caused by a virus.

[0176] In embodiments, the disclosure provides methods of treating a bacterial infection or a fungal infection in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating the bacterial infection or the fungal infection in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of treating a bacterial infection or a fungal infection in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating the bacterial infection or the fungal infection. In embodiments, the methods comprise treating a bacterial infection. In embodiments, the methods comprise treating a fungal infection.

[0177] In embodiments, the disclosure provides methods of treating a viral infection in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating the viral infection in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of treating a viral infection in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g.,

comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating the viral infection. In embodiments, the viral infection is a Flavivirus infection. In embodiments, the Flavivirus infection is caused by West Nile virus, dengue virus, tick-borne encephalitis virus, yellow fever virus, or Zika virus. In embodiments, the disclosure provides methods of treating a Flavivirus infection in a subject in need thereof by the methods described herein. In embodiments, the disclosure provides methods of treating a West Nile virus infection in a subject in need thereof by the methods described herein. In embodiments, the disclosure provides methods of treating a dengue virus infection in a subject in need thereof by the methods described herein. In embodiments, the disclosure provides methods of treating a tick-borne encephalitis virus infection in a subject in need thereof by the methods described herein. In embodiments, the disclosure provides methods of treating a yellow fever virus infection in a subject in need thereof by the methods described herein. In embodiments, the disclosure provides methods of treating a Zika virus infection in a subject in need thereof by the methods described herein.

[0178] The term “viral infection” or “viral disease” refers to a disease or condition that is caused by a virus. Non-limiting examples of viral infections include hepatic viral diseases (e.g., hepatitis A, B, C, D, E), herpes virus infection (e.g., HSV-1, HSV-2, herpes zoster), flavivirus infection (e.g., Zika virus infection, dengue virus infection, yellow fever virus infection, West Nile virus infection tick-borne encephalitis virus infection), cytomegalovirus infection, a respiratory viral infection (e.g., adenovirus infection, influenza, severe acute respiratory syndrome, coronavirus infection (e.g., SARS-CoV-1, SARS-CoV-2, MERS-CoV, COVID-19, MERS)), a gastrointestinal viral infection (e.g., norovirus infection, rotavirus infection, astrovirus infection), an exanthematous viral infection (e.g., measles, shingles, smallpox, rubella), viral hemorrhagic disease (e.g., Ebola, Lassa fever, dengue fever, yellow fever), a neurologic viral infection (e.g., West Nile viral infection, polio, viral meningitis, viral encephalitis, Japanese encephalitis, rabies), and human papilloma viral infection.

[0179] In embodiments, the disclosure provides methods of treating a tau pathology in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating the tau pathology in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the

disclosure provides methods of treating a tau pathology in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating the tau pathology. Tau pathology refers to neurodegenerative diseases characterized by pathological tau aggregation in neurofibrillary tangles (NFTs). Diseases with this typical pathological feature are called tauopathies and include, for example Alzheimer's disease, Parkinson's disease, progressive supranuclear palsy, Huntington's disease, amyotrophic lateral sclerosis, Pick's disease, dementia pugilistica, and frontotemporal dementia. In embodiments, the tau pathology is Alzheimer's disease. In embodiments, the tau pathology is Parkinson's disease. In embodiments, the tau pathology is Parkinson's disease linked to chromosome 17. In embodiments, the tau pathology is progressive supranuclear palsy. In embodiments, the tau pathology is Huntington's disease. In embodiments, the tau pathology is amyotrophic lateral sclerosis. In embodiments, the tau pathology is Pick's disease. In embodiments, the tau pathology is dementia pugilistica. In embodiments, the tau pathology is frontotemporal dementia. In embodiments, the tau pathology is frontotemporal dementia linked to chromosome 17.

[0180] In embodiments, the disclosure provides methods of treating a neurodegenerative disease in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating the neurodegenerative disease in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA (e.g., gRNA). In embodiments, the disclosure provides methods of treating a neurodegenerative disease in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating the neurodegenerative disease.

[0181] As used herein, the term "neurodegenerative disorder" or "neurodegenerative disease" refers to a disease or condition in which the function of a subject's nervous system becomes impaired. Examples of neurodegenerative diseases that may be treated with the fusion proteins and methods described herein include Alexander's disease, Alper's disease, Alzheimer's disease,

amyotrophic lateral sclerosis, ataxia telangiectasia, Batten disease (also known as Spielmeyer-Vogt-Sjogren-Batten disease), bovine spongiform encephalopathy (BSE), canavan disease, chronic fatigue syndrome, cockayne syndrome, corticobasal degeneration, Creutzfeldt-Jakob disease, frontotemporal dementia, Gerstmann-Sträussler-Scheinker syndrome, Huntington's disease, HIV-associated dementia, Kennedy's disease, Krabbe's disease, kuru, lewy body dementia, Machado-Joseph disease (Spinocerebellar ataxia type 3), multiple sclerosis, multiple system atrophy, myalgic encephalomyelitis, narcolepsy, neuroborreliosis, Parkinson's disease, Pelizaeus-Merzbacher Disease, Pick's disease, primary lateral sclerosis, prion diseases, Refsum's disease, Sandhoff's disease, Schilder's disease, subacute combined degeneration of spinal cord secondary to pernicious anaemia, schizophrenia, spinocerebellar ataxia (multiple types with varying characteristics), spinal muscular atrophy, Steele-Richardson-Olszewski disease, progressive supranuclear palsy, and tabes dorsalis.

[0182] Angelman syndrome (AS) is a neurological genetic disorder caused by loss of expression of the maternal copy of *UBE3A* in the brain. Due to brain-specific genetic imprinting at this locus, the paternal *UBE3A* is also silenced by an antisense transcript. The methods described herein can be used to inhibit the antisense transcript, thereby unsilencing the paternal *UBE3A*, which results in the treatment of Angelman syndrome. *See, e.g.,* Bailus et al, *Mol Ther*, 24(3):548-555 (2016). In other words, the methods described herein silence the negative regulator of the silenced paternal *UBE3A*, which equates to unsilencing the paternal *UBE3A*.

[0183] In embodiments, the disclosure provides methods of treating Angelman syndrome in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme); and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating Angelman syndrome in the subject. In embodiments, the second polynucleotide comprises sgRNA. In embodiments, the second polynucleotide comprises two different sgRNA. In embodiments, the disclosure provides methods of treating Angelman syndrome in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein, as described herein, including all embodiments and aspects thereof (e.g., comprising a nuclease-deficient DNA endonuclease enzyme); thereby treating Angelman syndrome in the subject

[0184] In aspects, the sequence is within the target nucleic acid sequence is methylated. In aspects, the sequence that is within about 3000, 2900, 2800, 2700, 2600, 2500, 2400, 2300, 2200, 2100, 2000, 1900, 1800, 1700, 1600, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700,

600, 500, 400, 300, 200, 100, 50, 20, or 10 base pairs of the target nucleic acid sequence is methylated. Without intending to be bound by any theory, methylating a chromatin means that DNA is methylated at the C nucleotide of CG sequences found in CpG islands or non-CpG islands (i.e., adding methyl marks at the C nucleotide of CG DNA sites found in CpG islands).

[0185] The term “repressive chromatin markers” as used herein refers to modifications made to the chromatin that result in silencing (e.g., decreasing or inhibiting of transcription) of the target nucleic acid sequence (e.g., a gene). Examples of repressive chromatin markers include, but are not limited to, mono-, di-, and/or tri-methylation, acetylation/deacetylation, phosphorylation, and ubiquitination of histones (e.g., H3K9, H3K27, H3K79, H2BK5).

[0186] The term “CpG island” is used in its customary sense to refer to regions in a nucleic acid that have a high frequency of the nucleotides G and C next to one another (i.e., CpG dinucleotides). In aspects, a CpG island refers to a region of a nucleic acid sequence having a region with a GC content greater than 50%, with an observed-to-expected CpG ratio greater than 60%. In aspects, a CpG island refers to a region of a nucleic acid sequence having at least 50 base pair, and a GC content greater than 50%, with an observed-to-expected CpG ratio greater than 60%. In aspects, a CpG island refers to a region of a nucleic acid sequence having at least 100 base pair, and a GC content greater than 50%, with an observed-to-expected CpG ratio greater than 60%. In aspects, a CpG island refers to a region of a nucleic acid sequence having at least 150 base pair, and a GC content greater than 50%, with an observed-to-expected CpG ratio greater than 60%. In aspects, a CpG island refers to a region of a nucleic acid sequence having at least 200 base pair, and a GC content greater than 50%, with an observed-to-expected CpG ratio greater than 60%. The percentage CpG is the ratio of CpG nucleotide bases (twice the CpG count) to the length. The ratio of observed to expected CpG is calculated according to the formula:

$$\text{Obs/Exp CpG} = \text{Number of CpG} * N / (\text{Number of C} * \text{Number of G}),$$

where N = length of sequence. *See* Gardiner-Garden et al, *Journal of Molecular Biology*, 196(2):261-282 (1987)).

[0187] The phrase “target nucleic acid does not comprise a CpG island” or “target nucleic acid that does not comprise a CpG island” or “non-CpG island” refers to a target nucleic acid that does not contain a “CpG island” as that term is defined herein. This region can be any region encoded by a mammalian (e.g., human) genome. In aspects, the phrase “target nucleic acid does not comprise a CpG island” refers to regions in a target nucleic acid that have do not have the nucleotides G and C next to one another (i.e., CpG dinucleotides) or that have a low frequency of the nucleotides G and C next to one another. In aspects, a non-CpG island refers to regions of

a target nucleic acid having a region with a GC dinucleotide content less than 50%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 50%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 50%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 50%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 50%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 45%, with an observed-to-expected CpG ratio less than 55%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content less than 40%, with an observed-to-expected CpG ratio less than 50%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 1% to 45%, with an observed-to-expected CpG ratio of less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 1% to 45%, with an observed-to-expected CpG ratio less than 55%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 1% to 45%, with an observed-to-expected CpG ratio less than 50%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 5% to 40%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 5% to 40%, with an observed-to-expected CpG ratio less than 55%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 5% to 40%, with an observed-to-expected CpG ratio less than 50%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 10% to 40%, with an observed-to-expected CpG ratio less than 60%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 10% to 40%, with an observed-to-expected CpG ratio less than 55%. In aspects, a non-CpG island refers to regions of a target nucleic acid having a GC dinucleotide content of 10% to 40%, with an observed-to-expected CpG ratio less than 50%. In aspects, the target nucleic acid that does not comprise a CpG island has at least 10 base pairs. In aspects, the target nucleic acid that does not comprise a CpG island has at least 50 base pairs. In aspects, the target nucleic acid that does not comprise a CpG island has at least 100 base pairs. In aspects, the target nucleic acid that does not comprise a CpG island has at least 150 base pairs. In aspects, the target nucleic acid that does not comprise a CpG island has at least 200 base pairs.

[0188] In embodiments, silencing refers to a complete suppression of transcription. In aspects, silencing refers to a significant decrease in transcription compared to control levels of transcription.

[0189] In embodiments, the first polynucleotide is contained within a first vector. In aspects, the first polynucleotide is contained within a second vector. In aspects, the first vector and the second vector are the same. In aspects, the first vector is different from the second vector.

[0190] In embodiments, the polynucleotide described herein is delivered into the cell by any method known in the art, for example, by transfection, electroporation or transduction.

[0191] Alternatively, in an aspect is provided a method of silencing a target nucleic acid sequence in a cell, including delivering a complex as described herein, including embodiments and aspects thereof, to a cell containing the target nucleic acid. Without intending to be bound by any theory, the complex silences the target nucleic acid sequence in the cell by methylating a chromatin containing the target nucleic acid sequence and/or by introducing repressive chromatin marks to a chromatin containing the target nucleic acid sequence.

[0192] In embodiments, the method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence. In aspects, the method has a specificity that is at least 2-fold (e.g., 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 15-, 20-, 25-fold) higher than a specificity to a non-target nucleic acid sequence. Methods for determining specificity are well known in the art and include, but are not limited to, RNA-seq, bisulfite sequencing, chromatin immunoprecipitation, flow cytometry, and qPCR. Thus, in aspects, specificity is determined by RNA-seq. In aspects, specificity is determined by bisulfite sequencing. In aspects, specificity is determined by chromatin immunoprecipitation. In aspects, specificity is determined by flow cytometry. In aspects, specificity is determined by qPCR.

[0193] In aspects, the complex is delivered into the cell via any method known in the art. In aspects, the complex is delivered to the cell via RNA, DNA, or ribonucleoprotein (RNP) delivery. In aspects, the complex is delivered into the cell via RNA. In aspects, the complex is delivered to the cell via DNA. In aspects, the complex is delivered to the cell via transfection, virus, lipid nanoparticle (LNP) or viral-like particles. In aspects, the complex is delivered to the cell via transfection. In aspects, the complex is delivered to the cell via virus. In aspects, the complex is delivered to the cell via lipid nanoparticle. Methods for delivery complexes into a cell are well known in the art.

[0194] Embodiments 1 to 76

[0195] Embodiment 1. A fusion protein comprising, from N-terminus to C-terminus, a DNA

methyltransferase domain, a first XTEN linker, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain.

[0196] Embodiment 2. The fusion protein of Embodiment 1, wherein the first XTEN linker comprises from about 5 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to about 864 amino acid residues.

[0197] Embodiment 3. The fusion protein of Embodiment 2, wherein the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues, and a Krüppel-associated box domain.

[0198] Embodiment 4. The fusion protein of Embodiment 3, wherein the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues.

[0199] Embodiment 5. The fusion protein of Embodiment 4, wherein the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues.

[0200] Embodiment 6. The fusion protein of any one of Embodiments 1 to 5, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.

[0201] Embodiment 7. The fusion protein of Embodiment 6, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.

[0202] Embodiment 8. The fusion protein of Embodiment 6, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.

[0203] Embodiment 9. The fusion protein of any one of Embodiments 1 to 8, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

[0204] Embodiment 10. The fusion protein of Embodiment 9, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).

[0205] Embodiment 11. The fusion protein of any one of Embodiments 1 to 10, further comprising an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0206] Embodiment 12. The fusion protein of Embodiment 11, comprising, from N-terminus

to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, a nuclear localization signal peptide, and the Krüppel-associated box domain.

[0207] Embodiment 13. The fusion protein of Embodiment 11, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, a nuclear localization signal peptide, the second XTEN linker, and the Krüppel-associated box domain.

[0208] Embodiment 14. The fusion protein of Embodiment 11, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, the second XTEN linker, the Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag.

[0209] Embodiment 15. A fusion protein having at least 85% sequence identity to the amino acid sequence of Formula (A); where the amino acid sequence of Formula (A) is, from N-terminus to C-terminus: $C_1-R^3-C_2-R^2-A-R^1-R^4-B$ (A), wherein: C_1 comprises SEQ ID NO:26 or SEQ ID NO:106; R^3 is absent or R^3 comprises SEQ ID NO:27; C_2 comprises SEQ ID NO:28; R^2 is absent or R^2 comprises SEQ ID NO:32; A comprises SEQ ID NO:23; R^1 is absent or R^1 comprises SEQ ID NO:25; R^4 is absent or R^4 comprises SEQ ID NO:31; and B comprises SEQ ID NO:16, SEQ ID NO:103, SEQ ID NO:104, or SEQ ID NO:105.

[0210] Embodiment 16. The fusion protein of Embodiment 15 having at least 90% sequence identity to the amino acid sequence of Formula (A).

[0211] Embodiment 17. The fusion protein of Embodiment 16 having at least 95% sequence identity to the amino acid sequence of Formula (A).

[0212] Embodiment 18. A fusion protein having at least 85% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110.

[0213] Embodiment 19. The fusion protein of Embodiment 18 having at least 90% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110.

[0214] Embodiment 20. A cell comprising the fusion protein of any one of Embodiments 1 to 19.

[0215] Embodiment 21. The cell of Embodiment 20, wherein the cell is a eukaryotic cell, a mammalian cell, or a stem cell.

[0216] Embodiment 22. A method of silencing a target nucleic acid sequence in a cell, the

method comprising: (i) delivering a first polynucleotide encoding a fusion protein of any one of Embodiments 1 to 19 to a cell containing the target nucleic acid; and (ii) delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence.

[0217] Embodiment 23. The method of Embodiment 22, wherein the target nucleic acid comprises a CpG island

[0218] Embodiment 24. The method of Embodiment 22, wherein the target nucleic acid does not comprise a CpG island.

[0219] Embodiment 25. The method of any one of Embodiments 22 to 24, wherein the second polynucleotide comprises sgRNA.

[0220] Embodiment 26. A method of silencing a target nucleic acid sequence in a cell, the method comprising: (i) delivering a first polynucleotide encoding a fusion protein to a cell containing the target nucleic acid, wherein the target nucleic acid does not comprise a CpG island; wherein the fusion protein comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain; and (ii) delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence in the cell.

[0221] Embodiment 27. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain; and (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating Angelman syndrome, the infectious disease, the tau pathology, or the neurodegenerative disease.

[0222] Embodiment 28. The method of Embodiment 26 or 27, wherein the second polynucleotide comprises sgRNA.

[0223] Embodiment 29. The method of any one of Embodiments 26 to 28, wherein the fusion protein comprises, from N-terminus to C-terminus, the DNA methyltransferase domain, the nuclease-deficient RNA-guided DNA endonuclease enzyme, and the Krüppel associated box domain.

[0224] Embodiment 30. The method of any one of Embodiments 26 to 29, wherein the

nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.

[0225] Embodiment 31. The method of Embodiment 30, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.

[0226] Embodiment 32. The method of Embodiment 30, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.

[0227] Embodiment 33. The method of any one of Embodiments 26 to 32, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

[0228] Embodiment 34. The method of Embodiment 33, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).

[0229] Embodiment 35. The method of any one of Embodiments 26 to 34, wherein the dCas9 is covalently linked to the Dnmt3A domain via a peptide linker and wherein the Dnmt3A domain is covalently linked to the Krüppel associated box domain via a peptide linker.

[0230] Embodiment 36. The method of Embodiment 35, wherein the peptide linker is a XTEN linker.

[0231] Embodiment 37. The method of any one of Embodiments 26 to 28, wherein the fusion protein comprises, from N-terminus to C-terminus, the Krüppel associated box, the nuclease-deficient RNA-guided DNA endonuclease enzyme, and the DNA methyltransferase domain.

[0232] Embodiment 38. The method of Embodiment 37, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.

[0233] Embodiment 39. The method of Embodiment 38, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.

[0234] Embodiment 40. The method of Embodiment 38, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.

[0235] Embodiment 41. The method of any one of Embodiments 37 to 39, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

[0236] Embodiment 42. The method of Embodiment 41, wherein the Dnmt3A domain is

linked to a Dnmt3L domain (Dnmt3A-3L domain).

[0237] Embodiment 43. The method of any one of Embodiments 39 to 42, wherein the dCas9 is covalently linked to the Dnmt3A domain via a peptide linker and wherein the Krüppel associated box domain is covalently linked to the dCas9 via a peptide linker.

[0238] Embodiment 44. The method of Embodiment 43, wherein the peptide linker is a XTEN linker.

[0239] Embodiment 45. The method of any one of Embodiments 37 to 44, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the Krüppel associated box domain via a peptide linker.

[0240] Embodiment 46. The method of any one of Embodiments 37 to 45, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0241] Embodiment 47. The method of any one of Embodiments 37 to 46, wherein the Krüppel associated box domain is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0242] Embodiment 48. The method of any one of Embodiments 26 to 47, wherein the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0243] Embodiment 49. The method of any one of Embodiments 26 to 47, wherein the fusion protein further comprises a nuclear localization signal peptide.

[0244] Embodiment 50. The method of any one of Embodiments 26 to 28, wherein the fusion protein having at least 85% sequence identity to the amino acid sequence of Formula (A); where the amino acid sequence of Formula (A) is, from N-terminus to C-terminus: C₁-R³-C₂-R²-A-R¹-R⁴-B (A), wherein: C₁ comprises SEQ ID NO:26 or SEQ ID NO:106; R³ is absent or R³ comprises SEQ ID NO:27; C₂ comprises SEQ ID NO:28; R² is absent or R² comprises SEQ ID NO:32; A comprises SEQ ID NO:23; R¹ is absent or R¹ comprises SEQ ID NO:25; R⁴ is absent or R⁴ comprises SEQ ID NO:31; and B comprises SEQ ID NO:16, SEQ ID NO:103, SEQ ID NO:104, or SEQ ID NO:105.

[0245] Embodiment 51. The method of Embodiment 50, wherein the fusion protein has at least 90% sequence identity to the amino acid sequence of Formula (A).

[0246] Embodiment 52. The method of Embodiment 51, wherein the fusion protein has at least 95% sequence identity to the amino acid sequence of Formula (A).

[0247] Embodiment 53. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising: (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein of any one of Embodiments 1 to 14; and (ii) delivering to the subject an effective amount of second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating Angelman syndrome, the infectious disease, the tau pathology, or the neurodegenerative disease in the subject.

[0248] Embodiment 54. The method of any one of Embodiments 26 to 53, wherein the infectious disease is a viral infections disease.

[0249] Embodiment 55. The method of Embodiment 54, wherein the infectious disease is a Flavivirus infectious disease.

[0250] Embodiment 56. A fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain.

[0251] Embodiment 57. The fusion protein of Embodiment 56, wherein the first XTEN linker comprises from about 5 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to about 864 amino acid residues.

[0252] Embodiment 58. The fusion protein of Embodiment 57, wherein the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues.

[0253] Embodiment 59. The fusion protein of Embodiment 58, wherein the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues.

[0254] Embodiment 60. The fusion protein of Embodiment 59, wherein the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues.

[0255] Embodiment 61. The fusion protein of any one of Embodiments 56 to 60, wherein the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or a transcription activator-like effector.

[0256] Embodiment 62. The fusion protein of Embodiment 62, wherein the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain.

[0257] Embodiment 63. The fusion protein of Embodiment 62, wherein the nuclease-deficient

DNA endonuclease enzyme is transcription activator-like effector.

[0258] Embodiment 64. The fusion protein of any one of Embodiments 56 to 63, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

[0259] Embodiment 65. The fusion protein of Embodiment 64, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).

[0260] Embodiment 66. The fusion protein of any one of Embodiments 56 to 65, further comprising an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

[0261] Embodiment 67. The fusion protein of any one of Embodiments 56 to 65, further comprising a nuclear localization signal peptide.

[0262] Embodiment 68. The fusion protein of Embodiment 56, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, the second XTEN linker, the Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag.

[0263] Embodiment 69. A cell comprising the fusion protein of any one of Embodiments 56 to 68.

[0264] Embodiment 70. The cell of Embodiment 69, wherein the cell is a eukaryotic cell, a mammalian cell, or a stem cell.

[0265] Embodiment 71. A method of silencing a target nucleic acid sequence in a cell, the method comprising delivering a first polynucleotide encoding a fusion protein of any one of Embodiments 56 to 68 to a cell containing the target nucleic acid; thereby silencing the target nucleic acid sequence.

[0266] Embodiment 72. The method of Embodiment 71, wherein the target nucleic acid comprises a CpG island.

[0267] Embodiment 73. The method of Embodiment 71, wherein the target nucleic acid does not comprise a CpG island.

[0268] Embodiment 74. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein of any one of Embodiments 56 to 68; thereby treating Angelman syndrome, the infectious disease, the tau pathology, or the neurodegenerative disease.

[0269] Embodiment 75. The method of Embodiment 74, wherein the infectious disease is a viral infectious disease.

[0270] Embodiment 76. The method of Embodiment 75, wherein the viral infectious disease is a Flavivirus infectious disease.

EXAMPLES

[0271] Embodiments and aspects herein are further illustrated by the following examples. The examples are merely intended to illustrate embodiments and aspects, and are not to be construed to limit the scope herein.

[0272] The technology described herein allows for, inter alia, permanent silencing of genes in mammalian cells without generating double stranded DNA breaks in the host genome. In embodiments, the central component is a single polypeptide chain composed of catalytically inactive Cas9 (dCas9) fused to Dnmt3A, Dnmt3L, and a KRAB domain (herein referred to as an “all-in-one protein”). In embodiments, the central component is a single polypeptide chain composed of a zinc finger domain fused to Dnmt3A, Dnmt3L, and a KRAB domain (herein referred to as an “all-in-one protein”). This fusion proteins provided herein can be directed to a specific site in a mammalian genome using a single guide RNA (sgRNA) and may add DNA methylation and/or repressive chromatin marks to the site. The result is gene silencing that is inheritable across subsequent cell divisions. The fusion protein provided herein (and sgRNA) are only expressed transiently, bypassing the use of viral delivery methods to induce permanent silencing.

[0273] The fusion proteins provided herein provide a robust long-term or permanent silencing of endogenous gene expression by epigenome editing rather than genome editing. Both alleles of a gene may be targeted or a single pathogenic allele may be selectively targeted. An advantage of the fusion protein provided herein is that epigenetic editing is reversible and therefore inherently safer than genome editing. Thus, the fusion proteins provided herein are useful in prophylactic applications. For example, gene silencing can enable acute protection from an infection/biologic toxin and then be reversed after the risk of infection or intoxication is absent. Thus, the fusion proteins provided herein is useful for viral or toxin that enters a cell through interaction with a protein that is required for long term organ function or homeostasis. The fusion proteins provided herein are useful in genome editing based therapeutics.

[0274] Permanent gene silencing in mammalian cells can be accomplished with two components: a single polypeptide chain composed of dCas9 fused to three epigenetic modulators and a single guide RNA that directs the protein to a specific site in the host genome. In

embodiments, permanent gene silencing in mammalian cells can be accomplished with two components: a single polypeptide chain composed of a zinc finger domain fused to three epigenetic modulators and a single guide RNA that directs the protein to a specific site in the host genome. In embodiments, the components are only expressed transiently in the host cell, thus reducing toxicity and off-target events. In embodiments, the fusion protein provided herein does not induce DNA breaks in the host cell for permanent gene silencing. In embodiments, the epigenetic marks that are added to the genomic site of interest are reversible, thus allowing for removal of any off-target events that may occur.

[0275] Example 1

[0276] Advances in gene editing have transformed our ability to modify the human genome. In particular, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated protein 9) and other CRISPR systems can be programmed with a single guide (sg)RNA to introduce DNA breaks at a specified site to inactivate gene function or to stimulate precise DNA editing by homology-directed repair (Knott and Doudna, 2018). Additionally, base and prime editing strategies allow for precise DNA sequence modifications but generally rely on one or more DNA single strand nicks (Anzalone et al., 2020). These technologies have been optimized for targeted changes in the underlying DNA sequence and are therefore ideally suited for repairing or introducing pathogenic mutations. However, the reliance on endogenous DNA repair machinery presents challenges as the complexity of these pathways can make it difficult to limit the outcome to a single desired change (Yeh et al., 2019).

[0277] An alternative modality for modulating gene function is to rewrite the epigenetic landscape to control gene expression without changing the underlying DNA sequence. Fusing protein scaffold or enzyme domains to catalytically inactive dCas9 can enhance (CRISPRa) or repress (CRISPRi) transcription in mammalian cells (Holtzman and Gersbach, 2018; Xu and Qi, 2019). Programmable epigenome editing is tunable, reversible, and does not require DNA breaks, effectively bypassing the cellular toxicity associated with gene editing (Jost et al., 2020). However, current programmable epigenome editing technologies typically rely on constitutive expression of dCas9-fusion proteins to maintain transcriptional control. As such, these modalities remain less suitable for therapeutic cell and organismal engineering.

[0278] Recent work has demonstrated that it is possible for epigenome editing to write a stable transcriptional program that is remembered and propagated by human cells without constitutive expression of the programmable epigenetic modulators (Amabile et al., 2016; Bintu et al., 2016; Park et al., 2019; Van et al., 2021). In particular, Amabile et al. showed that it was possible to

heritably silence human genes by recruitment of a cocktail of DNA methyltransferase and KRAB domains. However, to date only a small number of endogenous human loci have been tested for silencing by epigenetic memory writers (Amabile et al., 2016; O'Geen et al., 2019; Tarjan et al., 2019). Moreover, previous designs of programmable epigenetic silencers utilize either two or three fusion proteins for each target gene, which is experimentally cumbersome—especially for multiplexed gene targeting—and complicates an optimal gene targeting strategy. Furthermore, a TALE-based fusion of KRAB and the DNMT3A and DNMT3L domains resulted in low efficacy long-term gene silencing (Mlambo et al., 2018). Thus, it is unclear how generalizable these approaches are for establishing heritable gene silencing and whether there are genomic features that are required for writing and maintaining heritable epigenetic silencing. We hypothesized that an epigenetic editor composed of a single dead Cas9 fusion would enable us to broadly explore the biology and utility of heritable epigenetic gene silencing.

[0279] Here, we present the design, development, and characterization of CRISPRoff, a programmable epigenetic memory writer protein that can durably silence gene expression. Transient expression of CRISPRoff writes an epigenetic program that human cells maintain for more than 450 cell divisions, highlighting that this form of gene silencing is stable and heritable. Using genome-wide CRISPRoff screens, we show that this approach can durably and specifically silence the large majority of protein-coding genes and has a wide targeting window across gene promoters. Surprisingly, canonical CpG island annotations are not necessary for stable gene silencing by CRISPRoff. Lastly, we demonstrate that CRISPRoff can be used for silencing enhancers and engineering gene silencing programs in human stem cells that persist through differentiation to neurons. More generally, this system allows us to broadly explore the biological rules underlying epigenetic silencing and provides a robust tool for controlling gene expression, targeting enhancers, and exploring the principles of epigenetic inheritance.

[0280] Results

[0281] Rational design of a single fusion epigenome memory editor

[0282] We designed a CRISPR-based programmable epigenome editor protein, termed CRISPRoff-V1, composed of *ZNF10* KRAB, Dnmt3A (D3A), and DNMT3L (D3L) protein domains fused to catalytically inactive *S. pyogenes* dCas9 (**FIG. 1A**). To test whether a transient pulse of CRISPRoff epigenetic editing could silence gene expression durably, we transiently co-transfected HEK293T cells stably expressing a DNA methylation-sensitive *GAPDH-Snrpn* GFP reporter with either CRISPRoff-V1, dCas9-KRAB (CRISPRi), or dCas9-D3A-3L, along with sgRNAs targeting the *GAPDH-Snrpn* synthetic promoter (Liu et al., 2016; Stelzer et al., 2015)

(**FIG. 1B**). All three epigenetic editor proteins transiently silenced the GFP reporter (**FIG. 1C**). As expected for a transient transfection, expression of each epigenetic editor protein was lost over time, which for dCas9-KRAB and dCas9-D3A-3L resulted in restored expression of GFP. By contrast, for CRISPRoff-V1, gene silencing memory and CpG island (CGI) methylation was maintained long after CRISPRoff expression was lost (**FIG. 1C**).

[0283] Silencing by CRISPRoff-V1 appeared to be meta-stable as gene expression of the reporter gradually increased with time (**FIG. 1C**). To stabilize gene silencing memory, we encoded CRISPRoff with proteolysis-resistant linkers to minimize proteolysis that could result in untethered D3A-D3L and off-target DNA methylation as previously reported (Galonska et al., 2018; Hofacker et al., 2020). CRISPRoff variants programmed variably durable gene silencing (**FIGS. 8A-8B**). Second, we hypothesized that positioning D3A-3L at the N-terminus of dCas9 would allow Dnmt3A optimal access to CpG sites for DNA methylation (Zhang et al., 2018) (**FIG. 8C**). The CRISPRoff-V2 epigenetic editors we engineered each had similar gene silencing stability so we used CRISPRoff-V2.1 in all subsequent experiments (**FIGS. 8D-8F**).

[0284] Transient expression of CRISPRoff-V2 programmed a durable memory of gene silencing for at least 50 days post-transfection, with over 80% of transfected cells silencing the *Snrpn*-GFP reporter and over 90% silencing the endogenously GFP-tagged gene *HIST2H2BE* (*H2B*) (**FIGS. 1E-1F**, and **8G**). H2B silencing was accompanied by CGI methylation (**FIG. 1G**). Notably, starting at 10 days post-transfection, no CRISPR-off protein was detected (**FIG. 1E**). Transfection of CRISPRoff-V2 mRNA also silenced expression of the endogenously GFP-tagged gene *CLTA*, supporting that transient expression of CRISPRoff leads to effective gene silencing (**FIG. 8H**). These results demonstrate that CRISPRoff epigenetic memory does not depend on sustained transgene expression.

[0285] To further compare CRISPRoff-V1 and V2, we silenced three cell surface-localized proteins (ITGB1, CD81, CD151) that are not required for cell proliferation or survival. Transfection of CRISPRoff-V1 with one sgRNA silenced each target gene in a fraction of cells and a pool of three sgRNAs improved silencing of *ITGB1* and *CD81* (**FIG. 1H**). CRISPRoff-V2 improves silencing of each gene, with at least 80% silencing at 3 weeks post-transfection (**FIG. 1H**).

[0286] Durable and multiplexed silencing of endogenous genes

[0287] We demonstrated the efficacy of CRISPRoff-V2 in a variety of cell types, namely induced pluripotent stem cells (iPSCs), HeLa, U2OS, and K562 (as a doxycycline-inducible system) (**FIGS. 9A-9D**). We further show that CRISPRoff can be programmed by orthogonal

DNA binding proteins: dCas9 from *S. aureus* (dSauCas9) and dCas12a from *Lachnospiracea* (dLbCas12a) (**FIGS. 9E-9F**). Silencing with dLbCas12a was improved when three crRNAs (CRISPR RNAs) were encoded as a single transcript that can be processed by dLbCas12a into individual crRNAs, indicating a route to multiplexed gene silencing.

[0288] To explore multiplexed silencing of endogenous human genes with *S. pyogenes*-based CRISPRoff, we targeted *ITGB1*, *CD81*, and *CD151* in two, three, and four gene combinations (**FIGS. 11-1K** and **8I-8K**). At 30 days post-transfection, we observed robust multiplexed gene silencing of each gene combination (**FIG. 11**). We observed that cells that silenced one gene have a higher likelihood of silencing the other targeted genes (**FIGS. 11-1K, 8I, and 8K**). For example, when co-targeting *ITGB1* and *CD81*, cells that successfully silenced *ITGB1* had a 25-fold higher percentage of cells that also silenced *CD81* compared to cells that failed to silence *ITGB1* (**FIG. 11**).

[0289] To measure long term maintenance of epigenetic memory, we targeted the endogenous *CLTA* gene and followed *CLTA* expression in single cell clones (Leonetti et al., 2016a). Remarkably, at 15 months post-transfection or after ~450 cell divisions, 38 out of 39 clones maintained silencing of *CLTA* (**FIG. 1L**).

[0290] Epigenome editing is highly specific

[0291] To assess the specificity of CRISPRoff, we performed RNA-seq of cells 33 days post-transfection of CRISPRoff and sgRNAs targeting *ITGB1*, *CD81*, and *CD151* or a negative control sgRNA. Comparison of untransfected cells with cells transfected with CRISPRoff show minimal off-target gene knockdown (**FIG. 2A**). CRISPRoff targeting of *ITGB1*, *CD81*, and *CD151* were highly specific and showed near complete repression of the targeted gene (**FIGS. 2B-2D** and **10A**). RNA-seq analyses of three other cell lines with an endogenously GFP-tagged gene repressed by CRISPRoff (*RAB11A*, *CLTA*, and *H2B*) also showed robust and highly specific transcript knockdown (**FIGS. 10B-10D**). Analysis of neighboring genes within a 1 megabase window from the target gene showed no significant changes in gene expression (**FIGS. 10E-10F**). When analyzing the datasets from CRISPRoff targeting of *ITGB1*, *CD81*, and *CD151*, we observed 1-3 non-target transcripts with a log2 fold-change > 2 and adjusted p-value < 0.5 in each gene knockdown experiment, albeit at much lower magnitude compared to the targeted gene. Differential expression of non-targeted transcripts may be due to indirect effects associated with target gene knockdown or off-target CRISPRoff activity.

[0292] We assessed CRISPRoff DNA methylation specificity by whole genome bisulfite sequencing (WGBS) 30 days post-silencing of *CLTA* (**FIG. 10G**). We detected a single

dominant gain in DNA methylation at the *CLTA* promoter, in a 1.5 kb window across the *CLTA* promoter, highlighting the high specificity of CRISPRoff (**FIG. 2E**). We did not detect spreading of DNA methylation into the closest neighboring genes (**FIG. 2F**).

[0293] Consistent with previous analyses of DNA methylation in cells treated with DNMT-based epigenetic editors, we observed modestly higher global DNA methylation in CRISPRoff-transfected cells (<2%; **FIGS. 10H-10I**) (Galonska et al., 2018; O'Geen et al., 2019). However, global DNA methylation also varied between cell clones not exposed to CRISPRoff to a similar degree and the differences in local DNA methylation patterns at non-target DNA sites between cell clones was greater than any of the modest, non-specific changes in methylation seen following expression of CRISPRoff (**FIGS. S3I-S3J**). To examine whether possible CRISPRoff sgRNA dependent or independent off-target DNA methylation could alter gene expression, we inspected the top 10 most differentially methylated DNA regions. We did not detect any transcriptional differences for genes at or near the top 10 non-target differentially methylated regions (**FIGS. 10K-10L**). Thus, our WGBS data indicate that any off-target methylation differences are infrequent and unlikely to modulate gene expression or cellular phenotypes.

[0294] Lastly, we used ChIP-seq to profile changes in repressive H3K9me3 modifications after CRISPRoff targeting of the *HIST2H2BE* (*H2B*) gene. We detected a strong increase in H3K9me3 within a ~5 kb region across the *H2B* promoter at 5 days post CRISPRoff transfection that was maintained at 30 days, demonstrating the stable propagation of H3K9me3 and DNA methylation as discussed further below (**FIG. 2G**). Comparing *H2B*-targeting and non-targeting sgRNA conditions showed that the most significant gain of H3K9me3 occurred at the *H2B* locus and three neighboring genes: *HIST2H2AC*, *HIST2H2AB*, and *BOLA1* (**FIG. 2H**). We detected knockdown of *HIST2H2AC* expression whereas sequencing reads mapping to *HIST2H2AB* were not detected in our RNA-seq data. We did not detect transcriptional knockdown of *BOLA1*. These data, coupled with whole genome bisulfite sequencing data that showed confinement of CpG methylation, highlight the transcriptional and epigenomic specificity of CRISPRoff, while also documenting local epigenetic spreading and maintenance from the target site of establishment.

[0295] Genome-wide targeting of CRISPRoff

[0296] The simple design of CRISPRoff motivated us to perform pooled, genome-wide screens to determine its generalizability for silencing genes in the human genome. We designed a sgRNA library that targets over 20,000 protein-coding genes and includes about 1,000 non-targeting sgRNAs (Horlbeck et al., 2016a; Replogle et al., 2020). We constructed the sgRNA

library to encode two unique protospacers targeting the same gene per lentiviral vector, as our experiments show improvement in CRISPRoff activity when using multiple sgRNAs targeting the same gene (Replogle et al., 2020) (**FIG. 4A**).

[0297] We performed growth-based pooled screens since gene essentiality datasets are available from previous functional genomics efforts, allowing us to compare the performance of CRISPRoff to other genome-wide dropout screens. To perform a CRISPRoff pooled screen, we packaged the sgRNA library into lentiviral particles then transduced and selected HEK293T cells such that on average each cell expresses one sgRNA vector. We then transiently transfected this pool of cells with plasmids encoding CRISPRoff and sorted cells that expressed the CRISPRoff protein. We harvested a population of CRISPRoff-transfected cells as a time zero (T0) sample and continued to passage a population of CRISPRoff-transfected cells for at least 10 cell doublings (T10), followed by deep sequencing of genomic DNA at both time points to read out and quantify the sgRNA sequences as a proxy for cell count. We inferred that sgRNAs that were depleted in the T10 population relative to T0 are active, as these sgRNAs effectively silenced the expression of essential genes and drop out of the population (**FIG. 4B**). As a control, we performed in parallel an identical screen with a CRISPRoff variant encoding the Dnmt3A^{E765A} mutation, which is catalytically inactive and thus unable to maintain durable gene silencing and instead mirrors the transient silencing effect of CRISPRi (**FIG. 4C**). By comparing these two screens, we identified sgRNAs in the CRISPRoff screen that silenced gene expression in a manner that is DNA methylation dependent.

[0298] Analysis of the phenotype score (γ , with a more negative score indicating a stronger growth defect) for each gene showed that CRISPRoff expression reproducibly led to a more pronounced growth defect phenotypes compared to the CRISPRoff mutant (**FIGS. 4D-4E, 11A-11C**). A large set of genes showed drastic growth defects that were specific to CRISPRoff-mediated knockdown, highlighting the durable gene silencing effect of CRISPRoff. We also detected a subset of genes with comparable phenotypes between the two screens, likely due to their essentiality upon transient knockdown by the CRISPRoff mutant (**FIG. 4D**). We evaluated the specificity of silencing across the screens by analyzing the phenotype scores of control sgRNAs. Almost all negative control sgRNAs or sgRNAs targeting unexpressed genes (olfactory or Y chromosome genes) had little to no measured phenotype (1% with $\gamma < -0.2$) (**FIG. 4D**).

[0299] To evaluate the generality of CRISPRoff for gene silencing, we assessed the phenotype scores of genes that we expect to have growth phenotypes upon knockdown. Analysis of genes associated with DNA replication and the ribosome, which are predicted to be highly essential for

cell proliferation, were among the most severe growth phenotypes (**FIGS. 11D-11E**). We then analyzed the phenotypes for common essential genes that are required for cell proliferation or survival in most cancer cell lines (Meyers et al., 2017). The growth defects of these genes were far more pronounced in CRISPRoff (median $\gamma=-0.2$) compared to the CRISPRoff mutant (median $\gamma=-0.05$), whereas the majority of nonessential genes did not produce growth phenotypes (**FIG. 4E**). The CRISPRoff mutant resulted in weak phenotype scores due to the lack of DNA methylation-dependent durable gene silencing. Collectively, the CRISPRoff screen resulted in high positive rate of calling essential genes with low false positive gene hits, suggesting that CRISPRoff has the ability to silence the majority of genes in the human genome (**FIG. 4F**).

[0300] Programmable epigenome editors can initiate epigenetic marks that spread from the site of establishment (Hathaway et al., 2012; Stepper et al., 2017). We wondered whether some CRISPRoff gene hits were due to a “neighboring gene effect” caused by DNA methylation spreading from a nearby essential gene. We catalogued gene “hits” (defined by genes with phenotype scores of -0.2 or lower) and determined their linear distance on the genome from the nearest gene hit. Although a subset of gene hits were within a 1 kb window distance, the majority of CRISPRoff hits were over 10 kb from the nearest gene hit (**FIG. 4G**). Since CpG islands are largely restricted within a 1-2 kb window (Deaton and Bird, 2011), we postulate that the majority of our observed gene hits are specific to the targeted gene promoter, consistent with the specificity demonstrated in our CRISPRoff RNA-seq, WGBS, and ChIP-seq experiments.

[0301] CRISPRoff silencing of genes that lack CGI annotations

[0302] It is estimated that about 30% of human genes are not associated with a promoter CpG island (CGI) (Deaton and Bird, 2011). Given the observed generality of CRISPRoff for gene silencing, we wondered whether genes that lack CGI annotations can be silenced durably by CRISPRoff. Surprisingly, we found over 300 genes without CGI annotations with growth defects upon knockdown ($\gamma \leq -0.1$) by CRISPRoff, with 160 producing growth phenotypes $\gamma \leq -0.2$ (**FIG. 5A**). The majority of these genes had weak to no phenotype in the CRISPRoff mutant screen, indicating that their knockdown is DNA methylation dependent despite the absence of an annotated CGI.

[0303] To validate our observation that CRISPRoff can silence genes without annotated CGIs, we endogenously tagged five genes with no annotated CGI – *CALDI*, *DYNC2LI1*, *LAMP2*, *MYL6*, and *VPS25* – in HEK293T with mNeonGreen (mNG) and assessed durable silencing by CRISPRoff (**FIG. 5B**). At 14 days post-transfection of CRISPRoff, we detected a large

percentage of cells that turned off *DYNC2LI1*, *LAMP2*, *MYL6*, and *VPS25* (**FIG. 5C**). We did not detect stable silencing of *CALD1* at 14 days, potentially due to its promoter being almost completely devoid of CpG dinucleotides or non-optimal sgRNAs used in the experiment (**FIG. 5C**). Transfection of the CRISPRoff mutant did not silence gene expression durably. Treatment of *DYNC2LI1* and *LAMP2*-off cells with TETv4 led to reactivation of gene expression in about 70% of cells (**FIG. 5D**).

[0304] We isolated *LAMP2*, *DYNC2LI1*, and *MYL6* silenced cells and profiled the DNA methylation status of the gene promoter by bisulfite sequencing. Cytosines within a CG context were highly methylated in silenced cells (**FIG. 5E**). We passaged *DYNC2LI1*, *LAMP2*, and *MYL6*-silenced cells for 30 days and observed stable silencing of *DYNC2LI1* and *LAMP2* (**FIGS. 5F-5I**). We also followed 33 single cell clones with *DYNC2LI1* silenced and all clones repressed the gene by 50 days post-transfection with CRISPRoff (**FIG. 5I**). Although *MYL6* underwent silencing associated with DNA methylation at an early time point, gene expression reactivated to near pre-CRISPRoff level by day 30.

[0305] Lastly, to probe the extent of DNA methylation across a non-CGI annotated gene, we performed WGBS of cells with *DYNC2LI1* silenced after 30 days. We detected a single dominant gain in DNA methylation at the *DYNC2LI1* promoter (**FIG. 5J**) consisting of a about 1.2 kb region of the promoter (**FIG. 5K**). Together, these data establish that epigenetic editing using CRISPRoff is not limited to genes with canonical CGI annotations and can be targeted to most genes encoded in the human genome. Moreover, based on these findings, we propose that the theoretical framework of CGI gene annotation does not always predict the presence of functional CpG sites, bolstering the power of CRISPRoff and CRISPRon for functional testing of CpG methylation in modulating gene expression.

[0306] Targeting rules for CRISPRoff platform

[0307] We next explored the targeting landscape of CRISPRoff within gene promoters. Previously, we and others used sgRNA tiling screens and machine learning approaches to show that active sgRNAs for CRISPRi are localized in a narrow window at gene promoters, particularly at a nucleosome-depleted region immediately downstream of the transcription start site (TSS) (Gilbert et al., 2014). Despite successfully using these CRISPRi rules to design the genome-wide CRISPRoff essentiality screens, it remained untested whether the location of effective guides for CRISPRoff was similarly limited to this narrow window.

[0308] To empirically determine the targeting window of CRISPRoff, we designed a pooled sgRNA promoter tiling library against a subset of genes that are essential for cell growth based

on previous CRISPRi screens in K562 cells and our genome-scale CRISPRoff screen (**FIG. 12A**). The library tiles PAM-containing sequences +/- 1 kb from the TSS of 520 genes (425 with one annotated CGI, 56 with multiple CGIs, and 39 with no annotated CGI; defined by the presence of a CGI within 2.5 kb of the TSS), totaling ~116,000 unique sgRNAs (**FIGS. 6A-6B**). We performed the CRISPRoff screens in HEK293T cells using the same experimental workflow as the genome-wide screens, in parallel with the CRISPRoff D3A^{E765A} methyltransferase mutant. We also transduced the sgRNA tiling library into K562 cells stably expressing dCas9-KRAB and performed a CRISPRi screen to compare gene silencing mediated by dCas9-KRAB alone.

[0309] To evaluate the screens, we calculated the phenotype score (γ) for the three most active sgRNAs per gene and compared phenotypes across the three screens. We first focused on the 425 genes with one annotated CGI, as these were predicted to be canonical targets for CRISPRoff-mediated silencing. The phenotypes for the CRISPRoff screen (median $\gamma = -0.33$) were more pronounced compared to the CRISPRoff mutant screen phenotypes (median $\gamma = -0.15$), establishing that strong phenotypes observed in the CRISPRoff screen are DNA methylation-dependent (**FIG. 6C**).

[0310] To compare the optimal sgRNAs between CRISPRoff and CRISPRi, we normalized the phenotypes across the screens and generated an aggregate plot of sgRNA activities relative to the TSS (**FIG. 6D**). Consistent with our previous work, highly active sgRNAs for CRISPRi were centered on a narrow window (about 75 bp) directly downstream of the TSS. Similarly, active sgRNAs for the CRISPRoff mutant mirrored CRISPRi, which we expected because the KRAB domain remains functional in this fusion protein despite the lack of DNA methylation activity. In contrast, the active CRISPRoff sgRNAs were broadly distributed across the TSS, notably within a 1 kb window centered on the TSS. Representative gene analysis of *DKC1*, *GPN2*, and *ZCCHC9* shows that even within a single promoter, active sgRNAs for CRISPRoff are distributed across -500 to +500 bp from the TSS—a greatly widened targeting window for silencing compared to CRISPRi (**FIGS. 6E-6G**). We also observed effective sgRNAs outside the CGI that were DNA methylation-dependent (**FIG. 6G**), indicating that functional CpGs are not necessarily confined to canonical CGIs as observed in our WGBS data. Our aggregate plot analysis of active sgRNAs targeting the 56 genes with multiple annotated CGIs also shows a broad targeting window, similar to genes with one annotated CGI, and centered at the TSS (**FIGS. 12B-12C**).

[0311] Analysis of the 39 genes without promoter CGIs showed many highly active sgRNAs of comparable phenotype strength to genes with annotated CGIs (**FIG. 6C**, colored red) and the phenotypes were strongly diminished in the CRISPRoff mutant screen. A representative gene

plot of *ORC5* shows that similarly to genes with annotated CGIs, active CRISPRoff sgRNAs are spread across -500 to +500 bp from the TSS (**FIG. 6H**). Moreover, we observed that CRISPRoff has a broadened targeting window despite the lack of an annotated CGI for these 39 genes (**FIG. 6I**). Our experiments demonstrate that the optimal window for CRISPRoff gene silencing is similarly broad for genes with and without annotated CGIs, likely due to low density CpG sites that are functional for methylation-dependent gene silencing as we demonstrated for *DYNC2L1*, *LAMP2*, *MYL6*, and *VPS25* (**FIG. 5C**).

[0312] We observed that active sgRNAs are not evenly distributed but instead appear in a periodic pattern within the -500 to +500 bp window, as shown for *DKC1*, *GPN2*, and *ZCCHC9* (**FIGS. 6E-6G**). Overlaying nucleosome occupancy with CRISPRoff sgRNA activity scores for all genes showed that the most active sgRNAs are located in nucleosome-depleted regions of gene promoters, as we and others have shown previously for Cas9 and dCas9-based tools (**FIGS. 6J-6K, 12D and 12E**) (Horlbeck et al., 2016b; Isaac et al., 2016).

[0313] To validate that the CRISPRoff targeting window is similar for genes that do not have a growth phenotype upon knockdown, we designed tiling sgRNA libraries spanning +/- 2.5 kb from the TSS to target four endogenous genes: *CLTA*, *H2B*, *RAB11A*, and *VIM*. For each custom sgRNA library screen, we utilized the corresponding HEK293T cell line that expresses the endogenously GFP-tagged gene (Leonetti et al., 2016b). Each cell line was transduced with the respective sgRNA library, transfected with CRISPRoff, and the cells were passaged for 4 weeks to ensure that gene silencing was durable (**FIGS. 12F**). We then sorted GFP positive and GFP negative cell populations for each screen and processed the samples as described above. We calculated sgRNA efficacy by identifying sgRNAs in the gene-off (GFP-) population compared to the gene-on (GFP+) population.

[0314] Similar to the growth screens, active sgRNAs for *CLTA*, *H2B*, and *VIM* spanned a large window across the TSS (**FIGS. 6L and 12G-12I**). Active CRISPRoff sgRNAs for *CLTA* were within two distinct regions, with one region upstream of the TSS outside of the annotated CGI (**FIG. 12G**). Unexpectedly, sgRNAs targeting ~2 kb upstream of the *H2B* TSS were highly active (**FIG. 6L**). Similarly, for *VIM*, active sgRNAs spanned a 2 kb window +/- 1 kb from the TSS. By contrast, active sgRNAs for *RAB11A* were constricted to a narrow window at the TSS. Overlaying nucleosome occupancy with sgRNA activity showed that the *RAB11A* promoter is nucleosome-dense (**FIG. 12H**). From these data, we interpret that CRISPRoff accessibility is restricted by nucleosomes; however, once bound, CRISPRoff can silence gene expression even when distal to the TSS.

[0315] Durable gene silencing is dependent on H3K9me3 and DNA methylation maintenance

[0316] To explore the mechanism underlying CRISPRoff-mediated heritable memory, we made use of the wide targeting window across the *H2B* promoter to investigate the establishment, spreading, and maintenance of H3K9me3 histone modifications and CpG methylation marks. We targeted CRISPRoff to the TSS (sgRNA-A) and to a distal site ~2 kb upstream of the TSS (sgRNA-B) (**FIG. 13A**). At 30 days post CRISPRoff transfection, 89% of cells maintained H2B silencing when sgRNA-A was delivered compared to 76% with sgRNA-B (**Figure 13B**). Using ChIP-seq, we showed that both sgRNAs induced deposition at day 5 and maintenance at day 30 of H3K9me3 across the locus despite the ~2 kb distance between the sgRNA binding sites (**FIG. 13C**). The acquired H3K9me3 modifications in CRISPRoff-treated cells overlapped with the unmethylated CpG region in untreated parental cells (bottom track in **FIG. 13C**). In contrast, deposition and maintenance of H3K9me3 was far weaker with CRISPRoff bearing the D3A mutation, consistent with the failure to sustain gene repression with the mutant (**FIG. 13D**).

[0317] We next profiled CRISPRoff-mediated CpG methylation at the targeted distal and TSS regions. At day 5, we detected establishment and spreading of CpG methylation from the site of initiation to an about 2 kb site from the sgRNA binding site (labeled site 1 and site 3, **FIGS. 13E-13F**). By day 30, we detected stable maintenance of DNA methylation at both sgRNA binding sites (**FIGS. 13G-13H**). We also detected a high degree of CpG methylation between the sgRNA binding sites, suggesting a linear movement of spreading from the site of initiation (site 2, **FIG. 13I**). These data, together with our WGBS data, highlight the orchestration of histone and DNA methylation deposition, spreading, and maintenance and indicate that there are underlying regulatory principles that likely depend on the genomic context.

[0318] CRISPRoff gene silencing in iPSCs and iPSC-derived neurons

[0319] Due to the utility of stem cells for studying the development and function of specific cell types, we employed CRISPRoff in induced pluripotent stem cells. We transfected iPSCs with CRISPRoff and sgRNAs targeting *CD81* or a non-targeting control and found that at 30 days post-transfection, many iPSCs had stably silenced CD81 (**FIGD. 7A-7B**). Thus, CRISPRoff-encoded memory of silencing is stably maintained in stem cells.

[0320] We isolated CD81-off iPSCs and differentiated the cells into neurons, as previously described (Tian et al., 2019) (**FIG. 7A**). We then measured CD81 protein levels at the neural precursor cell stage (day 0 of differentiation) and after cells had differentiated into neurons (8 days post-differentiation). We observed that CD81 remained silenced during and after neuronal

differentiation in 90% of cells (**FIGS. 7C-7D**). A similar fraction of undifferentiated iPSCs maintained CD81 silencing during the same time course, suggesting that the reactivation of CD81 in about 10% of cells was not due to the differentiation process. We harvested genomic DNA from CD81-off neurons and detected heavily methylated promoter CpG dinucleotides compared to neurons treated with CRISPRoff and a non-targeting sgRNA (**FIG. 7E**).

[0321] We next applied CRISPRoff-mediated editing of iPSC-derived neurons to silence *MAPT*, a gene that encodes the Tau protein and is implicated in various neurological diseases (Iqbal et al., 2016). *MAPT* is transcriptionally repressed in iPSCs by H3K27me3 rather than by DNA methylation and H3K9me3 and its expression increases substantially during neuronal differentiation (Guenther et al., 2010). We hypothesized that CRISPRoff could write an epigenetic memory of silencing at the *MAPT* locus that would persist through neuronal differentiation to silence *MAPT* in neurons. We transiently transfected CRISPRoff into iPSCs along with sgRNAs targeting *MAPT* or a non-targeting control (**FIG. 7F**). At day 10 of the differentiation protocol, we measured Tau protein levels and found ~30% of cells with reduced Tau expression compared to a non-targeting control (**FIGS. 7G-7H**). Together, these data support CRISPRoff-mediated epigenome editing as an applicable technology for rewriting gene expression programs in iPSC-derived cells, especially for modulating gene expression in cells where delivery of gene editing platforms remains a challenge.

[0322] CRISPRoff targeting of enhancer elements

[0323] Finally, we explored the potential utility of CRISPRoff for silencing promoter-distal elements by targeting enhancers that control the expression of the *PVT1* long noncoding RNA (Cho et al., 2018; Fulco et al., 2016). We transiently expressed CRISPRoff in the MDA-MB-231 breast cancer cell line with sgRNAs targeting either the *PVT1* promoter or at four previously identified enhancer elements downstream of the *PVT1* promoter: E1 (+15.5 kb), E2 (+60 kb), E3 (+105 kb), and E4 (+113 kb) (**FIG. 7I**). We detected a significant reduction of *PVT1* transcript levels with sgRNAs targeting E1, E3, and E4 (~40-60%) compared to 80% knockdown with promoter-targeting sgRNAs (**FIG. 7J**, left). In contrast, parallel experiments using the CRISPRoff-Dnmt3A^{E765A} mutant resulted in less robust knockdown (**FIG. 7J**, right). These results highlight the potential use of CRISPRoff for mapping and dissecting the functions of enhancer elements and noncoding regulatory elements in the human genome.

[0324] Discussion

[0325] Here, we present CRISPRoff and CRISPRon, two technologies for programmably writing and erasing epigenetic memories to control gene expression programs. Transient

expression of CRISPRoff writes a robust, specific, and multiplexable gene silencing program that is memorized by cells through cell division and differentiation, which can be rapidly reversed by CRISPRon. We show that CRISPRoff can specifically and robustly silence the large majority of human genes. Our experiments demonstrate CRISPRoff can perturb enhancers, opening the potential to target genome elements that control tissue-specific gene expression (Fulco et al., 2016; Tarjan et al., 2019).

[0326] Our finding that targeted DNA methylation outside of annotated CGIs can lead to robust memorized gene silencing extends the canonical model of methylation-based gene silencing, which posits that a high density of CpG methylation is a requirement for stable propagation of silencing (Boyes and Bird, 1992). Although CRISPRoff-mediated writing of DNA methylation and histone modification are artificially programmed, our results motivate the need to define functional DNA methylation and dissect regulatory DNA elements and other host factors that mediate initiation, spreading, and maintenance of histone and DNA methylation marks. For example, targeting CRISPRoff 2 kb upstream of the *H2B* TSS leads to acquisition and maintenance of H3K9me3 and DNA methylation marks at the same genomic positions as targeting CRISPRoff directly proximal to the TSS, pointing to the existence of preexisting boundaries that restrict epigenetic spreading. By allowing the initiation of silencing at a defined time and genomic location, CRISPRoff provides a unique tool for addressing these and other fundamental questions regarding the mechanism and biological role of heritable gene silencing in mammalian cells (Audergon et al., 2015; Iglesias et al., 2018; Ragunathan et al., 2015; Yu et al., 2018).

[0327] CRISPRoff provides a valuable complement to existing CRISPRi and CRISPR nuclease approaches (Doench, 2018; Hanna and Doench, 2020; Shalem et al., 2015). CRISPRoff gene silencing can lead to effectively complete nulls without inducing a DNA damage response facilitating multigene targeting screens or therapeutic cell engineering (Ihry et al., 2018). The ability to target CRISPRoff to a large window upstream of the TSS allows access to promoter SNPs that can be utilized for allele-specific targeting of disease-associated mutations. This will broadly enable approaches to silence dominant negative alleles. Similarly, silencing of long noncoding RNAs and regulatory RNAs provides a new avenue for stable reprogramming of gene expression. Silencing of inhibitory elements such as antisense transcripts, will result in a heritable increase in expression of some genes, enabling therapeutic efforts to mitigate haploinsufficiency or imprinting disorders (Buiting et al., 2016). More broadly, heritable epigenetic silencing provides a general tool for rewiring human gene expression programs.

[0328] Example 2

[0329] Blocking *Flavivirus* infection by epigenome editing

[0330] Medical interventions that promote immunity to viruses, such as vaccines, are one of the most promising classes of antiviral therapies. However, vaccines can take years to develop and deploy. We reasoned that epigenome editing is an attractive platform for blocking viral infections due to the transient pulse of editor expression, durability and reversibility of silencing, and unlike genome editing, there is no permanent change in the underlying DNA sequence (FIG. 3A).

[0331] As an initial test to assess the efficacy of epigenome editing in blocking viral infections, we challenged CRISPRoff-treated HEK293T cells that have maintained silencing of CLTA for >1 year with the live attenuated yellow fever virus (YFV) vaccine 17D. YFV is a *Flavivirus* that enter host cells through clathrin-mediated endocytosis. We harvested cells two days after YFV infection and detected a 50% reduction in infection of CLTA-off cells compared to WT HEK293T (FIG. 3B). CLTA is one of two clathrin light chains that can form the functional clathrin triskelion and we reasoned that the infected cells are due to the virus utilizing CLTB light chain-containing clathrin triskelion molecules for entry.

[0332] We next applied epigenome editing to block Dengue Virus (DENV-2) infection by turning off SPCS1 and STT3A, which are key factors in DENV pathogenesis. We transfected CRISPRoff into HEK293T and infected the cells with DENV-2 at 14 days post transfection. Infection of DENV-2 is reduced significantly in SPCS1 and STT3A targeted cells (FIG. 3C). Gene silencing of SPCS1 and STT3A was measured at about 60% knockdown, mirroring the DENV-2 infection in edited cells (FIG. 3D). CLTA-silenced cells were infected at the same level as unedited WT cells, suggesting that DENV-2 entry is CLTA-independent.

[0333] Example 3

[0334] The MAPT knockdown experiments were performed by transfecting CRISPRoff with *MAPT*-targeting sgRNAs, as described below. To increase signal, iPSCs were co-transfected with *MAPT*-targeting sgRNAs along with *CD81*-targeting sgRNAs. As controls, iPSCs were transfected with CRISPRoff and non-targeting sgRNAs. One week post-transfection, the iPSCs were stained with CD81 antibody and cells with CD81 knocked down were FACS-sorted and passaged for 1 week prior to differentiation into neurons. The neuronal differentiations were performed as described below. At 10 days post-differentiation, neurons were harvested for qPCR analysis and antibody body staining for Tau protein and analyzed by flow cytometry. The quantification of MAPT/Tau knockdowns were normalized to WT untransfected cells. The results are shown in FIGS. 8A-8C.

[0335] Materials and Methods

[0336] Plasmid Design and Construction

[0337] The dCas9 and KRAB sequences were obtained from a previous CRISPRi construct (Gilbert et al., 2013). The D3A and D3L sequences, including the D3A-D3L fusion, originated from (Stepper et al., 2017) and were assembled with dCas9 and KRAB DNA sequences into a CAG-expression plasmid using NEBuilder[®] HiFi DNA Assembly (NEB). All CRISPRoff fusion proteins include BFP as either a direct fusion or with a P2A-cleavage sequence to measure transfection efficiency by flow cytometry. The dSaCas9 (D10A, N508A) sequence was PCR amplified from pX603 (Addgene #61594) and the dLbCas12a sequence was PCR amplified from (Tak et al., 2017). The GAPDH-Snrpn-GFP lentiviral reporter originated from Addgene #70148 (Liu et al., 2016; Stelzer et al., 2015).

[0338] The sgRNA plasmids were constructed by restriction cloning of protospacers downstream of a U6 promoter using BstXI and BlnI cut sites, as previously described. The sgRNA expression plasmids also express a T2A-mCherry marker to measure transfection efficiency. The sgRNA sequences used for CRISPRoff experiments are listed in Table 5. The sgRNA sequences were chosen based on our previous algorithm to predict active CRISPRi sgRNAs (Horlbeck et al., 2016a).

[0339] All mRNA constructs were synthesized using the mMESSAGE mMachin[™] T7 Ultra Transcription Kit (Thermo Fisher Scientific). The T7 promoter sequence (SEQ ID NO:111) was first cloned upstream of the CRISPRoff sequence. The T7-CRISPRoff sequence was PCR amplified and used as template for in vitro synthesis reactions. Following the manufacturer protocol for synthesis, the reactions were cleaned by chloroform extraction and isopropanol precipitation.

[0340] Cell culture, DNA transfections, and flow cytometry

[0341] All cell lines were cultured at 37 °C with 5% CO₂ tissue culture incubators. HEK293T (female), HeLa (female), and U2OS (female) cells were cultured in Dulbecco's modified eagle medium (DMEM) in 10% FBS (HyClone), 100 units/mL streptomycin, 100 µg/ml penicillin, and 2 mM glutamine. K562 (female) cells were maintained in RPMI-1640 with 25 mM HEPES and 2.0 g/L NaHCO₃ in 10% FBS, 2 mM glutamine, 100 units/mL streptomycin, and 100 mg/mL penicillin. WTC Gen1c iPSCs (male) were cultured in mTESR media (STEMCELL Technologies) under feeder-free conditions on growth factor-reduced Matrigel (BD Biosciences). Cells were passaged using Accutase (STEMCELL Technologies) and seeded on Matrigel coated plates with mTESR media supplemented with p16-Rho-associated coiled-coil

kinase (ROCK) inhibitor Y-27632 (10 μ M; Selleckchem).

[0342] Lentiviral particles were produced by transfecting standard packaging vectors into HEK293T using *TransIT*-LT1 Transfection Reagent (Mirus, MIR2306). Media was changed 24 hours post-transfection with complete DMEM supplemented with 15 mM HEPES. Viral supernatants were harvested 48-60 hours after transfection and filtered through a 0.45 μ m PVDF syringe filter. Lentiviral infections included polybrene (8 μ g/ml).

[0343] CRISPRoff transfections and analysis

[0344] Transient transfection experiments in HEK293T were performed in 24-well plates using *TransIT*-LT1 Transfection Reagent (Mirus) and Opti-MEM™ 1 Reduced Serum Medium (Thermo Fisher Scientific). Cells at 70-80% confluency were transfected with 300 ng of plasmid encoding CRISPRoff, dCas9-KRAB, or dCas9-D3A-3L and 150 ng of plasmids encoding sgRNAs. CRISPRoff experiments in HeLa and U2OS cells were performed by nucleofection of plasmids using the SE Cell Line 96-well Nucleofector Kit (Lonza) and a 96-well Shuttle™ Device (Lonza), per manufacturer protocol. Transfected cells were sorted 2 days after transfection using a BD FACSaria II or FACSaria Fusion and sorted cells were passaged every 2-3 days to measure durability of gene silencing. Experiments that compare the silencing activity of different CRISPRoff fusions (FIGS. 1E, 4C, 8B, and 8F) were performed in cells that stably express the targeting sgRNA to normalize sgRNA expression. To generate cell lines stably expressing sgRNAs, cells were transduced with lentiviral particles that express the sgRNAs and sorted for sgRNA-positive cells 2-3 days after transduction.

[0345] Quantification of ITGB1, CD81, and CD151 protein levels were measured by cell surface antibody staining of live cells. Cells were incubated with APC- or PE-labeled antibody (BioLegend) for ~30 min in the dark at RT, washed twice with PBS containing 10% FBS, and protein expression was measured on a BD LSR II flow cytometer.

[0346] All flow cytometry data were analyzed using FlowJo and the raw FACS plots presented in the figures are in log10 scale.

[0347] PVT1 enhancer targeting

[0348] Quantitative RT-PCR quantification of *PVT1* expression was done as described in (Cho et al., 2018). Briefly, MB-MDA-231 cells were transfected with CRISPR DNA together with a sgRNA vector using Neon (1400 volt, 10 ms, 4 pulse). Double positive cells were sorted after 2 days and continued to culture for 3 days. RNA were extracted with Zymo spin column and gene expression was quantified with SYBR qPCR mix (LightCycler 480) using 45 ng of RNA. The expression of *PVT1* were normalized to GAPDH gene in ddCt method. T-test was

used to calculate the statistical significance based on 3-5 biological replicates per condition. The primer sequence used are: *PVT1* forward (SEQ ID NO:112), *PVT1* reverse (SEQ ID NO:113), *GAPDH* forward (SEQ ID NO:114), *GAPDH* reverse (SEQ ID NO:115). The sgRNAs targeting *PVT1* enhancers 1-4, promoter, and lambda (controls) were from (Cho et al., 2018) and listed on Table 5.

[0349] RNA sequencing

[0350] HEK293T cells that have maintained stable silencing of target genes were harvested 33 days (*ITGB1*, *CD81*, and *CD151*) or 28 days (*CLTA*, *HIST2H2BE*, *RAB11A*, and *VIM*) post CRISPRoff transfection. Cells were dislodged from plates with PBS, centrifuged at $500 \times g$ for 5 min and washed again with PBS. Total RNA was extracted using Direct-zol RNA MiniPrep (Zymo R2051). Library preparations were carried out using TruSeq Stranded mRNA Library Preparation Kit (Illumina RS-111-2101), starting with 1000 ng total RNA. Final libraries were assessed using a 2100 Bioanalyzer (Agilent), quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and sequenced as single end 50 base pair reads on a HiSeq 4000 (Illumina). For processing the sequencing reads, linker sequences (SEQ ID NO:116) were removed using FASTX-clipper (FASTX-Toolkit). The reads were then aligned to the human genome (GRCh37) using the STAR (Spliced Transcripts Alignment to a Reference, version 2.5) aligner against the Gencode Gene V24lift37 transcriptome annotation. Read quantification was carried out with featureCounts (Liao et al., 2014). All downstream analyses were performed with Python (version 2.7) using a combination of Numpy (v1.12.1), Pandas (v0.17.1), and Scipy (v0.17.0) libraries. Knockdown efficiency was calculated by normalizing gene Transcripts per Million (TPM) for the experimental samples with the mean TPM of the control (non-targeting) samples. Differential expression analysis was performed using DESeq2 (Love et al., 2014). We note that non-target differentially expressed transcripts are lowly expressed genes.

[0351] Chromatin immunoprecipitation and analysis

[0352] At 30 days post transfection, 10×10^6 cells were crosslinked with 1% formaldehyde for 10 min at room temperature and quenched with 1.25 M glycine. Crosslinked cells were washed twice with cold PBS containing 1% Halt™ protease inhibitors (Thermo Fisher Scientific) and the cell pellets were flash frozen at -80°C until sample preparation. Cells were lysed in lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 1% Igepal, 1% protease inhibitors) for 10 min on ice. Nuclei were isolated after spinning the suspension at 2000 rpm at 4°C for 5 min. Nuclei were lysed at 4°C for 10 min in nuclei lysis buffer (50 mM Tris pH 8, 10 mM EDTA, 1% SDS, 1% protease inhibitors). Chromatin shearing was performed at 4°C using a Diagenode Bioruptor®

Pico sonication device in 1.5 ml Bioruptor[®] Pico Microtubes. The shearing program was optimized to obtain 200-700 bp fragments (30 seconds on, 30 seconds off for 10 cycles). The sonicated samples were centrifuged at 13,000 rpm at 4 °C for 10 min and the supernatant was collected and diluted 5-fold in IP dilution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Igepal, 0.25% deoxycholate, 1% protease inhibitors). A fraction of the input was saved and frozen (4% of total) prior to proceeding to immunoprecipitation.

Immunoprecipitations were performed overnight at 4 °C using 5 µg of anti-H3K9me3 antibody (abcam ab8898). The washing steps were performed with Pierce[™] Protein A/G magnetic beads (Thermo Fisher Scientific) using the following protocol: once with IP wash buffer 1 (20 mM Tris pH 8, 2 mM EDTA, 50 mM KCl, 1% Triton X-100, 0.1% SDS), twice with high salt buffer (20 mM Tris pH 8, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS), once with IP wash buffer 2 (10 mM Tris pH 8, 1 mM EDTA, 0.25 lithium chloride, 1% Igepal, 1% deoxycholate), and twice with TE buffer (10 mM Tris pH 8, 1 mM EDTA). Samples were eluted in elution buffer (50 mM sodium bicarbonate, 1% SDS) at 65 °C for 1 hour and reversed crosslinked initially in 300 mM NaCl and RNase A for 1 hour at 37 °C, followed by 63 °C overnight with Proteinase K (Thermo Fisher). The DNA samples were purified using a Zymo Clean & Concentrator kit and libraries were prepared using the NEBNext[®] Ultra[™] II DNA Library Prep kit (NEB).

[0353] Reads were aligned to the human genome (hg19) using bowtie v2.3.2 (Langmead and Salzberg, 2012). Alignments were processed using deepTools2 bamCoverage (Ramírez et al., 2016), normalizing reads to 1x average coverage. The resulting bigWig files were visualized on the Integrative Genomics Viewer (IGV). Peak analysis was performed using MACS (<https://github.com/macs3-project/MACS>). Downstream analysis and enrichment at promoter regions, which were defined as +/- 2 kb of each TSS (with the TSSs based on previously published annotations (Horlbeck et al., 2016) were performed using Python 3.6 with the deepTools2 package. Differential H3K9me3 enrichment was analyzed using DESeq2 (Love et al., 2014), treating distinct sgRNAs against the same TSS as replicate samples. We note that *BOLA1* contains two TSS annotations and our ChIP-seq data show enrichment for H3K9me3 near TSS1, the closest to the *H2B* promoter, and no enrichment for TSS2 located 15 kb away (FIGS. 2G-2H). We do not detect transcriptional knockdown of *BOLA1* in our RNA-seq data.

[0354] Western blotting

[0355] Western blots of CRISPRoff constructs were performed by harvesting HEK293T cells 2 days post-transfection of CRISPRoff constructs. Cells were washed with cold PBS and lysed with RIPA buffer (Thermo Fisher Scientific) supplemented with Halt[™] Protease Inhibitor

Cocktail (Thermo Fisher Scientific). After 30 min of lysis at 4 °C, the samples were centrifuged at $20,000 \times g$ for 20 min at 4 °C. The soluble fractions were collected and protein concentrations were quantified by Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). 40 µg of total protein was mixed and heated with SDS loading buffer, separated on Bolt™ 4-12% Bis-Tris Plus Gels (Thermo Fisher Scientific), and wet transferred into PVDF membrane in buffer containing $1 \times$ MOPS and 10% methanol. Membranes were blocked with Odyssey® Blocking Buffer (LI-COR), incubated with antibodies against *S. pyogenes* Cas9 (Active Motif 61577) and calnexin (Abcam ab22595) at 4 °C overnight. Membranes were washed at least 3 times with blocking buffer before incubation with IRDye® secondary antibodies against Cas9 and calnexin. Blots were imaged using Odyssey® CLx (LI-COR).

[0356] Bisulfite sequencing PCR

[0357] Genomic DNA was extracted from $\sim 1-2 \times 10^6$ cells according to manufacturer's instructions using the PureLink Genomic DNA Mini Kit (Invitrogen). For each condition, 1 µg genomic DNA underwent bisulfite conversion and cleanup according to manufacturer's instructions using the EpiTect Bisulfite kit (Qiagen). Purified bisulfite-converted DNA was amplified using EpiMark Hot Start Taq (NEB). Amplicons were gel purified using a Gel DNA Recovery Kit (Zymo) and PCR amplified again using EpiMark Hot Start Taq. Amplicons were cloned into the pCR2.1 TOPO vector according to manufacturer's instructions using the TOPO TA Cloning Kit (Invitrogen). Cloning products were transformed into Stellar *E. coli* cells (Takara) and plated on carbenicillin plates with X-gal for blue-white screening. Colonies were picked per condition and sequenced by Sanger sequencing. Primer sequences for bisulfite-PCR amplification are listed in Table 6. The primer sequences for amplifying the GAPDH-Snrpn fragment was obtained from (Liu et al., 2016).

[0358] Table 6

Name	Sequence (5' to 3')	SEQ ID NO:
CLTA, forward primer	TATTTGTTGATTGGGTAGTTTTGAATT	SEQ ID NO:184
CLTA, reverse primer	ACTCCCTAACCTCCTAATTCAACAAAAA	SEQ ID NO:185
LAMP2, forward primer	ATAGGAAGGGTTGTGAATTA AAAAGTTAGG	SEQ ID NO:186
LAMP2, reverse primer	ACA ACTCACCCAAA ACTAAACAAACCAA	SEQ ID NO:187
DYNC2LI1, forward primer	GAGATGTAAAAGGATTCCTGAAAATTATATT	SEQ ID NO:188
DYNC2LI1, reverse primer	CTCTCTCCAAAATACTCTCAATAACCTTTCC	SEQ ID NO:189

MYL6, forward primer, fragment 1	ATTTTAAGCGTTTGAGTGTTCAGGTAGGG	SEQ ID NO:190
MYL6, reverse primer, fragment 1	TAAAACAATAAATAACCTCCTAATAACTAAAA	SEQ ID NO:191
MYL6, forward primer, fragment 2	TTTGATTTTTTTTGCTGTTGGGAGGTGTAGAT	SEQ ID NO:192
MYL6, reverse primer, fragment 2	CAACTCTCCGAAACCTTTTCCTACAATAAT	SEQ ID NO:193
CD81, forward primer	AAGTTGTGGGGtAttTGTGGGtTtAGGAG	SEQ ID NO:194
CD81, reverse primer	CCCaCaACCACCaCACCCATCACCACCACA	SEQ ID NO:195
MAPT, forward primer	AtAAAGAtTtAAtTAtAGGAGGTGGAGAAAG	SEQ ID NO:196
MAPT, reverse primer	aaAaCCTTCTCCTCCaaCCACTAaT	SEQ ID NO:197
H2B, site 1, forward primer	tAGTGGtATGGGGGAGtAGGATAAGAAAAG	SEQ ID NO:198
H2B, site 1, reverse primer	aATCCCCTTaAaCTCTAAACTaCCAAaTC	SEQ ID NO:199
H2B, site 2, forward primer	TCCTTGTAAGAAGtTGTAAtTgTAATTTTAGt	SEQ ID NO:200
H2B, site 2, reverse primer	TaaTaACACCCCaAaTAACTTaTTaAaCTCTT	SEQ ID NO:201
H2B, site 3, forward primer	AttAAGAGGAAAGtTTTATAAGtTTTtTAGGGG	SEQ ID NO:202
H2B, site 3, reverse primer	CCTaATTCTTTTATAACCACCTTATaCAAATTAAa aCTC	SEQ ID NO:203

[0359] Whole genome bisulfite sequencing and analysis

[0360] We generated whole genome bisulfite sequencing (WGBS) libraries for 12 samples, corresponding to WT (untreated), NT (non-targeting), and T (targeting) for CLTA and DYNC2LI1 experiments and profiled in two replicates. After DNA extraction and RNase A treatment using the PureLink Genomic DNA Mini Kit (Invitrogen), 1 µg of DNA was diluted to 7.7 ng/µL in 130 µL and sheared to 450-550 bp length using a Covaris E220 evolution with intensifier for 50s at 140V, 7°C, 10% amplitude, 200 cycles. The sonicated DNA was recovered and concentrated using Ampure XP beads and sizes of the sheared DNA were checked on an Agilent TapeStation device with a D1000 HS DNA ScreenTape. Next, the sheared DNA was bisulfite converted using the EZ DNA Lightning kit (Zymo, Cat. No. D5030) according to the manufacturer's instructions and a desulphonation step of 16 minutes. Then, 500 ng of sheared and converted DNA was subjected to library preparation using a Swift Accel®-NGS Methyl-Seq DNA Library Kit (Cat. No. 30024) and Methyl-Seq Unique Dual Indexing Primers (Cat. No. 39096). The prepared libraries were quantified using a KAPA Library Quantification kit (Roche,

Cat. No. KK4873) and sequenced using paired-end 150bp reads (300 cycles) on an Illumina NovaSeq6000 instrument with an S4 flow cell and a 35% spike-in from another non-WGBS library to diversify the sample pools. We obtained on average 707M paired-end reads (range 629-835M) across all 12 libraries.

[0361] Prior to alignment, sequencing reads were trimmed using fastp (version 0.21.0, (Chen et al., 2018)) and the following parameters: --adapter_sequence=(SEQ ID NO:117) --adapter_sequence_r2=(SEQ ID NO:118) --trim_front1=0 --trim_tail1=20 --trim_front2=20 --trim_tail2=0. Trimming is required to remove bases that are added during the Adaptase reaction that could affect alignment and DNA methylation calling. Processed reads were aligned to the hg38 reference genome using methylCtools (version 1.0.0, <https://github.com/hovestadt/methylCtools>, (Hovestadt et al., 2014)) and bwa mem (version 0.7.17, arXiv: 1303.3997v1) using default parameters. Over 98% of reads were aligned as proper pairs across samples. After marking of PCR duplicates using sambamba (version 0.8.0, (Tarasov et al., 2015)), genome-wide CpG methylation values were called using methylCtools using the --trimPE parameter. Average CpG coverage was ~25-fold across samples. Bisulfite conversion efficiency was estimated to be greater than 99.5% based on non-CpG methylation.

[0362] Downstream analyses were performed in R (version 4.0.2, <https://www.r-project.org/>) using the bsseq (version 1.24.4, (Hansen et al., 2012)) and DSS (version 2.36.0, (Park and Wu, 2016)) packages. Specifically, we applied the DMLtest function to first call differentially methylated loci (using 500bp smoothing windows) between treatments, and then the callDMR function to define differentially methylated regions. Results were visualized by plotting log10-transformed p-values associated with individual loci (positive values for loci that gained methylation, negative values for loci that lost methylation). Loci were colored by their difference in beta-values (-1: blue, 0: white, +1: red). Close-ups of genomic regions were generated by visualizing beta-values of individual loci in IGV (Robinson et al., 2011). Data was displayed as bar charts (min/0: blue, mid/0.5, max/1: red).

[0363] Genome-wide CRISPRoff screen and analysis

[0364] For genome-wide CRISPRoff screens, we constructed a compact library to maximize on-target knockdown while minimizing overall library size. We targeted each gene in the human genome with two unique sgRNAs expressed from tandem U6 expression cassettes in a single vector. To select the optimal sgRNAs targeting each gene, we relied on our previously published hCRISPRi v2.1 library (Horlbeck et al., 2016a). A three tiered approach was used to balance empirical data with computational predictions and select the most active sgRNA pair for each

gene. First, for strong essential genes ($p\text{-value} < 0.001$ and $\gamma < -0.2$ in hCRISPRi v2 growth screen), sgRNAs were ranked by growth. Next, for genes that were identified as a significant hit in previous CRISPRi screens, sgRNAs were ranked by the sum of Z-scored phenotypes across screens. Finally, for all other genes, sgRNAs were ranked by the regression scores in hCRISPRi v2.1. Using this ranking scheme, we designed a genome-wide library consisting of only 21378 elements (20360 targeting elements plus 1018 non-targeting controls).

[0365] To clone our libraries, we began by generating a modified single sgRNA lentiviral expression vector, pJR104, from the parental pJR85 (Addgene 140095) by: (i) replacing the BFP fluorescent marker with a BsmBI-negative GFP, (ii) replacing the sgRNA constant region with an unmodified constant region (i.e. without a Perturb-seq capture sequence), and (iii) incorporating a UCOE element upstream of the EF1 α promoter to prevent silencing. Dual-sgRNA oligos were synthesized as an oligonucleotide pool (Twist Biosciences) with the structure: 5'-PCR adapter-SEQ ID NO:126-protospacer A-SEQ ID NO:119-protospacer B-SEQ ID NO:120-PCR adapter-3'. Oligo pools were PCR-amplified, digested with BstXI/BlpI, gel extracted, ligated into the sgRNA lentiviral vector pJR104, and transformed to generate an intermediate library as previously described (Replogle et al., 2020). An insert, pJR98, consisting of a sgRNA constant region variant 3 (Adamson et al., 2016) and a hU6 promoter was synthesized (IDT), BsmBI-digested, and ligated into the BsmBI-digested intermediate library. The final dual-guide library was then transformed for amplification and sequenced by next-generation sequencing to ensure library representation and uniformity.

[0366] Pooled tiling sgRNA screens in HEK293T cells were performed by first transducing cells with lentiviral particles encoding the sgRNA library. The infection efficiency was measured 2 days post-infection by flow cytometry, aiming for 20-30% sgRNA-positive cells. The screens were performed with two technical replicates and each sgRNA was represented by at least 1000 cells throughout the duration of the screens. Two days post transduction, cells were treated with puromycin until the cell population was 90% sgRNA positive, as marked by mCherry encoded in the lentiviral vector. For transient transfection of CRISPRoff, $\sim 8 \times 10^6$ cells were first seeded on 15 cm² plates. About 20-24 hr later (70-80% confluency), each 15 cm² plate of cells were transfected with 20 μ g of plasmids encoding CRISPRoff or CRISPRoff-Dnmt3A^{E765A} catalytic mutant. Two days post-transfection, cells were sorted for CRISPRoff expression (BFP) and plated on 15 cm² plates. Four days post-sorting, cells were trypsinized and an aliquot of cells ($\sim 110 \times 10^6$) was harvested as an initial time point T(0) and the rest of the cell population was passaged for at least 10 more cell doublings. Cells were then collected as a final time point (T10).

[0367] DNA libraries of T(0) and T(10) were prepared for deep sequencing essentially as previously described (Jost et al., 2020). Briefly, genomic DNA was isolated using a NucleoSpin Blood XL kit (Macherey–Nagel). Then, isolated gDNA was directly amplified by 23 cycles of PCR using NEBNext Ultra II Q5 PCR MasterMix (NEB), appending Illumina adaptors and unique sample indices (oJR234 forward primer: SEQ ID NO:121; index primers SEQ ID NO:122). Sequencing was performed on a NovaSeq 6000 (Illumina) using a 19 bp read 1, 19 bp read 2, and 5 bp index read 1 with custom sequencing primers oJR326 (custom read 1, SEQ ID NO:123), oJR328 (custom read 2, SEQ ID NO:124), and oJR327 (custom index read 1, SEQ ID NO:125).

[0368] Sequencing counts from CRISPR screens were processed to calculate gene phenotypes using a custom Python script, similar to as previously described (Horlbeck et al., 2016a), except now two protospacer sequences are matched instead of just one. In this case, gene phenotype is the same as sgRNA phenotype, and is defined by $\log_2$ sgRNA enrichment / cell doublings. All additional CRISPR screen data analyses and plotting were performed in Python 3.6 using a combination of Numpy (v1.16.2), Pandas (v0.23.4), Scipy (v1.4.1), and sklearn (v0.22.2). The DepMap essential and nonessential genes were downloaded from DepMap Public 20Q2 at <https://depmap.org/portal/download/> (Blomen et al., 2015; Hart et al., 2014). Gene set enrichment analysis (GSEA) was performed using GSEAPY (v0.9.19) in Python using the 2019 Human KEGG Pathway database.

[0369] Tiling screen library design, experimental specifications, and analysis

[0370] For the growth-based screen, a tiling sgRNA library targeting essential genes was designed based on our previously published genome-wide CRISPRi screen in K562s. For genes with no canonical CGIs (as defined by no CGIs within 2.5 kb of TSS, with the TSSs based on previously published annotations (Horlbeck et al., 2016a) and CGI annotations from the UCSC Genome Browser), all genes with an average growth phenotype score less than -0.2 were picked. For genes with one or multiple CGIs (also defined as within 2.5 kb of TSS), genes with an average growth phenotype score between -0.2 and -0.4 were selected. In total, 39 genes with no canonical CGIs, 425 genes with one annotated CGI, and 56 genes with multiple CGIs were chosen.

[0371] For each gene, all sequences +/-2.5 kb (or +/- 1 kb depending on the position and length of the CGI) of the TSS containing 19 bp followed by an NGG PAM were extracted as potential sgRNAs. All sequences were prepended with a 5' G to enable robust transcription from the U6 promoter, whether or not this base was present in the genomic sequence. The sgRNAs

were scored for off-target sites using weighted Bowtie, as previously described (Horlbeck et al., 2016a). Briefly, sgRNAs were scored by uniqueness in the genome, as determined by an empirically derived and experimentally verified scoring metric: PAM G1 = 40, PAM G2 = 19, PAM N = 0, the next 7 bases from the PAM = 28, the next 5 bases = 19, and the last 7 bases = 10. A mismatch score was then calculated by the sum of the mismatches with the scoring metric. This mismatch score was implemented using the Phred score threshold feature of Bowtie using the --nomaqround, -n 3, -l 15, -a, and --best flags. For the most stringent threshold, sgRNAs were required to have no more than 1 alignment (the sgRNA target site itself) in the genome with a mismatch score of 39. Control non-targeting sgRNAs were extracted from a previously tested list of control sgRNAs (Horlbeck et al., 2016a).

[0372] The tiling libraries for endogenously GFP-tagged CLTA, HIST2H2BE (H2B), RAB11A, and VIM were designed similarly, selecting for sgRNAs +/- 2.5 kb from the TSS and yielding ~600 sgRNAs per gene. The protospacer sequences for each GFP-tagged gene library are available in Table 5.

[0373] Oligonucleotide pools were designed with flanking PCR and restriction sites (BstXI and BpI), synthesized by Agilent Technologies, and cloned into the sgRNA expression vector pCRISPRia-v2 (Addgene #84832), as described previously (Horlbeck et al., 2016a). The expression vector contains a U6 promoter driving the sgRNA expression, as well as an EF1 α promoter driving puromycin T2A-mCherry.

[0374] The tiling screens in HEK293Ts were performed in a similar workflow as the genome-wide CRISPRoff screens. To perform the tiling screen in K562s, cells stably expressing dCas9-KRAB were first transduced with lentiviral particles of the tiling sgRNA library. Two days post transduction (20-30% infection), cells were treated with puromycin until the population consisted of 90% sgRNA-expressing cells. A T(0) time point was then collected and cells were continued to passage for 10 more cell doublings to obtain the T(10) time point.

[0375] Sequencing counts from CRISPR screens were processed using the Python-based ScreenProcessing pipeline (<https://github.com/mhorlbeck/ScreenProcessing>), as previously described (Horlbeck et al., 2016a) to calculate sgRNA phenotypes. sgRNA phenotype score is defined by $\log_2$ sgRNA enrichment / cell doublings. All additional CRISPR screen data analyses and plotting were performed in Python 2.7 using a combination of Numpy (v1.12.1), Pandas (v0.17.1), and Scipy (v0.17.0). K562 and GM12878 MNase-seq data was obtained from the ENCODE consortium as processed continuous signal data (BigWig file format; Michael Snyder lab, Stanford University). The average of the K562 and GM12878 MNase-seq data was used.

We note that the phenotypes in K562 CRISPRi (median $\gamma = -0.46$) are more pronounced compared to the HEK293T CRISPRoff screen (median $\gamma = -0.33$). However, as we have previously demonstrated, because the genes were chosen based on essentiality in K562s, this difference likely can be attributed to cell type variability between K562 and HEK293T.

[0376] GFP-tagged sgRNA tiling screen

[0377] The tiling sgRNA screens in HEK293T GFP-tagged cell lines were performed in a similar workflow as the growth-based screens described above. The previously published endogenously GFP-tagged cell lines (*CLTA*, *HIST2H2BE*, *RAB11A*, *VIM*) were further FACS sorted to yield >99% GFP-positive cells to minimize background GFP-negative cells. After generating cell lines that stably express the respective sgRNA library, plasmids expressing CRISPRoff were transfected. Two days later, the transfected cells were sorted and subsequently passaged for 4 weeks by trypsinization every 2-3 days. At the 4 weeks time point, each cell line had the following detectable GFP-silenced population: 21.8% *CLTA*, 22.7% *HIST2H2BE*, 3.05% *RAB11A*, and 24.7% *VIM*. The GFP-on and GFP-off populations were FACS sorted into separate bins, collecting $\sim 2 \times 10^6$ cells per population for each cell line. The log2 fold change in sgRNA abundance was quantified by the presence of each sgRNA in the GFP-off population compared to the total population. Analysis was performed using Python 2.7, similar to the other tiling screens described previously.

[0378] iPSC manipulation and neuronal differentiation

[0379] Transient transfections of iPSCs were performed in 6-well plates using *TransIT-LT1* Transfection Reagent (Mirus). First, a mixture of 0.5 ml of mTeSR1 and 2 μ l of 10 mM Y-27632 ROCK inhibitor were added to each well of a Matrigel coated 6-well plate. Then, a mixture of plasmids encoding dCas9-KRAB or CRISPRoff (1 μ g), 1 μ g of sgRNA plasmids, and 200 ng of plasmid encoding BCL-XL (Li et al., 2018) were added to 0.4 ml of Opti-MEMTM. *TransIT-LT1* (12 μ l) was added to the DNA-Opti-MEMTM mixture and added to each well of a 6-well plate. Cells at 70-80% confluence were lifted with Accutase, washed with DPBS, and counted using a Countess (Thermo Fisher AMQAX1000). About 1.5×10^6 cells were resuspended in 1 ml of mTeSR1 and added to each well containing the transfection mixtures. Transfected cells were sorted 3 days post-transfection on a BD FACS Fusion and plated in mTeSR media supplemented with 10 μ M Y-27632 ROCK inhibitor.

[0380] Neuronal differentiations were performed on passage number 46 iPSCs using doxycycline-inducible NGN2 (Tian et al., 2019). On day -3, cells at 70-80% confluency were lifted with Accutase and washed with DPBS. About 7.5×10^5 cells were resuspended in 2 ml of

N2 Pre-differentiation Media each well of a Matrigel coated 6-well plate. On day 0, cells were lifted with Accutase and washed with DPBS. About 5×10^5 cells were resuspended in 2 ml classic N2/B27 Differentiation Media and plated onto Poly-D-Lysine coated plates (Corning 354413). On day 3, the media in each well were aspirated and replaced with 2 ml of fresh N2/B27 Differentiation Media. On day 7, 1 ml of media was removed from each well and replaced with 1 ml of fresh N2/B27 Differentiation Media. N2 Pre-differentiation Media was made with 1X Knockout DMEM/F12 (Thermo Fisher 11320-033), 1X NEAA (Thermo Fisher 11140-050), 1X N2 Supplement (Thermo Fisher 17502-048), 10 ng/ml NT-3 (PreproTech 450-03), 10 ng/ml BDNF (PreproTech 450-02), 1 μ g/ml Mouse Laminin (Thermo Fisher 23017-015), 10 nM Y-27632 ROCK inhibitor, and 2 μ g/ml doxycycline (Sigma-Aldrich D3447-500MG). Classic N2/B27 Differentiation Media was made with 0.5X DMEM/F12 (Thermo Fisher 10888-033), 0.5X Neurobasal-A (10888-022), 1X NEAA, 0.5X GlutaMAX (Thermo Fisher 35050-061), 0.5X N2 Supplement, 0.5X B27-VA Supplement (Thermo Fisher 12587010), 10 ng/ml NT-3, 10 ng/ml BDNF, 1 μ g/ml Mouse Laminin, and 2 μ g/ml doxycycline.

[0381] We used iNeuron RNA-Seq (<https://kampmannlab.ucsf.edu/ineuron-rna-seq>) to support the activation of *MAPT* gene expression through neuronal differentiation of iPS cells.

[0382] Informal Sequence Listing

[0383] In the sequences listed herein, the skilled artisan will appreciate that a methionine (M) can be present on the N-terminus of protein in order to initiate translation. Thus, SEQ ID NOS: 1-15 can optionally further comprise a methionone on the N-terminus.

[0384] SEQ ID NO: 1 - CRISPRoff-V1 or V1 or P76 (described in WO 2019/204766): KRAB (bold; from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A (bold italics; residues 612-912; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGSMDDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLG
 NTDHRSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKF
 RGHFLIEGDLNPDNSDVKLFQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQ

LPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDXYDDDLNLLAQIGDQYADLF
 LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFDQOS
 KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGE
 LHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVV
 DKGASASQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
 KKAIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFL
 DNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
 LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
 TRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
 QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYH
 HAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEI GKATAKYFFYSNIMNF
 FKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS
 SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHR
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSTGLYETRIDL SQ
 LGGDSRADypdydpdyasGS~~pkkrkv~~EASGSGRASPGIPGSTRNHDQEFDPKVPVPAEKR
 KPIRVLSLFDGIATGLLVKDLGIQVDRIASEVCEDSITVGMVRHQGKIMYVGDVRSVT
 QKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDR
 PFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGMNRPLA
 STVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEME
 RVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSGNSNANSRGPS
 FSSGLVPLSLRGSHMGPMEIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLESGSGSG
GGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQ
YALPROESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV
WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSLDAPKVDLLVKNCLLPLREYFKYFS
QNSLPLSRAD~~pkkrkv~~GSGatnfsllkqagdveenpgpselikenmhmkllymegtvdnhhfktsegegkpyeggt
mrikvveggplpfadilatsflygsktfinhtqgipdfkqspegftwervttvedggvltatqdtlsldgclinyvkirgvnftsnpgv
mqqktlgweaftetlypadgglegrndmalklvggshlianikttyskpkaknlkmpgvvyvdyrlerikeannetyveqhevav
arycdlpsklghkln*

[0385] SEQ ID NO:2 (p90 (KRAB-dCas9-XTEN16-Dnmt3A-Dnmt3L-P2A-BFP): KRAB
 (bold, from Gilbert et al., Cell, 2013, 2014); Linkers (underlined), dCas9 (italics); HA tag

(lowercase), SV40 NLS (lowercase italics), XTEN16 (uppercase, 16 amino acid sequence), Dnmt3A (bold italics; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGS**MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLG**
*NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS***NEMAKVDDSSFFHR**
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKF
*RGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEPNINASGVD***AKAILSARLSKSRRLNLI****Q**
*LPGEKKNGLFGNLI***ALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF**
*LA***AKNLS***DAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS*
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYFPLKDNREKIEKILTRIPYVVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFL
DNEENEDILEDIVLTTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
*RDKQSGKTILDFLKSDFANRNF***QLIHDDSLTFKEDIQKAQVSGQGDSLHEHLANLAGSPA**
*IKKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGKELG***SQI**
LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
*TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT***KAERGGSEL****DKAGFIKR**
*QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS***KLVSDFRKDFQFYKVINNYH**
*HAHDAYLNAVVG***TALIKKYPKLESEFVYGDKVYDVRKMI****AKSEQ****EIGKATAKYFFYSNIMNF**
*FKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVL***SMPQVNIVKKTEVQTGGFSKE**
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERS
*SFEKNPIDFLEAKGYKEVKKDLI***KLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF**
*LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR***VILADANLDKVL****SAYNKHR**
*DKPIREQAENIIHLFTLTNLGAPAAF***KYFDTTIDRKRYTSTKEVLDATLIHQ****SITGLYETRIDL****SQ**
*LG***GDS****SRAD***ypypdy***aSGS***pkkkrv***SPGSGSETPGTSESATPES****NHDQEFDP****PKVYPPVPAE**
KRKPIRVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRS
VTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPK**EGD**
DRPFFWL**FENVVAMGVSDKRD****ISRFLESNPVMIDAKEV****SAAHRARYFWGNLPGMNRP**
LASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQ**GKDQHFPVFMNEKEDILWCTE**
MERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSGNSNANSR

GPSFSSGLVPLSLRGSHMGPM EIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLESGS
SGGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHR
ILQYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNA
MRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYF
KYFSQNSLPLSRAD*pkkkkrkv*GS*Gatnfsllkqagdveenpgpselikenmhmklmegtvdnhhfktsegegk*
pyegtqtmrikvveggplpfafdilatsflygsktfinhgtqgipdffkqsfpdgftwervttvedggvltatqdtlsldgclinyvknirgvn
ftsngpvmqkktlgweaftetlypadgglegrmdmalklvggshlianikttyskpkaknlkmpgvvyvdyrlrikeannetyve
*qhevavarycdlpsklghkln**

[0386] SEQ ID NO:3 (p91 (KRAB-dCas9-Dnmt3A-Dnmt3L-P2A-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP(lowercase underlined))

DAKSLTAWSR TLVTFKDV FVDF TREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGSGGGG*MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLG*
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVV
DKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYH
HAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE

SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERS
 SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHR
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGDSRADypypdydyaSGSpkkkrkvEASGSGRASPGIPGSTRNHDQEFDPKVPVPAEKR
 KPIRVLSLF DGIATGLLV LKDLGIQVDRIASEVCEDSITVGMVRHQGKIMYVGDVRSVT
 QKHIQEWGPF DLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDR
 PFFWLFENVVAMGVSDKRDISRFLSNSPVMIDAKEVSAHRARYFWGNLPGMNRPLA
 STVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEME
 RVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSGNSNANSRGPS
FSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLESGSGSG
GGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQ
YALPROESORPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLODVRGRDYQONAMRV
WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSLDAPKVDLLVKNCLLPLREYFKYFS
QNSLPLSRADpkkkrkvGSGatnfsllkqagdveenpgpGSGatnfsllkqagdveenpgpselikenmhmklym
 egtvdnhhfktsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqspegftwervtyedggvltatq
 dtslqdgclinyvknirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttvrskkpaknlkmpgvv
 yvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0387] SEQ ID NO:4 (p92 (KRAB-dCas9-XTEN16-Dnmt3A-Dnmt3L-P2A-P2A-BFP):
 KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), dCas9 (italics), HA
 tag (lowercase), SV40 NLS (lowercase italics), XTEN16 (16 amino acid sequence), Dnmt3A
 (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
 (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage
 sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSR TLVTFKDV FVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEEPGSGGGSGMDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLG
 NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKF
 RGHFLIEGDLNPDNSD VDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQ
 LPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
 LAAKNLS DAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
 KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGE
 LHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVV

DKGASAQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
 KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFL
 DNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHLANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
 LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
 TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
 QLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYH
 HAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNIMNF
 FKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKE
 SILPKRNSDKLLARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS
 SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQITGLYETRIDL
 LGGDSRADypdydpdySGSpkkkrkvSPGSGSETPGTSESATPESNHDQEFDPKVPVPPVPAE
 KRKPIRVLSLFDGIATGLLVKDLGIQVDRYASEVCEDSITVGMVRHQGKIMYVGDVRS
 VTQKHQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGD
 DRPFFWL FENVVAMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGMNRP
 LASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTE
 MERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSR
 GPSFSSGLVPLSLRGSHMGPMIYKTVSAWKRQPVRLSLFRNIDKVLKSLGFLSGS
SGGGTCLKYVEDVTNVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHR
ILQYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNA
MRVWSNIPGLKSKHAPLTPKEEEYLOAQVRSRSLDAPKVDLLVKNCLLPLREYF
KYFSQNSLPLSRADpkkkrkvGSGatnfsllkqagdveenpgpGSGatnfsllkqagdveenpgpselikenmh
mklymegtvdnhhfktsegegkpyegqtmrikvveggplpfafdilatsflygsktfinhtqgipdfkqsfpegftwervttvedg
gvltatqdtslqdgcliynvkirgvnftsnpgvmqkktlgweaftetlypadgglegmdmalklvggshlianikttyrskkpaknlk
mpgvyyvdyrlrikeannetyveqhevavarycdlpsklghkl*

[0388] SEQ ID NO:5 (p93 (KRAB-dCas9-XTEN80-Dnmt3A-Dnmt3L-P2A-BFP): KRAB
 (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), dCas9 (italics), HA tag
 (lowercase), SV40 NLS (lowercase italics), XTEN80 (80 amino acid sequence), Dnmt3A (bold
 italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics
 underlined; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage

sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSR**TLVTFKDV****FVDF****TREEWKLLDTAQQIVYRNV****MLENYKNLV****SLGY**
QLTKPDVILRLEKGE**EP****CGSGGGSMDKKYSIGLAIGTNSVGWAVITDEYKVP****SKKFKVLG**
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD**SFFHR**
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD**KADLR****LIYLALAHMIKF**
RGHFLIEGDLNPDNSDVKLFIQLVQTYNQ**LFEENPINASGVDAKAILSARLSKSR****RLENLIAQ**
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ**LSKDTYDDDLDNLLAQIGDQYADLF**
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFDQ**S**
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV**KLNR****EDLLRKQRTFDNGSIPHQIHLGE**
LHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**EVV**
DKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAI**VDLLFK****TRNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT****YHDL****LKIIKDKDFL**
DNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ**LKRRRYTG****WGR****LSRK****LINGI**
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGD**SLHEHIANLAGSPA**
IKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIE**EGIKELG****SQI**
LKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD**DAIVPQ****SFLKDD****SIDNKVL**
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLNAKLITQRKFDNLT**KAERG****GLSELDKAGFIKR**
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS**KLVSDFR****KDFQFYK****REINNYH**
HAHDAYLNAVVG**TALIKKYPKLESEFVYG****DYK****VYDVRK****MI****AKSEQ****EIGKATAKYFFYS****NIMNF**
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL**SMPQVNIVKKTEVQTGGFSKE**
SILPKRNSDKLLARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPE**DNEQKQLFVEQH****KHYLDEIIEQISEFSKRVILADANLDKVL****SAYN****KHR**
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ**SITGLYETR****IDL****SQ**
LGGDSRADypdydpdyaSGSpkkkrkvSPGGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESG
PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEN**H****Q****E****F****D****P****P****K****V****Y****P****P****V**
AEKRKPIRVLSLFDGIATGLLVKDLGIQVDRYASEVCEDSITVGMVRH**Q****G****KIM****YV****G****DV**
RSVTQKH**Q****E****W****G****P****F****D****L****V****I****G****G****S****P****C****N****D****L****S****I****V****N****P****A****R****K****G****L****Y****E****G****T****G****R****L****F****F****E****F****Y****R****L****L****H****D****A****R****P****K****E****G**
DDRPF**FWL****FENVVAMGVSDKRD****ISRFLES****NPV****MIDAKEV****SAAHRARYFWGNLPGMNR**
PLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQ**GKDQ****HFPVFMNEKEDILWCT**
EMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACV**SSGNSNANS**
RGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLESG
SGSGGGT**LKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCDRCPGWYMFQFH**
RILQYALPROESQRPF**FWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN**

AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSLDAPKVDLLVKNCLLPREY
FKYFSQNSLPLSRAD*pkkrkv***GS****Gatnfsllkqagdveenpgps***elikenmhmklmegtvdnhhfktsegeg*
kpvegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegtwerwttyedggvltatqdtlqdgcliynvki
nftsnpgvmqkktlgweaftetlypadgglegmdmalklvqgshlianikttyskpkaknlkmpgvvyvdyrlerikeannetyv
*eqhevavarycdlpsklghkln**

[0389] SEQ ID NO:6 (p94 (KRAB-dCas9-XTEN80-Dnmt3A-Dnmt3L-P2A-P2A-BFP):
 KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), dCas9 (italics), HA
 tag (lowercase), SV40 NLS (lowercase italics), XTEN80 (80 amino acid sequence), Dnmt3A
 (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
 (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage
 sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPPGGSGGSGMDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKF
RGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQ
LPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVV
DKGASASQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKHKDKDFL
DNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEIGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYH
HAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEI GKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHR

DKPIREQAENIHFLFTLTNLGAPAAFKYFDDTTDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGDSRADypdydpdy aSGSpkkrkvSPGGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESG
 PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSENHDQEFDPKVPVPPVP
 AEKRKPIRVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDV
 RSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEG
 DDRPFFWL FENVVAMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGMNR
 PLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT
 EMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANS
 RGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLESG
SGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCDRCPGWYMFQFH
RILOYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN
AMRVWSNIPGLKSKHAPLTPKEEY LQAQVRSRSKLDAPKVDLLVKNCLLPLREY
FKYFSQNSLPLSRADpkkrkvGSGatnfsllkqagdveenpgpGSGatnfsllkqagdveenpgpselikenm
hmklymegtdvnhhfktsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhgtqgipdffkqsfpegftwervttyed
ggvltatqdtlsldgcliyvknirgvnftsngpvmqkktlgweaftety padgglegndmalklvvggshlianikttyskpkaknl
kmpgvyyvdrylerikeannetyveqhevavarycdlpsklghkln*

[0390] SEQ ID NO:7 (p95 (KRAB-XTEN16-dCas9-Dnmt3A-Dnmt3L-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN16 (16 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSR TLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEEPSGSETPGTSESATPESMDKKYSIGLAIGTNSVGWAVITDEYKVP
 KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
 DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLA
 LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
 LENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGD
 QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIP
 HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRF AWMTRKSEETITP
 WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKP
 AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII

KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLS
 RKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
 NLAGSPAIIKKGILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
 IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
 REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF
 YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQT
 GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELL
 GITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGKRMLASAGELQKGNELAL
 PSKYVNFLLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL
 AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQITGLYE
 TRIDLSQLGGDSRADypdydpdySGSpkkrkvEASGSGRASPPIGSTRNHDQEFDPKPYVP
 VPAEKRKPIRVLSLFDGIATGLLVKDLGIQVDRIASEVCEDSITVGMVRHQGKIMYVG
 DVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPK
 EGDDRPFVFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGM
 NRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILW
 CTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNA
NSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLE
SGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQ
FHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVGRDY
QNAMRVWSNIPGLKSKHAPLTPKEEYLYQAQVRSRSLDAPKVDLLVKNCLPLR
EYFKYFSQNSLPLSRADpkkrkvGSGatnfsllkqagdveenpgpselikenmhmklmegtdnhhfktse
gegkpyegtqtmrikvveggplpfafdilatsflygsktfinhqtgipdfkqsfpegftwervttvedggvltatqdtlsldgclivnki
rgvnftsngpvmqkktlgweaftetlypadgglegrmdmalklyggshlianiktyrskkpaknlkmpgvvyvdyrlerieanne
tyveqhevavarycdlpsklghkln*

[0391] SEQ ID NO:8 (p96 (KRAB-XTEN16-dCas9-Dnmt3A-Dnmt3L-P2A-P2A-BFP):

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN16 (16 amino
 acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A
 (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
 (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage
 sequence (lowercase bold); BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY

QLTKPDVILRLEKGEEPSGSETPGTSESATPESMDKKYSIGLAIGTNSVGWAVITDEYKVPS
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAQLQSKDXYDDDLNLLAQIGD
QYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
KDKDFLDNEENEDILEDIVLTLTFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWRSL
RKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIAKKGILQTVKVVDELVKVMGRHKKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDD
SIDNKVLTNRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGPEDNEQQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL
AYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQITGLYE
TRIDLSQLGGDSRAD^{pydvpdy}aSGS^{pkkrkv}EASGSGRASP GIPGSTR^{NHDQE}FDPPKVYPP
VPAEKRKPIRVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVG
DVRSVTQKHQIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPK
EGDDRPFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGM
NRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQKGDKQHFVFMNEKEDILW
CTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNA
NSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLE
SGSGSGGGTLKYVEDVTNVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQ
FHRILQYALPROESQRPFFWIFMDNLLLTEDDOETTTRFLOTEAVTLQDVRGRDY
QNAMRVWSNIPGLKSKHAPLTPKEEEYLOAQVRSRSKLDAPKVDLLVKNCLLPLR
EYFKYFSQNSLPLSRAD^{pkkrkv}GS^{Gatnfsllkqagdveenpgp}GS^{Gatnfsllkqagdveenpgp}selike
nmhmklmegtvdnhhfktsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhqtgipdfkqsfpegtwerwtt
yedggyltatqdtlsqdgcliynvkiirgvnfsngpvmqkktlgweaftetlypadgglegmdmalklvggshlianikttyskka

knlnmpgvyvdyrlrikeannetyveqhevavarycdlpsklghkln*

[0392] SEQ ID NO:9 (p97 (KRAB-XTEN80-dCas9-Dnmt3A-Dnmt3L-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (80 amino acid sequence), dCas9 (italics); HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

**DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSE
GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSHEMA
KVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDSTDKADLRIL
YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRENLIAQLPGEKKNGLFGNLLALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQ
IGDQYADFLAAKNLSDAILLSDLRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIHKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHE
HIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTAKERGGLS
ELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF
YKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVK
ELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNE
LALPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTIG
LYETRIDLSQLGGDSRADypdydpdySGSpkkrkvEASGSGRASPGIPGSTRNHDQEFDPKVK
YPPVPAEKRKPIRVLSLFDGIATGLLVKDLGIQVDRIASEVCEDSITVGMVRHQGKIM**

YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDA
 RPKEGDDRPFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSAAHRARYFWGNL
 PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKE
 DILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSG
NSNANSRGPSSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLFRNIDKVLKSL
GFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGW
YMFQFHRILQYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEYLAQAVRSRSLDAPKVDLLVKNC
LLPLREYFKYFSQNSLPLSRADpkkkrkvGSGatnfsllkqagdveenpgpselikenmhmklmegtvdn
hhfktsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhgtqgipdffkqspegftwervttvedggvltatqdtlsld
gclivnvmkngvntsnpgvmqkktlgweaftetlypadgglegndmalklvvgshlianikttvrskkpaknlkmpgvyvdyrl
erikeannetyveqhevavarycdlpsklghkln*

[0393] SEQ ID NO: 10 (p98 (KRAB-XTEN80-dCas9-Dnmt3A-Dnmt3L-P2A-P2A-BFP):
 KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (80 amino
 acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A
 (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
 (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage
 sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEPPGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSE
 GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEMDKKYSIGLAIGTNSVGWAVITDEYKV
 PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSHEMA
 KVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDSTDKADLRLLI
 YLALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEEPNINASGVDAKAILSARLSK
 SRLENLIAQLPGEKKNGLFGNLLALSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQ
 IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
 KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
 GSIPHQIHLGELHAILRRQEDFYFLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEE
 TITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGM
 RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
 LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHE
 HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI

EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
 KDDSIDNKVLRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLS
 ELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQF
 YKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDKVYDVVRKMLAKSEQEIGKATAK
 YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTE
 VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVK
 ELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNL
 LALPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDK
 VLSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTG
 LYETRIDLSQLGGDSRADypdydpdySGSpkkkrkvEASGSGRASPGIPGSTRNHDQEFDPKVK
 YPPVPAEKRKPIRVLSLFDGIATGLLVKDLGIQVDRIASEVCEDSITVGMVRHQGKIM
 YVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDA
 RPKEGDDRPFFWLFEENVVAMGVSDKRDISRFLSNPVMIDAKEVSAAHRARYFWGNL
 PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKE
 DILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSG
NSNANSRGPSSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLFRNIDKVLKSL
GFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGW
YMFQFHRILQYALPROESORPFFWIFMDNLLLTEDDQETTTRFLOTEAVTLQDVR
GRDYQONAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSLDAPKVDLLVKNC
LLPLREYFKYFSQNSLPLSRADpkkkrkvGSGatnfsllkqagdveenpgpGSGatnfsllkqagdveenp
gtselikenmhmklymegtdnhhfktsegegkpyegqtmrikvveggplpfadilatsflygsktfinhqtgipdfkqsfpeg
ftwervttvedggvltatqdtlsldgeliynvkiqgnvftsnpgvmqkktlgweaftetlypadgglegrmdmalklvggshlianikt
yrskkpaknlkmpgvvyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0394] SEQ ID NO:11 (p99 (KRAB-XTEN16-dCas9-XTEN80-Dnmt3A-Dnmt3L-P2A-BFP):
 KRAB (bold, from Gilbert et al., Cell, 2013, 2014), XTEN16 (16 amino acid sequence), dCas9
 (italics), HA tag (lowercase), Linkers (underlined), SV40 NLS (lowercase italics), XTEN80
 (lowercase italics bold, 80 amino acid sequence), Dnmt3A (bold italics, from Siddique et al.,
 JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique
 et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (old underlined, from Siddique et al.,
 JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP
 (lowercase underlined))

DAKSLTAWSRITLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEEPSGSETPGTSESATPESMDKKYSIGLAIGTNSVGWAVITDEYKVP
 KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV

DDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQ LFEENPINASGVDAKAILSARLSKSR
LENLIAQLPGEKKNGLFGNLLALSGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF AWMTRKSEETITP
WNFEVVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL KTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD
SIDNKVLTNRSDKNRGKSDNVPSEEVVKMKNYWRQLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEHIEQISEFSKRVLADANLDKVL S
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYE
TRIDLSQLGGDSRADypdydpdy aSGSpkkrkvSPGggpssgapppsggspagsptsteegtsesatpespggtste
psegsapgs pagsptsteegtsstepsegsapgtstepseNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIAT
GLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLV
IGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENNVAMG
VSDKRDISRFLESNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQECLE
HGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVS
NMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSH
MGPMEIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVT
NVVRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILOQYALPRQESQRPF
FWIFMDNLLLTEDDQETTTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKH
APLTPKEEYLOAQVRSRSLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRADpk
kkrkvGS Gatnfsllkqagdveenpgpselikenmhmklymegtdnhhfkctsegegkpyegqt mrikvveggplpfafdi
latsflygsktfinhtqgipdffkqs fpegftwervttyedggvltatqdt slqdgcliynv kirgvnftsngpvmqkktlgweaftetlyp
adgglegmdmalklvggshlianikttyrskkpaknlkmpgvvyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0395] SEQ ID NO:12 (p100 (KRAB-XTEN16-dCas9-XTEN80-Dnmt3A-Dnmt3L-P2A-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), XTEN16 (16 amino acid sequence), dCas9 (italics), HA tag (lowercase), Linkers (underlined), SV40 NLS (lowercase italics), XTEN80 (lowercase bold italics, 80 amino acid sequence), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPSGSETPGTSESATPESMDKKYSIGLAIGTNSVGWAVITDEYKVP
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENPINASGVDAKILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDAEDAQLQSKDITYDDDLNLLAQIGD
QYADFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAMWTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE
IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPOSFLKDD
SIDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIHKLPKYSLEFENGRKRLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL
AYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYE
TRIDLSQLGGDSRADypdydpdySGSpkklrkvSPGggpssgapppsggspagsptsteegtsesatpesgpgste
psegsapgsagsptsteegtsstepsegsapgtstepseNHDQEFDPKVPYPPVPAEKRRKPIRVL
SLFDGIAT
GLLVLKDGLIQVDRIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLV

IGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFWWLFENVVAMG
 VSDKRDISRFLESNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQECL
 HGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVS
 NMSRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSH
MGPMEIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVT
NVVRDVEKWWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPF
FWIFMDNLLLTEDDQETTTTRFLQTEAVTLQDVRGRDYQONAMRVWSNIPGLKSKH
APLTPKEEYQLAQVRSRSLDAPKVDLLVKNCLLPREYFKYFSQNSLPLSRAD*pk*
*kkrv*GS*Gatnfsllkqagdveenpgp*GS*Gatnfsllkqagdveenpgp*selikenmhmkllymegtdnhhfktsegeg
 kpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttvedggvltatqdtslqdgclinykirkv
 nftsnpgvmqkktlgweaftetlypadgglegmdmalklvggshlianikttvrskkpaknlkmpgvvyvdyrlrikeannetv
 eqhevavarycdlpsklghkln*

[0396] SEQ ID NO:13 (p101 (KRAB-XTEN80-dCas9-XTEN16-Dnmt3A-Dnmt3L-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), XTEN16 (16 amino acid sequence), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEEPggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsapgsapgsptsteegt
 stepsegsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGAL
 LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDTKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
 SDVDKLFILQLVQTYNQLEENPINASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFNGLI
 ALSLGLTPNFKSNEFLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLLAAKNLSDAILSDI
 LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
 QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYF
 FLKDNREKIEKILTRIPYYVGPLARGNSRFAMTRKSEETTPWNFEVVDKGASAQSFIERM
 TNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTYKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLT
 LTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
 DGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVDEL

VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
 KLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNP
 SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKHVAQ
 ILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVGTA
 LIKKYPKLESEFVYGDYKVYDVRKMLAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR
 PLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLIARK
 KDWDPPKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKG
 YKEVKKDLIIKLPKYSLEFLENKRKMLASAGELQKGNELALPSKYVNFYLYASHYEKLKSGP
 EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHARDKPIREQAENIIHLF
 TLTLNGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGDSRADypdydp
 dy aSGSpkkrkvSPGSGSETPGTSESATPESNHDQEFDPKVPYPPVPAEKRPKPIRVLSLFDGI
 ATGLLVKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDF
 LVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVA
 MGVSDKRDISRFLESNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQEC
 LEHGRIAKFSKVRITITRSNSIKQKGKQHFVPFMNEKEDILWCTEMERVFGFPVHYTD
 VSNMSRLARQRLGRSWSPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRG
SHMGPMIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVED
VTNVVRRDVEKWGPFDLVYGSTOPLGSSCDRCPGWYMFQFHRILLOYALPROESQ
RPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVGRDYQONAMRVWSNIPGLKS
KHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRA
DpkkrkvGSatnfsllkqagdvengpselikenmhmklymegtdnhhfktsegegkpyegqtmrkvveggplpf
afdilatsflygsktfinhqtgipdfkqsfegftwerwttyedggvltatqdtlsldgclinykirkvgnftsnpgvmqkktlgweafte
tlvpadgglegmdmalklvvgshlianikttvrskkpaknlmpgvvyvdyrlrikeannetyveqhevavarycdlpsklghkl
n*

[0397] SEQ ID NO:14 (p102 (KRAB-XTEN80-dCas9-XTEN16-Dnmt3A-Dnmt3L-P2A-BFP): KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), XTEN16 (16 amino acid sequence), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), P2A peptide cleavage sequence (lowercase bold), BFP (lowercase underlined))

DAKSLTAWSRITLVTFKDVFDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
 QLTKPDVILRLEKGEEPggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsagpsagsptsteegt

stepsegsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFVKLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYF
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERM
TNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT
LTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVDEL
VKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHPVENTQLONE
KLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
SEEVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTA
LIKYPKLESEFVYGDYKVYDVRKMLAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR
PLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLA SHYEKLKGSP
EDNEQKQLFVEQH KHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDSRADypdydp
dy aSGSpkkkrkvSPGSGSETPGTSESATPESNHDQEFDP PKVYPPVPAEKRKPIRVLSLFDGI
ATGLLV LKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKH IQEWGPF D
LVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPF FFWLFENVVA
MGVSDKRDISRFLESNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQEC
LEHGRIAKFSKVRTITTRSNSIKQ GKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTD
VSNM SRLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRG
SHMGPM EIYKTVSAWKROPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVED
VTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPROESQ
RPFFWIFMDNLLL TEDDOETTTRFLQTEAVTLQDV RGRDYONAMRVWSNIPGLKS
KHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRA
DpkkkrkvGS GatnfsllkqagdveenpgpGS Gatnfsllkqagdveenpgpselikenmhmklymegtv dnhhfktse
gegkpyegtqtmrikvveggplpfadilatsflygsktfinhtqgipdfkqsfpegftwervttyedggvltatqdtlsldgclinyvki
rgvnftsngpvmqkktlgweaftetlypadgglegmdmalklvggshlianikttyrskkpaknlkmpgvvyvdyrlerikeanne

tyveqhevavarycdlpsklghkln*

[0398] SEQ ID NO:15 (V2.1 or p112 (Dnmt3A-Dnmt3L-XTEN80-dCas9-BFP-KRAB); KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), Dnmt3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), BFP (lowercase underlined))

*NHDQEFDPKVPYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSIT
VGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGR
LFFEFYRLLHDARPKEGDDRPFFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSA
AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ
HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLFA
PLKEYFACVSSGNSNANSRGPSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSL
FRNIDKVLKSLGFLESGSGSGGGTLLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPL
GSSCDRCPGWYMFQFHRILQYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLO
TEAVTLQDVRGRDYQONAMRVWSNIPGLKSKHAPLTPKEEEYLOAQVRSRSKLDAP
KVDLLVKNCLLPLREYFKYFSQNSLPLggpssgapppsggspagsptsteegtsesatpesgpgtstepsegs
apgspagsptsteegstepsegsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRL
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAHMIKFRG
HFLIEGDLNPDNSDVKLFQVLVQTYNQFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVV
KGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKHKDKDFLD
NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYH*

HAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNF
 FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
 SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEHEQISEFSKRVILADANLDKVL SAYNKHR
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDL SQ
 LGGD ApydydpdyaSLGSGSpkkkrkvEDpkkkrkvDGIGSGSNGSSGSselikenmhm klymegtvdnhhf
kctsegegkpyegqtmmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdt slqdgcli
ynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvvgshlianikttyskpkpaknlkmpgvyyvdyrl erik
eannetyveqhevavarycdlpsklghklnGGGGGMDAKSLTAW SRTLVTFKDVFVDFTREEWKL
LDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP*

[0399] SEQ ID NO:16 (KRAB; from Gilbert et al., Cell, 2013, 2014)

DAKSLTAW SRTLVTFKDVFVDFTREEWKL LDTAQQIVYRNVMLENYKNLVSLGYQLT
 KPDVILRLEKGEEP

[0400] SEQ ID NO:17 (Linker)

GGSGGS

[0401] SEQ ID NO:18 (Linker)

SGS

[0402] SEQ ID NO:19 (Linker)

EASGSGRASPGIPGSTR

[0403] SEQ ID NO:20 (Linker)

SRAD

[0404] SEQ ID NO:21 (Linker)

GSG

[0405] SEQ ID NO:22 (Linker)

SPG

[0406] SEQ ID NO:23 (dCas9)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPIF
 GNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN

SDVDKLFQVLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
HDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQK
GQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGSELKAGFIKQQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLE
SEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLE
AKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHARDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDL
SQLGGD

[0407] SEQ ID NO:24 (HA tag)

YPYDVDPDYA

[0408] SEQ ID NO:25 (SV40 NLS)

PKKKRKV

[0409] SEQ ID NO:26 (Dnmt3A; residues 612-912; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

NHDQEFDPKVPVPAEKRPVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSITV
GMVRHQGKIMYVGDVRSVTQKHQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL
FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSAAH
RARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHF
PVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLK
EYFACV

[0410] SEQ ID NO:27 (27 amino acid linker; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

SSGNSNANSRGPSFSSGLVPLSLRGSH

[0411] SEQ ID NO:28 (Dnmt3L; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

MGPMEIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLESGSGGGTLKYVEDVTNVV
RRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMD
NLLLTEDDQETTTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEY
LQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL

[0412] SEQ ID NO:29 (P2A peptide cleave sequence)

ATNFSLLKQAGDVEENPGP

[0413] SEQ ID NO:30 (BFP)

SELIKENMHMKLYMEGTVDNHHFKCTSEGEKPYEGTQTMRIKVVEGGPLPFAFDILA
TSFLYGSKTFINHTQGIPDFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNV
KIRGVNFTSNGPVMQKKTGLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANIKTT
YRSKKPAKNLKMGPVYYVDYRLRIKEANNETYVEQHEVAVARYCDLPSKLGHKLN*

[0414] SEQ ID NO:31 (XTEN16 (16 amino acid sequence))

SGSETPGTSESATPES

[0415] SEQ ID NO:32 (XTEN80 (80 amino acid sequence))

GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT
STEPSEGSAPGTSTEPSE

[0416] SEQ ID NO:33 (Dnmt3A-Dnmt3L domain)

NHDQEFDPKVPVPAEKRPVLSLFDGIATGLLVKDLGIQVDRIASEVCEDSITV
GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL
FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAH
RARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHF
PVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAPLK
EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRPVRVLSLFRN
IDKVLKSLGFLESGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDR
CPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPL
REYFKYFSQNSLPL

[0417] SEQ ID NO:34 (ddAsCfp1)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYFIGRTDNLTD
INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNV
FSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSF
PFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL
FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLNRNENVLETAELFNELNSIDLTHIFISH
KKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIIISAAGK
ELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDES
NEVDPEFSARLTGIKLEMEPSLSFYNNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN
KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTTEKTSEGFDMYYDYFPDAAKMI
PKCSTQLKAVTAHFQTHTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ
KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIA
EKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQA
ELFYRPKSRMKRMAHRLGEKMLNKKLDQKTPIPDYLYQELYDYVNHRLSHDLSDEA
RALLPNVITKEVSHEIHKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETP
IIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAAARQAWSVVGTI
KDLKQGYLSQVIHEIVDLMIHYQAVVVLANLNFQKSKRTGIAEKAVYQQFEKMLIDK
LNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTSQSGFLFYVPAPYTSKIDPLTGFVD
PFVWTKIKNHESRKHFLLEGDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAKGTPFIAGKRIVPVIENTHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILP
KLENDSDSHAITMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPPEWP
MDADANGAYHIALKGQLLLNLKESKDLKLQNGISNQDWLAYIQELRN

[0418] SEQ ID NO:35 (ddLbCfp1)

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRYY
LSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKNGEGYKSLF
KKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAPFCINENLT
RYISNMDIFEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIG
GFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVL
EVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKW
NAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIQK
VDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLLDSVKSFENYKAFEGEGKET
NRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKD
KETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPK

VFFSKKWMAYYNPSEDIQKIYKNGTFFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAY
DFNFSETEKYKDIAGFYREVVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDK
SHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNP
NPKKTTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIA
RGERNLLYIVVVDGKGNIVEQYSLNEIINNFNIRIKTDYHSLLDKKEKERFEARQNWTS
IENIKELKAGYISQVVHKICELVEKYDAVIALADLNSGFKNSRVKVEKQVYQKFEKMLI
DKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGF
VNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYG
NRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFM
ALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNI
ARKVLWAIGQFKKAEDEKLDKVIAISNKEWLEYAQTsvkh

[0419] SEQ ID NO:36 (ddFnCfp1)

MYPYDVPDYASGSGMSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK
DYKKAKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKSDDDNLQKDFKSADTI
KKQISEYIKDSEKFNLFNQNLDIAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEAL
EIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPE
AINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGK
FVNGENTKRKGINEYINLYSQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDV
VTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSSQVFD
DYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDL
LDQTNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYVLGVMNKKNNKIFDDKAIKE
NKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLN
GEAELFYRKQSIPKKITHPAKEAIANKNKNDNPKKESVFEYDLIKDKRFTEDKFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGN
DRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVV
FEDLNFSGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKG
YFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTsvlntilQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKNLVIKNEEYFEFVQNR

NN

[0420] SEQ ID NO:97 = V2.2 (DNMT3A-DNMT3L-XTEN80-dCas9-HA-NLS-NLS-XTEN16-KRAB-P2A-BFP)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), DNMT3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), BFP (lowercase underlined)); XTEN16 linker is identified as: (SGSETPGTSESATPES); P2A peptide cleavage sequence is identified as ((ATNFSLLKQAGDVEENPGP)). The use of parentheses in the sequence is merely to identify the XTEN16 linker and the PA2 peptide cleavage sequence.

MNHDQEFDPKVPVPAEKRRPIRVLSLFDGIATGLLVLKDLGIQVDRIASEVCEDSI
TVGMVRHQGKIMYVGDVRSVTQKHQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
RLFFEFYRLLHDARPKEGDDRPFFFWL FENVVAMGVSDKRDISRFLSNPVMIDAKEVS
AAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKD
QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLF
APLKEYFACVSSGNSNANSRGPSSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLF
RNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCD
RCPGWYMFQFHRIOYALPROESORPFFWIFMDNLLLTEDDOETTTRFLOTEAVTLQD
VRGRDYONAMRVWSNIPGLKSKHAPLTPKEEYLOAOVRSRSKLDAPKVDLLVKNCLL
PLREYFKYFSONSLPLggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsapgsapgsptsteegstep
segsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAETRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLLAQLPGEKKNGLFGNLIASL
GLTPNFKSNFDLAEDAQLQSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKD
NREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFD
KNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
ANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

MGRHKPENIVITEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
 LYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNPSEE
 VVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKHVAQILD
 SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVGITALIK
 KYPKLESEFVYGDYKVVYDVRKMAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
 ETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD
 WDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYK
 EVKCDLIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFYLYASHYEKLKGSPEL
 NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTL
 TNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDApydvpdySLG
SGSpkkrkvEDpkkrkvDGIGSGSNGSSGS(SGSETPGTSESATPES)GGGGGMDAKSLTA
WSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVI
LRLEKGEEP((ATNFSLLKQAGDVEENPGP))selikenmhmklvmegtdvnhhfktsegegkpyegtqt
mrikvveggplpfadilatsflygsktfinhgtgipdffkqsfpegftwervttvedggvltatqdtlsldgclinyvkingvnftsngpy
mqkktlgweaftetlypadgglegrndmalklyggshlianikttyskpkaknlkmpgvvyvdyrlerieannetyveqhevav
arycdlpsklghkln*

[0421] SEQ ID NO:98 = V2.3 (DNMT3A-DNMT3L-XTEN80-NLS-dCas9-NLS-XTEN16-KRAB-P2A-BFP)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS (lowercase italics), DNMT3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), BFP (lowercase underlined)); XTEN16 linker is identified as (SGSETPGTSESATPES); P2A peptide cleavage sequence is identified as ((ATNFSLLKQAGDVEENPGP)). The use of parentheses in the sequence is merely to identify the XTEN16 linker and the PA2 peptide cleavage sequence.

MNHDQEFDPKVPVPAEKRRPIRVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSI
 TVGMVRHQGKIMYVGDVRSVTQKHQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
 RLFFEFYRLLHDARPKEGDDRPFWWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVS
 AAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQKGD
 QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLF
 APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLF
RNIDKVLKSLGFLSGSGGGTLKYVEDVTNVRRDVEKWGPFDLVYGSTOPLGSSCD

RCPGWYMF OFHRILOYALPROESORPFFWIFMDNLLLTEDDOETTTRFLOTEAVTL
ODVRGRDYONAMRVWSNIPGLKSKHAPLTPKEEEYLOAOVRSRSLDAPKVDLLVKNCLL
PLREYFKYFSQNSLPLggpssgappsggspagsptsteegtsesatpespgtstepsgsapgsagsptsteegstep
segsapgtstepsepkkrkvMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYLAHMIKFRGHFLIEGDLNP
 DNSDVDKLFQVLQVTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
 NLIASLSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILL
 SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
 FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSF
 ERMTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
 TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTL
 DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANLAGSPAIIKKGILQTVK
 VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENT
 QLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTNRSDKNRGK
 SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTAKERGGLSELDKAGFIKQVLVETROIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNA
 VVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDK
 LIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDF
 LEAGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLASHYEK
 LKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAE
 NIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDpkkrk
 v(SGSETPGTSESATPES)RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYK
 NLVSLGYQLTKPDVILRLEKGEEP((ATNFSLLKQAGDVEENPGP))selikenmhmklymeg
tvdnhhfktsegegkpyegtqtmrikvveggplpfadilatsflygsktfinhgtqgipdffkqspegftwervttvedggvltatqdt
slqdgcliyvkiqgvnftsnpgvmqkktlgweaftetlypadgglegndmalklyggshlianiktyrskkpaknlkmpgvyyv
dyrlrikeannetyveqhevavarycdlpsklghkln*

[0422] SEQ ID NO:99 = **V2.4** (DNMT3A-DNMT3L-XTEN80-NLS-dCas9-NLS-XTEN16-BFP-KRAB)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), HA tag (lowercase), SV40 NLS

(lowercase italics), DNMT3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), BFP (lowercase underlined)); XTEN16 linker is identified as (SGSETPGTSESATPES). The use of parentheses in the sequence is merely to identify the XTEN16 linker sequence.

*MNHDQEFDPKVPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRIASEVCEDSI
TVGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
RLFFEFYRLLHDARPKEGDDRPFFWL FENVVAMGVSDKRDISRFLSNPVMIDAKEVS
AAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKD
QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLF
APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLF
RNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCD
RCPGWYMFQFHRIOYALPROESORPFFWIFMDNLLLTEDDOETTTTRFLOTEAVTLQD
VRGRDYQONAMRVWSNIPGLKSKHAPLTPKEEYLOAQVRSRSLDAPKVDLLVKNCLL
PLREYFKYFSONSLPLggpssgapppsggspagsptsteegtsesatpespgtstepsgsapgsagsptsteegstep
segsapgtstepsepkkrkvMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADFLAAKNLSDAILL
SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
FYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSF
ERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVENT
QLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNA
VVGTAIIKKYPKLESEFVYGDYKVDYVRKMAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDK*

LIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDF
 LEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEK
 LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAE
 NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDpkkkrk
 v(SGSETPGTSESATPES)selikenmhm kly megtv dnhhf kctsegegkpyegtqtmrikvveggplpfafdilatsf
 lygsktfinh tqgipdffkqsfpegftwervttvedggvltatqdtslqdgcliynv kirgvnftsngpvmqkktlgweaftetly padg
 glegmdmalklvggshlianikttyskpkaknlkmpgvyyvdrylerikeannetyveqhevavarycdlpsklghklnGGGG
**GMDAKSLTAW SRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSL
 GYQLTKPDVILRLEKGEEP***

[0423] SEQ ID NO:100

GACGCTCAAATTTCCGCAGTGTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTT
 AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0424] SEQ ID NO:101

GTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTAT
 CAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

[0425] SEQ ID NO:102

GACGCTCAAATTTCCGCAGT

[0426] SEQ ID NO:103 (KRAB with initiating methionine)

MDAKSLTAW SRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQL
 TKPDVILRLEKGEEP

[0427] SEQ ID NO:104 (KRAB)

RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEK
 GEEP

[0428] SEQ ID NO:105 (KRAB)

MRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLE
 KGEEP

[0429] SEQ ID NO:106 (Dnmt3A with methionine on N-terminus)

MNHDQEFDPKVPVPAEKRPKPIRVL S LFDGIATGLLV LKDLGIQVDRIYASEVCEDSI
 TVGMVRHQGKIMYVG D VRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
 RLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
 AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ

HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFAP
LKEYFACV

[0430] SEQ ID NO:107 (V2.1 or p112 (Dnmt3A-Dnmt3L-XTEN80-dCas9-BFP-KRAB);
KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase
bold italics, 80 amino acid sequence), dCas9 (italics), SV40 NLS (lowercase italics), Dnmt3A
(bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
(italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), Dnmt3L (bold
underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016),

NHDQEFDPKVPVPAEKRPVLSLFDGIATGLLVLKDLGIQVDRYASEVCEDSIT
VGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGR
LFEEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVSA
AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ
HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLFA
PLKEYFACVSSGNSNANSRGPSSSGLVPLSLRGSHMGPMEIYKTVSAWKROPVRVLSL
FRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPL
GSSCDRCPGWYMFQFHRILQYALPROESQRPFFWIFMDNLLLTEDDQETTTRFLO
TEAVTLQDVRGRDYQONAMRVWSNIPGLKSKHAPLTPKEEEYLOAQVRSRSLDAP
KVDLLVKNCLLPLREYFKYFSQNSLPLggpssgapppsggspagsptsteegtsesatpesgpgtstepsegs
apgpagsptsteegstepsegsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRG
HFLIEGDLNPDNSDVKLFQVLVQTYNQFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKHKDKDFLD
NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYH

HAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKYYDVVRKMIKSEQEIGKATAKYFFYSNIMNF
 FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVEKSKKLKSVKELLGITIMERS
 SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEHEQISEFSKRVILADANLDKVL SAYNKHR
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETRIDL SQ
 LGGDASLGSGSpkkkrkvEDpkkkrkvDGIGSGSNGSSSGSGGGGGMDAKSLTAWSR TLVTFK
DVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP

*

[0431] SEQ ID NO:108 = V2.2 (DNMT3A-DNMT3L-XTEN80-dCas9-HA-NLS-NLS-
 XTEN16-KRAB-P2A-BFP)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase
 bold italics, 80 amino acid sequence), dCas9 (italics), SV40 NLS (lowercase italics), DNMT3A
 (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker
 (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold
 underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), BFP (lowercase
 underlined)); XTEN16 linker is identified as: (**SGSETPGTSESATPES**). The use of
 parentheses in the sequence is merely to identify the XTEN16 linker.

MNHDQEFDPPKVYPPVPAEKRRKPIRVLSLFDGIATGLLVLKDLGIQVDRIYASEVCEDSI
 TVGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
 RLFFEFYRLLHDARPKEGDDRPFFWL FENVVAMGVSDKRDISRFLSNPVMIDAKEVS
 AAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGGK
 QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSPVIRHLF
 APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLF
RNIDKVLKSLGFLESGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCD
RCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDOETTTRFLQTEAVTLQD
VRGRDYONAMRVWSNIPGLKSKHAPLTPKEEYLOAOVRSRSKLDAPKVDLLVKNCLL
PLREYFKYFSONSLPLggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsapgsapgsptsteegtstep
segsapgtstepseMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
 SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHP
 IFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
 DKLFIQLVQTYNQLEENPINASGVDAKAILSARLSKSRLENLLAQLPGEKKNGFLGNLIALSL
 GLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
 TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF

YKFIKPILEKMDGTEELLVKLNRDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKD
 NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFD
 KNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
 QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLF
 EDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
 ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANLAGSPAIAKKGILQTVKVVDELVKV
 MGRHKPENIVITEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
 LYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
 VVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKHVAQILD
 SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTA
 LIK KYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLI
 ETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD
 WDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYK
 EVKCDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
 D NEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTL
 TNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSLITGLYETRIDLSQLGGDASLGSGS
pkkkrv vEDpkkkrv DGIGSGSNGSSGS(SGSETPGTSESATPES)GGGGGMDAKSLTAWSR
**TLVT FKDVVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGE
 EP**

[0432] SEQ ID NO:109 = V2.3 (DNMT3A-DNMT3L-XTEN80-NLS-dCas9-NLS-XTEN16-KRAB-P2A-BFP)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), SV40 NLS (lowercase italics), DNMT3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), XTEN16 linker is identified as (SGSETPGTSESATPES). The use of parentheses in the sequence is merely to identify the XTEN16 linker.

MNHDQEFDPKVPYPPVPAEKRKPIRVLSLFDGIATGLLVKDLGIQVDRYIASEVCEDSI
 TVGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
 RLFFEFYRLLDARPKEGDDRPFVFWLFENVVAMGVSDKRDISRFLSNPVMIDAKEVS
 AAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKD
 QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVPVIRHLF
 APLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKROPVRVLSLF

RNIDKVLKSLGFLESGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTOPLGSSCD
RCPGWYMFQFHRILOYALPROESORPFFWIFMDNLLLTEDDOETTTRFLOTEAVTLQD
VRGRDYQONAMRVWSNIPGLKSKHAPLTPKEEEYLOAQVRSRSLDAPKVDLLVKNCLL
PLREYFKYFSONSLPL*ggpssgapppsggspagsptssteegtsesatpespgtstepsegsapgsapgsptssteegtstep*
*segsapgtstepsepkkrkv*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
 DNSDVDKLFQVLQVQTYNQVFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFG
 NLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILL
 SDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG
 GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
 FYPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF
 ERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
 TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILE
 DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTL
 DFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
 VVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENT
 QLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGK
 SDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDFQFYKVRINNYYHHAHDAYLNA
 VVGTAIIKKYPKLESEFVYGDYKVYDVVKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDK
 LIARKKDWDPPKKGFFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDF
 LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEK
 LKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAE
 NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGD*pkkrkv*
 v(SGSETPGTSESATPES)RTLVTFKDVVDFFTREEWKLLDTAQQIVYRNVMLENYK
 NLVSLGYQLTKPDVILRLEKGEEP

[0433] SEQ ID NO:110 = V2.4 (DNMT3A-DNMT3L-XTEN80-NLS-dCas9-NLS-XTEN16-BFP-KRAB)

KRAB (bold, from Gilbert et al., Cell, 2013, 2014), Linkers (underlined), XTEN80 (lowercase bold italics, 80 amino acid sequence), dCas9 (italics), SV40 NLS (lowercase italics), DNMT3A (bold italics, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), 27 amino acid linker (italics underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), DNMT3L (bold

underlined, from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016), XTEN16 linker is identified as (SGSETPGTSESATPES). The use of parentheses in the sequence is merely to identify the XTEN16 linker sequence.

*MNH**DQ**EF**DP**PKV**YPP**VPA**EKR**KPI**RVL**SL**FD**GI**ATG**LL**VL**KD**LGI**QV**DR**YI**ASE**V**CE**DSI*
*TV**GM**VR**HQ**GK**IM**YV**GD**VR**SV**TQ**KHI**Q**EW**GP**FD**LVI**GG**SP**CND**LS**IV**N**P**AR**K**GL**Y**E**GT**G*
*RL**FF**E**F**Y**RL**LH**D**AR**P**K**E**G**DD**RP**FF**WL**F**EN**V**A**M**G**V**S**D**K**R**D**I**S**R**F**L**E**S**N**P**M**I**D**A**K**E**V**S*
*AA**H**R**A**R**Y**F**W**G**N**L**P**G**M**N**R**P**L**A**S**T**V**N**D**K**L**E**L**Q**E**C**L**E**H**G**R**I*A**K**F**S**K**V**R**T**I*T**T**R**S**N**S**I**K**Q**G**K**D*
*Q**H**F**P**V**F**M**N**E**K**E**D**I**L**W**C**T**E**M**E**R**V**F**G**F**P**V**H**T**D**V**S**N**S**R**L**A**R**Q**R**L**L**G**R**S**W**S**V**P**V**R**H**L**F*
*A**P**L**K**E**Y**F**A**C**V**S**S**G**N**S**N**A**N**S**R**G**P**S**F**S**S**G**L**V**L**S**L**R**G**S**H**M**G**P**M**E**I**Y**K**T**V**S**A**W**K**R**O**P**V**R**V**L**S**L**F*
*R**N**I**D**K**V**L**K**S**L**G**F**L**E**S**G**S**G**S**G**G**T**L**K**Y**E**D**V**T**N**V**R**R**D**V**E**K**W**G**P**F**D**L**V**G**S**T**O**P**L**G**S**S**C**D*
*R**C**P**G**W**Y**M**F**O**F**H**R**I**O**A**L**P**R**O**E**S**O**R**P**F**F**W**I**F**M**D**N**L**L**L**T**E**D**D**O**E**T**T**T**R**F**L**O**T**E**A**V**T**L**O**D*
*V**R**G**R**D**Y**O**N**A**M**R**V**S**N**I**P**G**L**K**S**K**H**A**P**L**T**P**K**E**E**E**Y**L**O**A**Q**V**R**S**R**S**K**L**D**A**P**K**V**D**L**L**V**K**N**C**L**L*
*P**L**R**E**Y**F**K**Y**F**S**O**N**S**L**P**L**g**p**s**s**g**a**p**p**p**s**g**s**p**a**s**p**t**s**t**e**e**g**t**s**a**t**p**e**s**g**p**g**t**s**e**g**s**a**p**s**p**a**s**p**t**s**t**e**e**g**t**s**e**g**t**s*
*se**g**s**a**p**g**t**s**e**p**k**k**r**k**v**M**D**K**K**Y**S**I**G**L**A**I**G**T**N**S**V**G**W**A**V**I**T**D**E**Y**K**V**P**S**K**K**F**K**V**L**G**N**T**D**R**H**S**I**K**K**N**L**I**G*
*A**L**L**F**D**S**G**E**T**A**E**A**T**R**L**K**R**T**A**R**R**R**Y**T**R**R**K**N**R**I**C**Y**L**Q**E**I**F**S**N**E**M**A**K**V**D**D**S**F**F**H**R**L**E**E**S**F**L**V**E**E**D**K**K*
*H**E**R**H**P**I**F**G**N**I**V**D**E**V**A**Y**H**E**K**Y**P**T**I**Y**H**L**R**K**L**V**D**S**T**D**K**A**D**L**R**L**I**Y**L**A**H**M**I**K**F**R**G**H**F**L**I**E**G**D**L**N**P*
*D**N**S**D**V**D**K**L**F**I**Q**L**V**Q**T**Y**N**Q**L**F**E**E**N**P**I**N**A**S**G**V**D**A**K**I**L**S**A**R**L**S**K**S**R**R**L**E**N**L**I**A**Q**L**P**G**E**K**K**N**G**L**F**G*
*N**L**I**A**L**S**L**G**L**T**P**N**F**K**S**N**F**D**L**A**E**D**A**K**L**Q**L**S**K**D**T**Y**D**D**D**L**N**L**A**Q**I**G**D**Q**Y**A**D**L**F**L**A**A**K**N**L**S**D**A**I**L*
*S**D**I**L**R**V**N**T**E**I**T**K**A**P**L**S**A**M**I**K**R**Y**D**E**H**H**Q**D**L**T**L**K**A**L**V**R**Q**L**P**E**K**Y**K**E**I**F**D**Q**S**K**N**G**Y**A**G**Y**I**D**G*
*G**A**S**Q**E**E**F**Y**K**F**I**K**P**I**L**E**K**M**D**G**T**E**L**L**V**K**L**N**R**E**D**L**L**R**K**Q**R**T**F**D**N**G**S**I**P**H**Q**I**H**L**G**E**L**H**A**I**L**R**R**Q**E**D*
*F**Y**P**F**L**K**D**N**R**E**K**I**E**K**I**L**T**F**R**I**P**Y**Y**V**G**P**L**A**R**G**N**S**R**F**A**W**M**T**R**K**S**E**E**T**T**P**W**N**F**E**E**V**D**K**G**A**S**A**Q**S**F**I*
*E**R**M**T**N**F**D**K**N**L**P**N**E**K**V**L**P**K**H**S**L**L**Y**E**F**T**V**Y**N**E**L**T**K**V**K**Y**V**T**E**G**M**R**K**P**A**F**L**S**G**E**Q**K**K**A**I**V**D**L**L**F**K*
*T**N**R**K**V**T**V**K**Q**L**K**E**D**Y**F**K**K**I**E**C**F**D**S**V**E**I**S**G**V**E**D**R**F**N**A**S**L**G**T**Y**H**D**L**L**K**I**K**D**K**D**F**L**D**N**E**E**N**E**D**I**L*
*D**I**V**L**T**L**T**L**F**E**D**R**E**M**I**E**R**L**K**T**Y**A**H**L**F**D**D**K**V**M**K**Q**L**K**R**R**R**Y**T**G**W**G**R**L**S**R**K**L**I**N**G**I**R**D**K**Q**S**G**K**T**I*
*D**F**L**K**S**D**G**F**A**N**R**N**F**M**Q**L**H**D**D**S**L**T**F**K**E**D**I**Q**K**A**Q**V**S**G**Q**G**D**S**L**H**E**H**A**N**L**A**G**S**P**A**I**K**K**G**I**L**Q**T**V**K*
*V**V**D**E**L**V**K**V**M**G**R**H**K**P**E**N**I**V**E**M**A**R**E**N**Q**T**T**Q**K**G**Q**K**N**S**R**E**R**M**K**R**I**E**E**G**I**K**E**L**G**S**Q**I**L**K**E**H**P**V**E**N**T*
*Q**L**Q**N**E**K**L**Y**L**Y**L**Q**N**G**R**D**M**Y**V**D**Q**E**L**D**I**N**R**L**S**D**Y**D**V**D**A**I**V**Q**S**F**L**K**D**D**S**I**D**N</***

NIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI^vTGLYETRIDLSQLGGDpkkkrk
v(SGSETPGTSESATPES)GGGGGMDAKSLTAWSR^vTLVTFKDV^vFDFTREEWKLLDT
AQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP*

[0434] SEQ ID NO:111

5'-TAATACGACTCACTATAGG-3'

[0435] SEQ ID NO:112

CGAGCTGCGAGCAAAGAT

[0436] SEQ ID NO:113

CGTGTCTCCACAGGTCACAG

[0437] SEQ ID NO:114

GGGCTCTCCAGAACATCATC

[0438] SEQ ID NO:115

CCTGCTTCACCACCTTCTTG

[0439] SEQ ID NO:116

AGATCGGAAGAGCACACGTCTGAACTC

[0440] SEQ ID NO:117

AGATCGGAAGAGCACACGTCTGAACTCCAGTCA

[0441] SEQ ID NO:118

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTCCACCTTGTTG

[0442] SEQ ID NO:119

gtttcagagcgagacgtgcctgcaggatacgtctcagaaacatg

[0443] SEQ ID NO:120

GTTTAAGAGCTAAGCTG

[0444] SEQ ID NO:121

AATGATACGGCGACCACCGAGATCTACACCGCGGTCTGTATCCCTTGGAGAACCAC
CT

[0445] SEQ ID NO:122

CAAGCAGAAGACGGCATACGAGATXXXXXGCGGCCGGCTGTTTCCAGCTTAGCTCT

TAAA

[0446] SEQ ID NO:123

CGCGGTCTGTATCCCTTGGAGAACCACCTTGTGTTGG

[0447] SEQ ID NO:124

GCGGCCGGCTGTTTCCAGCTTAGCTCTTAAAC

[0448] SEQ ID NO:125

GTTTAAGAGCTAAGCTGGAAACAGCCGGCCGC

[0449] SEQ ID NO:126

CCACCTTGTG

[0450] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. All documents, or portions of documents, cited in the application including, without limitation, patents, patent applications, articles, books, manuals, and treatises are hereby expressly incorporated by reference in their entirety for any purpose.

[0451] While various embodiments and aspects of the present invention are shown and described herein, it will be obvious to those skilled in the art that such embodiments and aspects are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention.

[0452] References

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167, 1867-1882.e21. Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* 167, 219-232.e14. Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38, 824–844. Audergon, P.N.C.B., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A.L., and Allshire, R.C. (2015). Restricted epigenetic inheritance of H3K9 methylation. *Science* 348, 132–135. Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y.,

Uno, N., Oshimura, M., and Elowitz, M.B. (2016). Dynamics of epigenetic regulation at the single-cell level. *Science* 351, 720–724. Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096.

Boyes, J., and Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* 11, 327–333. Buiting, K., Williams, C., and Horsthemke, B. (2016). Angelman syndrome — insights into a rare neurogenetic disorder. *Nat. Rev. Neurol.* 12, 584–593. Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly-efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinforma. Oxf. Engl.* 34, i884–i890. Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* 173, 1398–1412.e22. Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022. Doench, J.G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* 19, 67–80. Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* 354, 769–773. Galonska, C., Charlton, J., Mattei, A.L., Donaghey, J., Clement, K., Gu, H., Mohammad, A.W., Stamenova, E.K., Cacchiarelli, D., Klages, S., et al. (2018). Genome-wide tracking of dCas9-methyltransferase footprints. *Nat. Commun.* 9. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* 154, 442–451. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 159, 647–661. Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells. *Cell Stem Cell* 7, 249–257. Hanna, R.E., and Doench, J.G. (2020). Design and analysis of CRISPR–Cas experiments. *Nat. Biotechnol.* 38, 813–823. Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSsmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* 13, R83. Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R., and Moffat, J. (2014). Measuring error rates in genomic

perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* *10*, 733.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. *Cell* *149*, 1447–1460.

Hofacker, D., Broche, J., Laistner, L., Adam, S., Bashtrykov, P., and Jeltsch, A. (2020). Engineering of Effector Domains for Targeted DNA Methylation with Reduced Off-Target Effects. *Int. J. Mol. Sci.* *21*.

Holtzman, L., and Gersbach, C.A. (2018). Editing the Epigenome: Reshaping the Genomic Landscape. *Annu. Rev. Genomics Hum. Genet.* *19*, 43–71.

Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., et al. (2016a). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *ELife* *5*, e19760.

Horlbeck, M.A., Witkowski, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., and Weissman, J.S. (2016b). Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *ELife* *5*.

Hovestadt, V., Jones, D.T.W., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., et al. (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* *510*, 537–541.

Iglesias, N., Currie, M.A., Jih, G., Paulo, J.A., Siuti, N., Kalocsay, M., Gygi, S.P., and Moazed, D. (2018). Automethylation-induced conformational switch in Ctr4 (Suv39h) maintains epigenetic stability. *Nature* *560*, 504–508.

Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* *24*, 939–946.

Iqbal, K., Liu, F., and Gong, C.-X. (2016). Tau and neurodegenerative disease: the story so far. *Nat. Rev. Neurol.* *12*, 15–27.

Isaac, R.S., Jiang, F., Doudna, J.A., Lim, W.A., Narlikar, G.J., and Almeida, R. (2016). Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *ELife* *5*, e13450.

Jost, M., Santos, D.A., Saunders, R.A., Horlbeck, M.A., Hawkins, J.S., Scaria, S.M., Norman, T.M., Hussmann, J.A., Liem, C.R., Gross, C.A., et al. (2020). Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* *38*, 355–364.

Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* *361*, 866–869.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* *517*, 583–588.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.

Leonetti, M.D., Sekine, S., Kamiyama, D., Weissman, J.S., and Huang, B. (2016a). A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U. S. A.* *113*, E3501–E3508.

Leonetti, M.D., Sekine, S., Kamiyama, D., Weissman, J.S., and Huang, B. (2016b). A scalable strategy for high-throughput

GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci.* *113*, E3501–E3508. Li, X.-L., Li, G.-H., Fu, J., Fu, Y.-W., Zhang, L., Chen, W., Arakaki, C., Zhang, J.-P., Wen, W., Zhao, M., et al. (2018). Highly efficient genome editing via CRISPR–Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res.* *46*, 10195–10215.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923–930. Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell* *167*, 233–247.e17.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550. Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., et al. (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* *31*, 1137–1142.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* *49*, 1779–1784.

Mlambo, T., Nitsch, S., Hildenbeutel, M., Romito, M., Müller, M., Bossen, C., Diederichs, S., Cornu, T.I., Cathomen, T., and Mussolino, C. (2018). Designer epigenome modifiers enable robust and sustained gene silencing in clinically relevant human cells. *Nucleic Acids Res.* *46*, 4456–4468.

O’Geen, H., Bates, S.L., Carter, S.S., Nisson, K.A., Halmi, J., Fink, K.D., Rhie, S.K., Farnham, P.J., and Segal, D.J. (2019). Ezh2-dCas9 and KRAB-dCas9 enable engineering of epigenetic memory in a context-dependent manner. *Epigenetics Chromatin* *12*, 26.

Park, Y., and Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinforma. Oxf. Engl.* *32*, 1446–1453.

Park, M., Patel, N., Keung, A.J., and Khalil, A.S. (2019). Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* *176*, 227–238.e20.

Ragunathan, K., Jih, G., and Moazed, D. (2015). Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* *348*, 1258699.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44*, W160–165.

Replogle, J.M., Norman, T.M., Xu, A., Hussmann, J.A., Chen, J., Cogan, J.Z., Meer, E.J., Terry, J.M., Riordan, D.P., Srinivas, N., et al. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* *38*, 954–961.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.

Schellenberger, V., Wang, C.-W., Geething, N.C., Spink, B.J.,

Campbell, A., To, W., Scholle, M.D., Yin, Y., Yao, Y., Bogin, O., et al. (2009). A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* 27, 1186–1190. Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16, 299–311. Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution. *Cell* 163, 218–229. Stepper, P., Kungulovski, G., Jurkowska, R.Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T.P. (2017). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. *Nucleic Acids Res.* 45, 1703–1713. Tak, Y.E., Kleinstiver, B.P., Nuñez, J.K., Hsu, J.Y., Horng, J.E., Gong, J., Weissman, J.S., and Joung, J.K. (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* 14, 1163–1166. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinforma. Oxf. Engl.* 31, 2032–2034. Tarjan, D.R., Flavahan, W.A., and Bernstein, B.E. (2019). Epigenome editing strategies for the functional annotation of CTCF insulators. *Nat. Commun.* 10, 4258. Tian, R., Gachechiladze, M.A., Ludwig, C.H., Laurie, M.T., Hong, J.Y., Nathaniel, D., Prabhu, A.V., Fernandopulle, M.S., Patel, R., Abshari, M., et al. (2019). CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* 104, 239–255.e12. Van, M.V., Fujimori, T., and Bintu, L. (2021). Nanobody-mediated control of gene expression and epigenetic memory. *Nat. Commun.* 12, 537. Xu, X., and Qi, L.S. (2019). A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. *J. Mol. Biol.* 431, 34–47. Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2, 1–12. Yeh, C.D., Richardson, C.D., and Corn, J.E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* 21, 1468–1478. Yu, R., Wang, X., and Moazed, D. (2018). Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* 558, 615–619. Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S.B., Wang, Y., et al. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* 554, 387–391.

CLAIMS

What is claimed is:

1. A fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient RNA-guided endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain.
2. The fusion protein of claim 1, wherein the first XTEN linker comprises from about 5 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to about 864 amino acid residues.
3. The fusion protein of claim 2, wherein the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues, and a Krüppel-associated box domain.
4. The fusion protein of claim 3, wherein the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues.
5. The fusion protein of claim 4, wherein the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues.
6. The fusion protein of claim 1, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.
7. The fusion protein of claim 6, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.
8. The fusion protein of claim 6, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.
9. The fusion protein of claim 1, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.
10. The fusion protein of claim 9, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).
11. The fusion protein of claim 1, further comprising an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more

thereof.

12. The fusion protein of claim 11, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, a nuclear localization signal peptide, and the Krüppel-associated box domain.

13. The fusion protein of claim 11, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, a nuclear localization signal peptide, the second XTEN linker, and the Krüppel-associated box domain.

14. The fusion protein of claim 11, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient RNA-guided endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, the second XTEN linker, the Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag.

15. A fusion protein having at least 85% sequence identity to the amino acid sequence of Formula (A); where the amino acid sequence of Formula (A) is, from N-terminus to C-terminus:



wherein:

C₁ comprises SEQ ID NO:26 or SEQ ID NO:106;

R³ is absent or R³ comprises SEQ ID NO:27;

C₂ comprises SEQ ID NO:28;

R² is absent or R² comprises SEQ ID NO:32;

A comprises SEQ ID NO:23;

R¹ is absent or R¹ comprises SEQ ID NO:25;

R⁴ is absent or R⁴ comprises SEQ ID NO:31; and

B comprises SEQ ID NO:16, SEQ ID NO:103, SEQ ID NO:104, or SEQ ID NO:105.

16. The fusion protein of claim 15 having at least 90% sequence identity to the amino acid sequence of Formula (A).

17. The fusion protein of claim 16 having at least 95% sequence identity to the amino acid sequence of Formula (A).

18. A fusion protein having at least 85% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110.

19. The fusion protein of claim 18 having at least 90% sequence identity to SEQ ID NO:97, 98, 99, 107, 108, 109, or 110.
20. A cell comprising the fusion protein of claim 1.
21. The cell of claim 20, wherein the cell is a eukaryotic cell, a mammalian cell, or a stem cell.
22. A method of silencing a target nucleic acid sequence in a cell, the method comprising:
- (i) delivering a first polynucleotide encoding a fusion protein of claim 1 to a cell containing the target nucleic acid; and
 - (ii) delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence.
23. The method of claim 22, wherein the target nucleic acid comprises a CpG island.
24. The method of claim 22, wherein the target nucleic acid does not comprise a CpG island.
25. The method of claim 22, wherein the second polynucleotide comprises sgRNA.
26. A method of silencing a target nucleic acid sequence in a cell, the method comprising:
- (i) delivering a first polynucleotide encoding a fusion protein to a cell containing the target nucleic acid, wherein the target nucleic acid does not comprise a CpG island; wherein the fusion protein comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain; and
 - (ii) delivering to the cell a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby silencing the target nucleic acid sequence in the cell.
27. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising:
- (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain; and
 - (ii) delivering to the subject an effective amount of a second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating Angelman syndrome,

the infectious disease, the tau pathology, or the neurodegenerative disease.

28. The method of claim 26, wherein the second polynucleotide comprises sgRNA.

29. The method of claim 26, wherein the fusion protein comprises, from N-terminus to C-terminus, the DNA methyltransferase domain, the nuclease-deficient RNA-guided DNA endonuclease enzyme, and the Krüppel associated box domain.

30. The method of claim 26, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.

31. The method of claim 30, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.

32. The method of claim 30, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.

33. The method of claim 26, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

34. The method of claim 33, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).

35. The method of claim 26, wherein the dCas9 is covalently linked to the Dnmt3A domain via a peptide linker and wherein the Dnmt3A domain is covalently linked to the Krüppel associated box domain via a peptide linker.

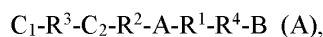
36. The method of claim 35, wherein the peptide linker is a XTEN linker.

37. The method of claim 26, wherein the fusion protein comprises, from N-terminus to C-terminus, the Krüppel associated box, the nuclease-deficient RNA-guided DNA endonuclease enzyme, and the DNA methyltransferase domain.

38. The method of claim 37, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is a CRISPR-associated protein, a leucine zipper domain, a winged helix domain, a helix-turn-helix motif, a helix-loop-helix domain, an HMB-box domain, a Wor3 domain, an OB-fold domain, an immunoglobulin domain, or a B3 domain.

39. The method of claim 38, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9.

40. The method of claim 38, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCpf1 or ddCpf1.
41. The method of claim 37, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.
42. The method of claim 41, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).
43. The method of claim 39, wherein the dCas9 is covalently linked to the Dnmt3A domain via a peptide linker and wherein the Krüppel associated box domain is covalently linked to the dCas9 via a peptide linker.
44. The method of claim 43, wherein the peptide linker is a XTEN linker.
45. The method of claim 37, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the Krüppel associated box domain via a peptide linker.
46. The method of claim 37, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the DNA methyltransferase domain via a peptide linker.
47. The method of claim 37, wherein the Krüppel associated box domain is covalently linked to the DNA methyltransferase domain via a peptide linker.
48. The method of claim 26, wherein the fusion protein further comprises an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.
49. The method of claim 26, wherein the fusion protein further comprises a nuclear localization signal peptide.
50. The method of claim 26, wherein the fusion protein having at least 85% sequence identity to the amino acid sequence of Formula (A); where the amino acid sequence of Formula (A) is, from N-terminus to C-terminus:



wherein:

C₁ comprises SEQ ID NO:26 or SEQ ID NO:106;

R³ is absent or R³ comprises SEQ ID NO:27;

C₂ comprises SEQ ID NO:28;

R² is absent or R² comprises SEQ ID NO:32;

A comprises SEQ ID NO:23;

R¹ is absent or R¹ comprises SEQ ID NO:25;

R⁴ is absent or R⁴ comprises SEQ ID NO:31; and

B comprises SEQ ID NO:16, SEQ ID NO:103, SEQ ID NO:104, or SEQ ID NO:105.

51. The method of claim 50, wherein the fusion protein has at least 90% sequence identity to the amino acid sequence of Formula (A).

52. The method of claim 51, wherein the fusion protein has at least 95% sequence identity to the amino acid sequence of Formula (A).

53. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising:

- (i) delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein of claim 1; and
- (ii) delivering to the subject an effective amount of second polynucleotide comprising sgRNA or cr:tracrRNA; thereby treating Angelman syndrome, the infectious disease, the tau pathology, or the neurodegenerative disease in the subject.

54. The method of claim 26, wherein the infectious disease is a viral infections disease.

55. The method of claim 54, wherein the infectious disease is a Flavivirus infectious disease.

56. A fusion protein comprising, from N-terminus to C-terminus, a DNA methyltransferase domain, a first XTEN linker, a nuclease-deficient endonuclease enzyme, a second XTEN linker, and a Krüppel-associated box domain.

57. The fusion protein of claim 56, wherein the first XTEN linker comprises from about 5 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to about 864 amino acid residues.

58. The fusion protein of claim 57, wherein the first XTEN linker comprises from greater than 50 to about 864 amino acid residues, and the second XTEN linker comprises from about 5 to 50 amino acid residues.

59. The fusion protein of claim 58, wherein the first XTEN linker comprises from about 60 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 40 amino acid residues.

60. The fusion protein of claim 59, wherein the first XTEN linker comprises from about 70 to about 864 amino acid residues, and the second XTEN linker comprises from about 10 to about 30 amino acid residues.

61. The fusion protein of claim 56, wherein the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain or a transcription activator-like effector.

62. The fusion protein of claim 62, wherein the nuclease-deficient DNA endonuclease enzyme is a zinc finger domain.

63. The fusion protein of claim 62, wherein the nuclease-deficient DNA endonuclease enzyme is transcription activator-like effector.

64. The fusion protein of claim 56, wherein the DNA methyltransferase domain comprises a Dnmt3A domain.

65. The fusion protein of claim 64, wherein the Dnmt3A domain is linked to a Dnmt3L domain (Dnmt3A-3L domain).

66. The fusion protein of claim 56, further comprising an epitope tag, a 2A peptide, a fluorescent protein tag, a nuclear localization signal peptide, or a combination of two or more thereof.

67. The fusion protein of claim 56, further comprising a nuclear localization signal peptide.

68. The fusion protein of claim 56, comprising, from N-terminus to C-terminus, the DNA methyltransferase domain, the first XTEN linker, the nuclease-deficient endonuclease enzyme, an epitope tag, a nuclear localization signal peptide, the second XTEN linker, the Krüppel-associated box domain, a 2A cleavable peptide, and a fluorescent protein tag.

69. A cell comprising the fusion protein of claim 56.

70. The cell of claim 69, wherein the cell is a eukaryotic cell, a mammalian cell, or a stem cell.

71. A method of silencing a target nucleic acid sequence in a cell, the method comprising delivering a first polynucleotide encoding a fusion protein of claim 56 to a cell containing the target nucleic acid; thereby silencing the target nucleic acid sequence.

72. The method of claim 71, wherein the target nucleic acid comprises a CpG island.

73. The method of claim 71, wherein the target nucleic acid does not comprise a CpG island.

74. A method of treating Angelman syndrome, an infectious disease, a tau pathology, or a neurodegenerative disease in a subject in need thereof, the method comprising delivering to the subject an effective amount of a first polynucleotide encoding a fusion protein of claim 56; thereby treating Angelman syndrome, the infectious disease, the tau pathology, or the neurodegenerative disease.

75. The method of claim 74, wherein the infectious disease is a viral infectious disease.

76. The method of claim 75, wherein the viral infectious disease is a Flavivirus infectious disease.