(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2019 (24.10.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/204766 A1

(51) International Patent Classification:
C12Q 1/68 (2018.01) *C12N 15/117* (2010.01)
C12N 9/22 (2006.01) *C12N 15/85* (2006.01)
C12N 15/63 (2006.01)

(21) International Application Number:
PCT/US2019/028377

(22) International Filing Date:
19 April 2019 (19.04.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/660,023 19 April 2018 (19.04.2018) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Twelfth Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **CHEN, Jin**; Office of Innovation, Technology, & Alliances, Genentech Hall, 600 16th Street, Suite S272, San Francisco, CA 94143-1209 (US). **GILBERT, Luke**; Office of Innovation, Technology, & Alliances, Genentech Hall, 600 16th Street, Suite S272, San Francisco, CA 94143-1209 (US). **NUNEZ, James**; Office of Innovation, Technology, & Alliances, Genentech Hall, 600 16th Street, Suite S272, San Francisco, CA 94143-1209 (US). **WEISSMAN, Jonathan**; Office of Innovation, Technology, & Alliances, Genentech Hall, 600 16th Street, Suite S272, San Francisco, CA 94143-1209 (US).

(74) Agent: **GRIEFF, Edward D.** et al.; Mintz Levin Cohn Ferris Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

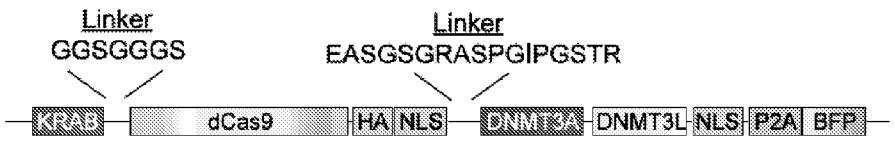
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: COMPOSITIONS AND METHODS FOR GENE EDITING

FIG. 1A



(57) Abstract: Provided herein are, *inter alia*, compositions and methods for manipulation of genomes of living organisms.

WO 2019/204766 A1

COMPOSITIONS AND METHODS FOR GENE EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to US Application No. 62/660,023 filed April 19, 2018, the disclosure of which is incorporated by reference herein in its 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 Department of Defense, Defense Advanced Research Projects Agency and grant no. R01 DA036858 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] Although considered a promising therapeutic approach for treatment of disease, genome editing carries inherent risks due to the 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.

[0004] Provided herein are solutions to these and other needs in the art.

BRIEF SUMMARY

[0005] In an aspect is provided a fusion protein including a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box (KRAB) domain and a DNA methyltransferase domain. In an aspect is provided a fusion protein of any one of SEQ ID NOS:1-15.

[0006] In an aspect is provided a nucleic acid sequence encoding the fusion protein as described herein, including embodiments and aspects thereof.

[0007] In an aspect is provided a complex including a fusion protein as described herein, including embodiments and aspects thereof, and 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.

[0008] In an aspect is provided a vector including the nucleic acid sequence of a fusion protein as described herein, including embodiments and aspects thereof.

[0009] 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.

[0010] In an aspect is provided a method 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 including (i) a DNA-targeting sequence that is complementary to the target nucleic acid sequence, and (ii) a binding sequence for the nuclease-deficient RNA-guide DNA endonuclease enzyme. Without intending to be bound by any theory, it is believed that the fusion protein 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. Thus, in aspects, the fusion protein 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.

[0011] 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, wherein the complex silences the target nucleic acid sequence in the cell. Without intending to be bound by any theory, it is believed that 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. Thus, in aspects, 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

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] FIG. 1A-1F describe engineering of an all-in-one protein for long-term gene silencing.

FIG. 1A is a schematic of an all-in-one protein (SEQ ID NO:1) of the disclosure that has the KRAB domain fused to the -N-terminus of dCas9 (SEQ ID NO:23), separated by a GGSGGGS (SEQ ID NO:17) linker, and Dnmt3A-Dnmt3L at the C-terminus of dCas9 (separated by a EASGSGRASPGIPGSTR (SEQ ID NO:19) linker). FIG. 1B provides schematics of dCas9-fused epigenetic modulators tested for permanent gene silencing. The dCas9-KRAB protein is adapted from Gilbert et al., Cell 2013 for CRISPR interference (CRISPRi) applications. The dCas9-Dnmt3A-Dnmt3L fusion is adapted from Stepper et al., Nucleic Acids Research, 2016. The inventors engineered a novel all-in-one protein that combines the KRAB domain (SEQ ID NO:16), dCas9 (D10A, H208A), Dnmt3A-Dnmt3L (SEQ ID NO:33; where Dnmt3A is SEQ ID NO:26 and Dnmt3L is SEQ ID NO:28) into one polypeptide. FIG. 1C provides schematics of a methylation-sensitive GFP reporter (adapted from Stelzer et al., Cell 2015) that was used to assess long-term silencing by the all-in-one protein. FIGS. 1D-1E provide a diagram and results of a hit-and-run experimental workflow in HEK293T cells expressing the GFP reporter shown in FIG. 3. FIG. 1D shows that plasmids were co-transfected into cells, one encoding the hit-and-run protein and the other plasmid encoding a sgRNA. FIG. 1E shows the results of the hit-and-run assay sorted for cells that were co-transfected with the all-in-one plasmid and sgRNA plasmid. FIG. 1F shows the results of the silencing of the GFP reporter is dependent on the sgRNA sequence.

[0013] FIGS. 2A-2F describe long-term silencing of endogenous genes. FIGS. 2A-2C are representative flow cytometry data shown taken 22 days post-transfection following gene (CD29, CD81, CD151) targeting for long term silencing using the all-in-one protein. Quadrant IV represents cells that have turned off the gene, indicated by the percentage of cells with the gene off (i.e., 45%, 66%, and 53%, respectively). FIG. 2D provides quantification of silencing of CD29, CD81, and CD151 with three different sgRNA. FIG. 2E provides quantification of silencing of two or three genes simultaneously to show that the all-in-protein can be multiplexed by co-delivery of sgRNAs targeting different genes. FIG. 2F provides a plot representing a time point taken 9 months post transfection of the all-in-one protein and sgRNA targeting the CLTA gene, signifying that the majority of cells have stably turned off the CLTA gene.

[0014] FIGS. 3A-3I describe long-term silencing of endogenous genes. FIGS. 3A-3C shows that harvested cells lost expression of CD29 (FIG. 3A), CD81 (FIG. 3B), and CD151 (FIG. 3C) thirty-six days post-transfection, as determined by their RNA expression profiles. FIGS. 3D-3F are volcano plots showing that the targeted genes CD29 (FIG. 3D), CD81 (FIG. 3E) and CD151 (FIG. 3F) is the only significant gene knocked down for each experiment, signifying high specificity of gene silencing. FIGS. 3G-3I provides quantification of transcript levels of CD151

(FIG. 3G), CD81 (GIF. 3H), and CD29 (FIG. 3I) showing more than 96% knockdown of each of the targeted genes.

[0015] FIGS. 4A-4H describe long-term gene silencing in different mammalian cell lines. FIGS. 4A-4F are flow cytometry plots showing BFP expression (which is fused to the all-in-one protein) in HeLa (cervical)(FIG. 4A), U2OS (bone)(FIG. 4B), and human induced pluripotent cells (iPSC)(FIG. 4C). FIGS. 4D-4F are the untransfected controls for FIGS. 4A-4C, respectively. FIG. 4G shows that stable silencing of endogenous genes in HeLa and U2OS cells, measured at 18 days post-transfection with the all-in-one protein, was achieved. In FIGS. 4A-4F, the x-axis is BFP (fused to all-in-one protein), and the y-axis is mCherry. FIG. 4H shows that gene silencing was detected 14 days post transfection by qPCR in AML12 mouse hepatocyte cell lines when targeting Pcsk9, Npc1, Spcs1 and Cd81.

[0016] FIG. 5 provides schematics of the fusion proteins p76, p90-p102, and p112 which correspond to SEQ ID NOS:1-15, respectively.

[0017] FIGS. 6A-6E describe gene silencing activities of all-in-one protein variants. FIGS. 6A-6-B shows the gene silencing results 18 days post-transfection of the fusion proteins of SEQ ID NOS:1-15 transfected into HEK293T cells for targeted silencing of the CLTA gene. The dCas9-KRAB and dCas9-Dnmt3A-Dnmt3L designs showed transient and lower efficiency of long term silencing. FIGS. 6C-6D provide a comparison of SEQ ID NO:1 (p76) and SEQ ID NO:15 (p112) for silencing the HIST2H2BE (H2B) endogenous gene (FIG. 6C) and a synthetic Snrpn-GFP reporter gene (FIG. 6D) stably expressed in HEK293T cells. FIG. 6E provides a plot of protein expression (dotted lines) of p76 and p112 over the 50 day time course to turn off the HIST2H2BE (H2B) gene. Protein levels were measured by flow cytometry detection of BFP, which is co-expressed with the all-in-one protein.

[0018] FIGS. 7A-7B provide Western blots of all-in-one-protein variants. FIG. 7A is a Western blot analysis of the all-in-one protein variants p76 and p90-p102 using an antibody against *Steptococcus pyogenes* Cas9. The top band represents full-length protein and smaller-sized bands represent proteolysis of the all-in-one protein. FIG. 7B is a Western blot analysis of all-in-one protein variants to detect free Dnmt3A that is cleaved from the fusion protein.

[0019] FIGS. 8A-8E describe pooled screen to determine optimal sgRNAs. FIG. 8A is a schematic of a pooled screen to determine the optimal sgRNAs that leads to long term gene silencing. FIGS. 8B-8E are flow cytometry histograms of the percent of cells undergoing gene silencing four weeks post-transfection. Four HEK293T cell lines were used, each with a different gene with a GFP tag, including CLTA (FIG. 8B), VIM (FIG. 8C), HIST2H2BE (H2B)

(FIG. 8D), and RAB11A (FIG. 8E).

[0020] FIGS. 9A-9D are maps of sgRNA functionality across the transcription start site of the targeted gene, including CLTA (FIG. 9A), H2B (FIG. 9B), RAB11 (FIG. 9C), and VIM (FIG. 9D). The transcription start site (TSS) and CpG island are annotated above each plot. Each dot represents one sgRNA and its efficacy in long term gene silencing is plotted as the log2 fold change in sgRNA abundance. Nucleosome occupancy (bottom plot) is plotted from MNase signal.

[0021] FIGS. 10A-10E describe functional sgRNAs for long term gene silencing. FIG. 10A is a workflow of a pooled screen in HEK293T cells to determine optimal sgRNA targeting positions for the all-in-one protein, adapted from a previous ricin tiling screen in K562 cells to determine optimal sgRNAs for dCas9-KRAB (Gilbert, Horlbeck et al., Cell 2014). FIGS. 10B-10E are representative plots showing growth phenotypes for four genes, including ARL1 (FIG. 10B), EIF6 (FIG. 10C), SMC3 (FIG. 10D), HEATR1 (FIG. 10E), from existing dCas9-KRAB/CRISPRi datasets in K562 cells (Gilbert, Horlbeck et al., 2014) and with the all-in-one protein (bottom plot). Each dot represents an sgRNA. The TSS and annotated CpG island are shown for each gene.

[0022] FIGS. 11A-11B provide a comparison of growth phenotypes and nucleosome positioning (from MNase signal) for VPS53 (FIG. 11A) and VPS54 (FIG. 11B) showing the location of functional sgRNAs at nucleosome-depleted regions.

[0023] FIGS. 12A-12C show the delivery of the all-in-one protein by mRNA expression FIG. 12A shows the in vitro transcription of two all-in-one variants (p102 and p112) show full length synthesis of each design. FIG. 12B provides a flow cytometry plot showing expression of p102 and p112 one day post-transfection of mRNA into HEK293T cells. FIG. 12C provides the time course of CLTA endogenous gene silencing in HEK293T cells after transfecting mRNA expressing the p102 and p112 all-in-one variants.

[0024] FIGS. 13A-13G describe controlled expression of the all-in-one protein by doxycycline induction. FIG. 13A provides flow cytometry plots showing induced expression of the all-in-one protein by addition of doxycycline in K562 cells that stably encode the all-in-one protein under a doxycycline-inducible promoter. The dotted line represents the baseline median BFP fluorescence without doxycycline administration. FIG. 13B provides a Western blot of cells to detect expression of the all-in-one protein before and after doxycycline treatment. FIGS. 13C-13F are flow cytometry plots of CD81 (FIGS. 13C-13D) and CD151 (FIGS. 13E-13F) knockdown 14 days post-doxycycline treatment of K562 cells. FIG. 13G shows the

quantification of CD81 and CD151 knockdown 14 days post-doxycycline treatment or without doxycycline treatment.

DETAILED DESCRIPTION

[0025] 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"). This fusion protein 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. In embodiments, the result is gene silencing that is inheritable across subsequent cell divisions. In embodiments, the fusion protein provided herein (and sgRNA) are only expressed transiently, bypassing the use of viral delivery methods to induce permanent silencing.

[0026] In embodiments, 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. In embodiments, an advantage of the fusion protein provided herein is that epigenetic editing is reversible and therefore inherently safer than genome editing. Thus, in embodiments, fusion protein provided herein is 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, in embodiments, fusion protein 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. In embodiments, fusion protein provided herein is useful in genome editing based therapeutics.

[0027] In embodiments, 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, the components are only expressed transiently in the host cell, thus reducing toxicity and off-target events.

[0028] 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.

[0029] Definitions

[0030] 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.

[0031] 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.

[0032] 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., R. 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.

[0033] 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. References identified herein are expressly incorporated herein by reference in their entireties unless otherwise indicated.

[0034] 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.

[0035] "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.

[0036] 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.

[0037] 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.

[0038] 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 amio acid on a protein or polypeptide through a covalent, non-covalent or other interaction.

[0039] 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 phosphothioate 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.

[0040] 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.

[0041] 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 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 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).

[0042] 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^{o}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.

[0043] 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.

[0044] 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.

[0045] 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.

[0046] 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.

[0047] 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.

[0048] 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.

[0049] 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 away from the transcription initiation site.

[0050] 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).

[0051] 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.

[0052] In embodiments, the polynucleotide (e.g., gRNA) is a single-stranded ribonucleic acid. In aspects, the polynucleotide (e.g., gRNA) is 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is from 10 to 30 nucleic acid residues in length. In aspects, the polynucleotide (e.g., gRNA) is 20 nucleic acid residues in length. In aspects, the length of the polynucleotide (e.g., gRNA) can be at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleic acid residues or sugar

residues in length. In aspects, the polynucleotide (e.g., gRNA) is from 5 to 50, 10 to 50, 15 to 50, 20 to 50, 25 to 50, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 5 to 75, 10 to 75, 15 to 75, 20 to 75, 25 to 75, 30 to 75, 35 to 75, 40 to 75, 45 to 75, 50 to 75, 55 to 75, 60 to 75, 65 to 75, 70 to 75, 5 to 100, 10 to 100, 15 to 100, 20 to 100, 25 to 100, 30 to 100, 35 to 100, 40 to 100, 45 to 100, 50 to 100, 55 to 100, 60 to 100, 65 to 100, 70 to 100, 75 to 100, 80 to 100, 85 to 100, 90 to 100, 95 to 100, or more residues in length. In aspects, the polynucleotide (e.g., gRNA) is from 10 to 15, 10 to 20, 10 to 30, 10 to 40, or 10 to 50 residues in length.

[0053] 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.

[0054] 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, likewise, may be referred to by their commonly accepted single-letter codes.

[0055] 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.

[0056] "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.

[0057] 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)).

[0058] "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.

[0059] 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 compliment 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.

[0060] 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.

[0061] 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.

[0062] 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.

[0063] 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), all of which are incorporated herein by reference in their entirety. 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.

[0064] 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, Dnmt3B, and Dnmt3L. 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 metyltransferases. 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 is 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. A description of Dnmt3A-3L domain structure and use may be found, 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, which are incorporated herein by reference in their entirety and for all purposes.

[0065] 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.

[0066] 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 enzyme or variants or homologs thereof that maintain Dnmt3L 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 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 (e.g., a 50, 100, 150 or 200 continuous amino acid portion) 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.

[0067] 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 97% 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.

[0068] 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).

[0069] 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.

[0070] 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.

[0071] 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 eukaroytic 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.

[0072] 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.

[0073] The terms "transfection", "transduction", "transfecting" or "transducing" can be used interchangeably and are defined as a process of introducing a nucleic acid molecule 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, magnetifection 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.

[0074] 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 peptide linker includes about 2 to about 13 amino acid residues. In aspects, the peptide linker includes about 2 to about 12 amino acid residues. In aspects, the peptide linker includes about 2 to about 11 amino acid residues. In aspects, the peptide linker includes about 2 to about 10 amino acid residues. In aspects, the peptide linker includes about 2 to about 9 amino acid residues. In aspects, the peptide linker includes about 2 to about 8 amino acid residues. In aspects, the peptide linker includes about 2 to about 7 amino acid residues. In aspects, the peptide linker includes about 2 to about 6 amino acid residues. In aspects, the peptide linker includes about 2 to about 5 amino acid residues. In aspects, the peptide linker includes about 2 to about 4 amino acid residues. In aspects, the peptide linker includes about 2 to about 3 amino acid residues. In aspects, the peptide linker includes about 3 to about 19 amino acid residues. In aspects, the peptide linker includes about 3 to about 18 amino acid residues. In aspects, the peptide linker includes about 3 to about 17 amino acid residues. In aspects, the peptide linker includes about 3 to about 16 amino acid residues. In aspects, the peptide linker includes about 3 to about 15 amino acid residues. In aspects, the peptide linker includes about 3 to about 14 amino acid residues. In aspects, the peptide linker includes about 3 to about 13 amino acid residues. In aspects, the peptide linker includes about 3 to about 12 amino acid residues. In aspects, the peptide linker includes about 3 to about 11 amino acid residues. In aspects, the peptide linker includes about 3 to about 10 amino acid residues. In aspects, the peptide linker includes about 3 to about 9 amino acid residues. In aspects, the peptide linker includes about 3 to

about 8 amino acid residues. In aspects, the peptide linker includes about 3 to about 7 amino acid residues. In aspects, the peptide linker includes about 3 to about 6 amino acid residues. In aspects, the peptide linker includes about 3 to about 5 amino acid residues. In aspects, the peptide linker includes about 3 to about 4 amino acid residues. In aspects, the peptide linker includes about 10 to about 20 amino acid residues. In aspects, the peptide linker includes about 15 to about 20 amino acid residues. In aspects, the peptide linker includes about 2 amino acid residues. In aspects, the peptide linker includes about 3 amino acid residues. In aspects, the peptide linker includes about 4 amino acid residues. In aspects, the peptide linker includes about 5 amino acid residues. In aspects, the peptide linker includes about 6 amino acid residues. In aspects, the peptide linker includes about 7 amino acid residues. In aspects, the peptide linker includes about 8 amino acid residues. In aspects, the peptide linker includes about 9 amino acid residues. In aspects, the peptide linker includes about 10 amino acid residues. In aspects, the peptide linker includes about 11 amino acid residues. In aspects, the peptide linker includes about 12 amino acid residues. In aspects, the peptide linker includes about 13 amino acid residues. In aspects, the peptide linker includes about 14 amino acid residues. In aspects, the peptide linker includes about 15 amino acid residues. In aspects, the peptide linker includes 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.

[0075] 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.

[0076] 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.

[0077] 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), which is incorporated herein by reference in its entirety and for all purposes. 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.

[0078] 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, ^{99m}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, $^{154\text{-}1581}Gd$, ^{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, monochrystalline iron oxide nanoparticles, monochrystalline 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 gas(es), perfluorcarbon, nitrogen, octafluoropropane, perflexane lipid microsphere, perflutren, etc.), 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.

[0079] 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 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. In aspects, the detectable agent 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.

[0080] 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, ^{99m}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, $^{154-1581}Gd$, ^{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.

[0081] "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.

[0082] 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).

[0083] 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 nuclei 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.

[0084] 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.

[0085] Fusion Proteins

[0086] 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 four proteins (e.g., catalytically inactive Cas9 (e.g., dCas9), a KRAB domain, Dnmt3A and Dnmt3L) which can be expressed transiently in cells. The fusion protein can be directed to a specific site in a mammalian genome using a polynucleotide complementary to a target nucleic acid sequence (e.g., DNA sequence) and that further includes a sequence (i.e., binding sequence) capable of binding the fusion protein. 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.

[0087] In embodiments, the fustion 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 fustion protein further comprises one or more peptide linkers. In aspects, the fusion protein further comprises one or more detectable tags. In aspects, the fusion ptoein 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 liner 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.

[0088] 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; 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).

[0089] In embodiments, the fusion protein comprises the structure: A-L_1-B-L_2-C, where A comprises a nuclease-deficient RNA-guided DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, L_1 is a covalent bond or a peptide linker, and L_2 is a covalent bond or a peptide linker; and where A is at the N-terminus and C 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, L_1 is a covalent bond, a peptide linker, a detectable tag, anuclear localization sequence, or a combination thereof. In aspects, L_2 is a covalent bond, a peptide linker, a detectable tag, a nuclear localization sequence, or a combination thereof.

[0090] In embodiments, the fusion protein comprises the structure: B-L_1-A-L_2-C, where A comprise a nuclease-deficient RNA-guided DNA endonuclease enzyme; B comprises a KRAB domain, C comprises a DNA methyltransferase domain, L_1 is a covalent bond or a peptide linker, and L_2 is a covalent bond or a peptide linker; and where B is at the N-terminus and C is at the C-terminus. In aspects, L_1 is a peptide linker. In aspects, L_1 is a covalent bond. In aspects, L_2 is a peptide linker. In aspects, L_2 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, L_1 is a covalent bond, a peptide linker, a detectable tag, or a combination thereof. In aspects, L_2 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. Exemplary fusion proteins comprising the structure: B-L_1-A-L_2-C include p76, p90, p91, p92, p93, p94, p95, p96, p97, p98, p99, p100, p101, and p102 (FIG. 5)

[0091] In embodiments, the fusion protein comprises the structure: B-L_3-A-L_4-C-L_5-D; where A comprises a nuclease-deficient RNA-guided 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, L_3 comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L_4 comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L_5 is absent or L_5 comprises a covalent bond or a peptide linker; and where B is at the N-terminus and D is at the C-terminus. In aspects, L_3 is a peptide linker. In aspects, L_3 is a covalent bond. In aspects, L_3 comprises a a peptide linker and a detectable tag. In aspects, L_3 comprises a detectable tag. In aspects, L_4 is a peptide linker. In aspects, L_4 comprises a peptide linker and a detectable tag. In aspects, L_4 is a covalent bond. In aspects, L_4 comprises a detectable tag. In aspects, L_5 is a peptide linker. In aspects, L_5 is a covalent bond. In aspects, D comprises one or a plurality of detectable tags. In aspects, D comprieses 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 L_5 and D are absent. When L_3, L_4, L_5, 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. Exemplary fusion proteins comprising the structure: B-L_3-A-L_4-C-L_5-D include p76, p90, p91, p92, p93, p94, p95, p96, p97, p98, p99, p100, p101, and p102, as shown in FIG. 5.

[0092] In embodiments, the fusion protein comprises the structure: C-L_3-A-L_4-B-L_5-D, where A comprises a nuclease-deficient RNA-guided 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, L_3 comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L_4 comprises a covalent bond, a peptide linker, a detectable tag, or a combination of two or more thereof, L_5 is absent or L_5 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, L_3 is a

peptide linker. In aspects, L_3 is a covalent bond. In aspects, L_3 comprises a detectable tag. In aspects, L_3 comprises a peptide linker and a detectable tag. In aspects, L_4 a peptide linker. In aspects, L_4 is a covalent bond. In aspects, L_4 comprises a detectable tag. In aspects, L_4 comprises a peptide linker and a detectable tag. In aspects, L_5 a peptide linker. In aspects, L_5 is a covalent bond. In aspects, D comprises one or a plurality of detectable tags. In aspects, D comprieses 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 L_5 and D are absent. When L_3, L_4, L_5, 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. Exemplary fusion proteins comprising the structure: C-L_3-A-L_4-B-L_5-D include p112, as shown in FIG. 5.

[0093] 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 dCas9, ddCpf1, a nuclease-deficient Cas9 variant, or a nuclease-deficient Class II CRISPR endonuclease.

[0094] 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.

[0095] 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.

[0096] 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.

[0097] 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.

[0098] 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.

[0099] 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.

[0100] 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, which are incorporated herein by reference in their entirety and for all purposes. Exemplary Cas9 variants are listed in the Table 4 below.

[0101] Table 4

Cas9 Variants	PAM domains	References
Strep pyogenes (Sp) Cas9	NGG	Hsu et al. 2014 Cell
Staph aureus (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
AsCpf1	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)

[0102] 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.

[0103] 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 and Dnmt3L. In aspects, the Dnmt3A and the Dnmt3L are covalently linked. In aspects, the Dnmt3A is covalently linked to the 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.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] 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.

[0108] In embodiments, the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain.

[0109] 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.

[0110] 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.

[0111] 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.

[0112] 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.

[0113] In embodiments, the fusion protein includes an amino acid sequence having at least 75% 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% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino

acid sequence having at least 75% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:13. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 75% sequence identity to SEQ ID NO:15.

[0114] In embodiments, the fusion protein includes an amino acid sequence having at least 80% 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 80% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:13. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 80% sequence identity to SEQ ID NO:15.

[0115] In embodiments, the fusion protein includes an amino acid sequence having at least 85% 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 85% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:13. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 85% sequence identity to SEQ ID NO:15.

[0116] In embodiments, the fusion protein includes an amino acid sequence having at least 90% 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 90% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an

amino acid sequence having at least 90% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:13. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 90% sequence identity to SEQ ID NO:15.

[0117] In embodiments, the fusion protein includes an amino acid sequence having at least 95% 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 95% sequence identity to SEQ ID NO:1. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:2. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:3. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:4. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:5. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:6. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:7. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:8. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:9. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:10. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:11. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:12. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:13. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:14. In aspects, the fusion protein includes an amino acid sequence having at least 95% sequence identity to SEQ ID NO:15.

[0118] Complexes

[0119] 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. 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 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).

[0120] 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, 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.

[0121] A "target polynucleotide sequence" as provided herein is a nucleic acid sequence preent 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 CRISPR complex. Full

complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a 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.

[0122] 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.

[0123] 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 of non-diseased) methylation levels.

[0124] In embodiments, the target polynucleotide sequence is within about 3000 base pairs (bp) flanking a transcription start site. In aspects, the target polynucleotide sequence 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, or 100 base pairs (bp) flanking a transcription start site.

[0125] 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.

[0126] In embodiments, exemplary target polynucleotide sequences include those described in Tables 1 and 2. In aspects, 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.

[0127] 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:37. 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:39. 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:41. 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:43. In aspects, the target polynucleotide sequence is SEQ ID NO:45. 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:47. 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:49. 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:51. 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:53. 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:55. 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:57. 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:59. 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:61. 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:63. 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:65. 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:67. 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: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.

[0128] In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:37. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:39. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:41. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:43. In aspects, the target polynucleotide sequence is SEQ ID NO:45. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:47. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:49. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:51. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:53. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:55. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:57. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:59. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:61. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:63. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:65. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:67. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:69. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:71. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:73. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:75. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:77. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:79. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:81. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:83. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:85. In aspects, the

target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:87. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:89. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:91. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:93. In aspects, the target polynucleotide sequence has at least 90% sequence identity to SEQ ID NO:95.

[0129] In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:37. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:39. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:41. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:43. In aspects, the target polynucleotide sequence is SEQ ID NO:45. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:47. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:49. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:51. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:53. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:55. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:57. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:59. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:61. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:63. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:65. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:67. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:69. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:71. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:73. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:75. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:77. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:79. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:81. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:83. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:85. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:87. In aspects,

the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:89. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:91. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:93. In aspects, the target polynucleotide sequence has at least 95% sequence identity to SEQ ID NO:95.

[0130] 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 30-100 nt, 35-50 nt, 37-47 nt, or 42 nt in length.

[0131] 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).

[0132] The binding sequence (e.g., Cas9-binding sequence) can have a length of from 10 nucleotides to 100 nucleotides, *e.g.*, from 10 nucleotides (nt) to 20 nt, from 20 nt to 30 nt, from 30 nt to 40 nt, from 40 nt to 50 nt, from 50 nt to 60 nt, from 60 nt to 70 nt, from 70 nt to 80 nt, from 80 nt to 90 nt, or from 90 nt to 100 nt. In aspects, the binding sequence has a length of from 15 nucleotides (nt) to 80 nt. In aspects, the binding sequence has a length of from 15 nt to 50 nt. In aspects, the binding sequence has a length of from 15 nt to 40 nt. In aspects, the binding sequence has a length of from 15 nt to 30 nt. In aspects, the binding sequence has a length of from 37 nt to 47 nt (*e.g.*, 42 nt). In aspects, the binding sequence has a length of from 15 nt to 25 nt.

[0133] The dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) can have a length from 6 base pairs (bp) to 50 bp. For example, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) can have a length from 6 bp to 40 bp, from 6 bp to 30 bp, from 6

bp to 25 bp, from 6 bp to 20 bp, from 6 bp to 15 bp, from 8 bp to 40 bp, from 8 bp to 30 bp, from 8 bp to 25 bp, from 8 bp to 20 bp or from 8 bp to 15 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 8 bp to 10 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 10 bp to 15 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 15 bp to 18 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 18 bp to 20 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 20 bp to 25 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 25 bp to 30 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 30 bp to 35 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 35 bp to 40 bp. In aspects, the dsRNA duplex of the binding sequence (e.g., Cas9-binding sequence) has a length from 40 bp to 50 bp.

[0134] In embodiments, the exemplary polynucleotide that forms a complex with a fusion protein described herein includes those described in Tables 1 and 2 as sgRNA. In aspects, 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.

[0135] Nucleic Acids and Vectors

[0136] 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, the messenger RNA is messenger RNP. 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 SEQ ID NO:1. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:2. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:3. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:4. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:5. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:6. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:7. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:8. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:9. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:10. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:11. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:12. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:13. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:14. In aspects, the nucleic acid sequence encodes for the fusion protein of SEQ ID NO:15.

[0137] 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 SEQ ID NO:1. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:2. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:3.In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:4. In aspects, the vector comprises a nucleic

acid sequence that encodes for the fusion protein of SEQ ID NO:5. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:6. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:7. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:8. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:9. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:10. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:11. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:12. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:13. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:14. In aspects, the vector comprises a nucleic acid sequence that encodes for the fusion protein of SEQ ID NO:15.

[0138] 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. Thus, one or more vectors may include all necessary components for preforming epigenome editing.

[0139] Cells

[0140] 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.

[0141] Methods

[0142] It is contemplated that the compositions described herein may be used for epigenome

editing, and more particularly epigenome editing resulting in the repression or silencing of target nucleic acid sequences (e.g., genes). Without intending to be bound by any theory, silencing may result from methylation of and/or the introduction of repressive chromatin markers (e.g., mono-, di-, or tri-methylation of specific histones (e.g., H3K9, H3K27), deacetylation, acetylation, phosphorylation, ubiquitination) on chromatin containing a target nucleic acid sequence. Without intending to be bound by any theory, the method can be used to change epigenetic state by, for example, closing chromatin via methylation or introducing repressive chromatin markers on chromatin containing the target nuclei acid sequence (e.g., gene). Without intending to be bound by any theory, it is contemplated that the Dnmt3A-3L fusion functions to add methyl marks at CG DNA sites found in CpG islands and the KRAB domain recruits epigenetic factors that modify the histones by introducing repressive marks. Without intending to be bound by any theory, DNA is methylated at the C nucleotide of CG sequences found in CpG islands (i.e., adding methyl marks at the C nucleotide of CG DNA sites found in CpG islands).

[0143] In an aspect is provided a method 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 including: (i) a DNA-targeting sequence that is complementary to the target nucleic acid sequence; and (ii) a binding sequence for the nuclease-deficient RNA-guide DNA endonuclease enzyme. Without intending to be bound by any theory, the fusion protein 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. 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 (i.e., adding methyl marks at the C nucleotide of CG DNA sites found in CpG islands). In aspects, the sequnce that is withinn about 3000 base pairs of the target nucleic acid sequence is is methylated. In aspects, the squence 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, or 100 base pairs of the target nucleic acid sequence is methylated.

[0144] 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).

[0145] 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.

[0146] 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.

[0147] 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.

[0148] 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.

[0149] In embodiments, the cell is a mammalian cell.

[0150] 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.

[0151] In aspects, the complex is delivered into the cell via any methods known in the art, for example, via ribonucleoprotein (RNP) delivery.

[0152] Embodiments N1-N41

[0153] Embodiment N1. A fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain.

[0154] Embodiment N2. The fusion protein of Embodiment N1, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, ddCpf1, a nuclease-deficient Cas9

variant, or a nuclease-deficient Class II CRISPR endonuclease.

[0155] Embodiment N3. The fusion protein of Embodiment N1 or N2, wherein the DNA methyltransferase domain is a Dnmt3A-3L domain.

[0156] Embodiment N4. The fusion protein of Embodiment N1, 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

[0157] Embodiment N5. The fusion protein of Embodiment N4, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain

[0158] Embodiment N6. The fusion protein of Embodiment N5, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain

[0159] Embodiment N7. The fusion protein of Embodiment N6, wherein the peptide linker is a XTEN linker.

[0160] Embodiment N8. The fusion protein of Embodiment N1, 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

[0161] Embodiment N9. The fusion protein of Embodiment N8, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain.

[0162] Embodiment N10. The fusion protein of Embodiment N9, wherein the dCas9 is covalently linked to the Dnmt3A-3L domain via a peptide linker and wherein the Krüppel associated box domain is covalently linked to the dCas9 via a peptide linker.

[0163] Embodiment N11. The fusion protein of Embodiment N10, wherein the peptide linker is a XTEN linker.

[0164] Embodiment N12. The fusion protein of any one of Embodiments N1 to N3, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the Krüppel associated box domain via a peptide linker.

[0165] Embodiment N13. The fusion protein of any one of Embodiments N1 to N3, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0166] Embodiment N14. The fusion protein of any one of Embodiments N1 to N3, wherein the Krüppel associated box domain is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0167] Embodiment N15. The fusion protein of any one of Embodiments N12 to N14, wherein the peptide linker is a XTEN linker.

[0168] Embodiment N16. The fusion protein of Embodiment N15, wherein the XTEN linker comprises about 16 to 80 amino acid residues.

[0169] Embodiment N17. The fusion protein of any one of Embodiments N1 to N16, further comprising a nuclear localization signal peptide.

[0170] Embodiment N18. The fusion protein of Embodiment N1, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 12, 13, 14, or 15.

[0171] Embodiment N19. A nucleic acid sequence encoding the fusion protein of any one of Embodiments N1 to N18.

[0172] Embodiment N20. The nucleic acid sequence of Embodiment N19, wherein the nucleic acid sequence is messenger RNA.

[0173] Embodiment N21. A complex comprising: (i) a fusion protein of any one of Embodiments N1 to N18; and (ii) a polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) 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.

[0174] Embodiment N22. The complex of Embodiment N21, wherein the target polynucleotide sequence is part of a gene.

[0175] Embodiment N23. The complex of Embodiment N21, wherein the target polynucleotide sequence is part of a transcriptional regulatory sequence.

[0176] Embodiment N24. The complex of Embodiment N21, wherein the target polynucleotide sequence is part of a promoter, enhancer, or silencer.

[0177] Embodiment N25. The complex of Embodiment N21, wherein the target polynucleotide sequence is within about 3000 bp flanking a transcription start site.

[0178] Embodiment N26. A vector comprising the nucleic acid sequence of Embodiment N19

or N20.

[0179] Embodiment N27. The vector of Embodiment N26, further comprising a polynucleotide, wherein the polynucleotide comprises: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for the nuclease-deficient RNA-guided DNA endonuclease enzyme.

[0180] Embodiment N28. A cell comprising the fusion protein of any one of Embodiments N1 to N18; the nucleic acid of Embodiment N19 or N20; the complex of any one of Embodiments N21 to N25, or the vector of Embodiment N26 or N27.

[0181] Embodiment N29. The cell of Embodiment N28, wherein the cell is a eukaryotic cell.

[0182] Embodiment N30. The cell of Embodiment N28, wherein the cell is a mammalian cell.

[0183] Embodiment N31. A method of silencing a target nucleic acid sequence in a cell, comprising: (i) delivering a first polynucleotide encoding a fusion protein of any one of Embodiments N1 to N18 to a cell containing the target nucleic acid; and (ii) delivering to the cell a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to the target nucleic acid sequence; and (b) a binding sequence for the nuclease-deficient RNA-guide DNA endonuclease enzyme

[0184] Embodiment N32. The method of Embodiment N31, wherein the fusion protein 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.

[0185] Embodiment N33. The method of Embodiment N31 or N32, wherein the first polynucleotide is contained within a first vector.

[0186] Embodiment N34. The method of any one of Embodiments N31 to N33, wherein the first polynucleotide is contained within a second vector.

[0187] Embodiment N35. The method of Embodiment N34, wherein the first vector and the second vector are the same.

[0188] Embodiment N36. The method of Embodiment N34, wherein the first vector is different from the second vector.

[0189] Embodiment N37. The method of Embodiment N31, wherein the cell is a mammalian cell.

[0190] Embodiment N38. The method of Embodiment N31, wherein the method has a

specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

[0191] Embodiment N39. A method of silencing a target nucleic acid sequence in a cell, the method comprising delivering the complex of any one of Embodiments N21 to N25 to a cell containing the target nucleic acid.

[0192] Embodiment N40. The method of Embodiment N39, wherein 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.

[0193] Embodiment N41. The method of Embodiment N39 or N40, wherein the cell is a mammalian cell.

[0194] Embodiment N42. The method of any one of Embodiments N39 to N41, wherein the method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

[0195] Embodiments 1 to 36

[0196] Embodiment 1. A fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box (KRAB) domain, and a DNA methyltransferase domain.

[0197] Embodiment 2. The fusion protein of Embodiment 1, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, ddCpf1, a nuclease-deficient Cas9 variant, or a nuclease-deficient Class II CRISPR endonuclease.

[0198] Embodiment 3. The fusion protein of Embodiment 1 or 2, wherein the DNA methyltransferase domain is a Dnmt3A-3L domain.

[0199] Embodiment 4. The fusion protein of any one of Embodiments 1 to 3, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the KRAB domain via a peptide linker.

[0200] Embodiment 5. The fusion protein of any one of Embodiments 1 to 4, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0201] Embodiment 6. The fusion protein of any one of Embodiments 1 to 5, wherein the KRAB domain is covalently linked to the DNA methyltransferase domain via a peptide linker.

[0202] Embodiment 7. The fusion protein of any one of Embodiments 4 to 6, wherein the

peptide linker is a XTEN linker.

[0203] Embodiment 8. The fusion protein of Embodiment 7, wherein the XTEN linker comprises about 16 to 80 amino acid residues.

[0204] Embodiment 9. The fusion protein of any one of Embodiments 1 to 8, further comprising a nuclear localization signal peptide.

[0205] Embodiment 10. The fusion protein of any one of Embodiments 1 to 9, wherein 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.

[0206] Embodiment 11. The fusion protein of any one of Embodiments 1 to 10, wherein the nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and the DNA methyltransferase domain is a Dnmt3A-3L domain.

[0207] Embodiment 12. The fusion protein of Embodiment 11, wherein 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.

[0208] Embodiment 13. The fusion protein of Embodiment 12, wherein the peptide linker is a XTEN linker.

[0209] Embodiment 14. The fusion protein of any one of Embodiments 1 to 13, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 12, 13, 14, or 15.

[0210] Embodiment 15. A nucleic acid sequence encoding the fusion protein of any one of Embodiments 1 to 14.

[0211] Embodiment 16. A complex comprising: (i) a fusion protein of any one of Embodiments 1 to 18; and (ii) a polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) 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.

[0212] Embodiment 17. The complex of Embodiment 21, wherein the target polynucleotide sequence is part of a gene.

[0213] Embodiment 18. The complex of Embodiment 21, wherein the target polynucleotide sequence is part of a transcriptional regulatory sequence.

[0214] Embodiment 19. The complex of Embodiment 21, wherein the target polynucleotide sequence is part of a promoter, enhancer, or silencer.

[0215] Embodiment 20. The complex of Embodiment 21, wherein the target polynucleotide sequence is a hypomethylated nucleic acid sequence.

[0216] Embodiment 21. The complex of Embodiment 21, wherein the target polynucleotide sequence is within about 3000 bp flanking a transcription start site.

[0217] Embodiment 22. A vector comprising the nucleic acid sequence of Embodiment 19.

[0218] Embodiment 23. The vector of Embodiment 26, further comprising a polynucleotide, wherein the polynucleotide comprises: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for the nuclease-deficient RNA-guided DNA endonuclease enzyme.

[0219] Embodiment 24. A cell comprising the fusion protein of any one of Embodiments 1 to 14; the nucleic acid of Embodiment 15; the complex of any one of Embodiments 16 to 21; or the vector of Embodiment 22 or 23.

[0220] Embodiment 25. The cell of Embodiment 28, wherein the cell is a eukaryotic cell.

[0221] Embodiment 26. The cell of Embodiment 28, wherein the cell is a mammalian cell.

[0222] Embodiment 27. A method of silencing a target nucleic acid sequence in a cell, comprising: (i) delivering a first polynucleotide encoding a fusion protein according to any one of Embodiments 1 to 14 to a cell containing the target nucleic acid; and (ii) delivering to the cell a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to the target nucleic acid sequence; and (b) a binding sequence for the nuclease-deficient RNA-guide DNA endonuclease enzyme, wherein the fusion protein 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

[0223] Embodiment 28. The method of Embodiment 27, wherein the first polynucleotide is contained within a first vector.

[0224] Embodiment 29. The method of Embodiment 27, wherein the first polynucleotide is contained within a second vector.

[0225] Embodiment 30. The method of Embodiment 28 or 29, wherein the first vector and the second vector are the same.

[0226] Embodiment 31. The method of Embodiment 28 or 29, wherein the first vector is different from the second vector.

[0227] Embodiment 32. The method of any one of Embodiments 27 to 31, wherein the cell is a mammalian cell.

[0228] Embodiment 33. The method of any one of Embodiments 27 to 32, wherein the method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

[0229] Embodiment 34. A method of silencing a target nucleic acid sequence in a cell, the method comprising delivering the complex of any one of Embodiments 16 to 20 to a cell containing the target nucleic acid, wherein the complex silences the target nucleic acid sequence in the cell by: (i) methylating a chromatin containing the target nucleic acid sequence, (ii) introducing repressive chromatin marks to a chromatin containing the target nucleic acid sequence, or (iii) methylating a chromatin containing the target nucleic acid sequence and introducing repressive chromatin marks to a chromatin containing the target nucleic acid sequence.

[0230] Embodiment 35. The method of Embodiment 34, wherein the cell is a mammalian cell.

[0231] Embodiment 36. The method of Embodiment 34 or 35, wherein the method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

EXAMPLES

[0232] Embodiments and aspects herein are further illustrated by the following examples. The examples are merely intended to illustrate embodiments and aspets, and are not to be construed to limit the scope herein.

[0233] Example 1

[0234] dCas9-fused epigenetic modulators tested for permanent gene silencing. The initial version (V1, p76 (SEQ ID NO:1)) of the all-in-one protein (FIG. 1A) has the KRAB domain fused to the -N-terminus of dCas9 (SEQ ID NO:23), separated by a GGSGGGS (SEQ ID NO:17) linker, and Dnmt3A-Dnmt3L at the C-terminus of dCas9 (separated by a EASGSGRASPGIPGSTR (SEQ ID NO:19) linker). Another all-in-one proteins that combined the KRAB domain (SEQ ID NO:16), dCas9 (D10A, H208A), Dnmt3A-Dnmt3L (SEQ ID NO:33; where SEQ ID NO:26 is Dnmt3A and SEQ ID NO:28 is Dnmt3L) into one polypeptide (FIG. 1B). With reference to FIG. 1B, the dCas9-KRAB protein was adapted from Gilbert et al., Cell 2013 for CRISPR interference (CRISPRi) applications, and the dCas9-Dnmt3A-Dnmt3L fusion was adapted from Stepper et al., Nucleic Acids Research, 2016.

[0235] The activity of the V1 epigenetic editor was tested in HEK293T cells using a DNA-methylation sensitive GFP reporter (adapted from Stelzer et al., Cell 2015) to assess long-term silencing by the all-in-one protein (FIG. 1C). A ubiquitous chromatin opening element (UCOE) was added upstream of the GAPDH CpG island (CGI) to prevent background silencing of the lentiviral vector in mammalian cells. The *gfp* gene is turned off when the GAPDH CGI is methylated. A, B and C denote positions we encoded single guide RNAs (sgRNA) to target in the promoter. These targeted sequences and corresponding sgRNA sequences are listed in the Table 1 below. Two plasmids were co-transfected into cells, one encoding the hit-and-run protein and the other plasmid encoding a sgRNA (FIG. 1D). Two days post-transfection, cells that express the hit-and-run protein and sgRNA-expressing vector are sorted. GFP fluorescence is assessed over time by flow cytometry. A population of cells undergoing long-term silencing of the GFP reporter was observed when the all-in-one protein is expressed with sgRNAs (FIG. 1E). The number of cells undergoing long-term silencing was higher than dCas9-Dnmt3A-Dnmt3L (lacking the KRAB domain).

[0236] Table 1

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
A (JKNg156)	ACTGCGGAAATTTGAGCGT (SEQ ID NO: 37)	ACGCTCAAATTTCCGCAGT (SEQ ID NO: 38)
B (JKNg158)	AGGCAATGGCTGCACATGC (SEQ ID NO: 39)	GCATGTGCAGCCATTGCCT (SEQ ID NO: 40)
C (JKNg160)	GACGCTTGGTTCTGAGGAG (SEQ ID NO: 41)	CTCCTCAGAACCAAGCGTC (SEQ ID NO: 42)

[0237] Silencing of the GFP reporter is dependent on the sgRNA sequence, with guide C resulting in the highest level of silencing among the three sgRNA sequences tested. Pooling the sgRNAs encoding different sequences did not have a significant change in gene silencing.

[0238] Example 2

[0239] Three genes (CD29, CD81, CD151) were targeted for long term silencing using the hit-and-run fusion protein. All three proteins are cell surface-localized and knockdown was assessed by cell surface antibody staining of cells, followed by flow cytometry. Representative flow cytometry data are shown in FIGS. 2A-2C taken 22 days post-transfection. Quadrant IV represents cells that have turned off the gene, indicated by the percentage of cells with the gene off. The lack of cells in Quadrants I and II signify that the hit-and-run protein (marked by BFP) is no longer present in the cells. FIG. 2D provides quantification of silencing of CD29, CD81 and CD151 with three different sgRNA sequences or a pool of all three sgRNAs. The targeted

DNA sequences and their sgRNAs used in this experiment are summarized in Table 2.

[0240] Table 2

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
CD29, sgRNA-A	TCCGGAAACGCATTCCTCT (SEQ ID NO: 43)	AGAGGAATGCGTTTCCGGA (SEQ ID NO: 44)
CD29, sgRNA-B	CCGCGTCAGCCCGGCCCGG (SEQ ID NO: 45)	CCGGGCCGGGCTGACGCGG (SEQ ID NO: 46)
CD29, sgRNA-C	CGACTCCCGCTGGGCCTCT (SEQ ID NO: 47)	AGAGGCCCAGCGGGAGTCG (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)	GGCCTGGCAGGATGCGCGG (SEQ ID NO: 52)
CD81, sgRNA-C	CCAACTTGGCGCGTTTCGG (SEQ ID NO:53)	CCGAAACGCGCCAAGTTGG (SEQ ID NO: 54)
CD151, sgRNA-A	ACCACGCGTCCGAGTCCGG (SEQ ID NO: 55)	CCGGACTCGGACGCGTGGT (SEQ ID NO: 56)
CD151, sgRNA-B	TGCTCATTGTCCCTGGACA (SEQ ID NO: 57)	TGTCCAGGGACAATGAGCA (SEQ ID NO: 58)
CD151, sgRNA-C	GGACACCCTGCTCATTGTC (SEQ ID NO: 59)	GACAATGAGCAGGGTGTCC (SEQ ID NO: 60)

[0241] Two or three genes were simultaneously targeted to show that the all-in-protein can be multiplexed by co-delivery of sgRNAs targeting different genes. NT sgRNA refers to non-targeting sgRNA control. The results are shown in FIG. 2E.

[0242] Gene silencing of cells that started as a single clone were followed and it was observed the majority of cells have maintained the targeted CLTA gene off (37 out of 39 clones). The plot in FIG. 2F represents a time point taken 9 months post transfection of the all-in-one protein and sgRNA targeting the CLTA gene.

[0243] The system described herein can target any genes in the mammalian genomes, especially those that contain CpG islands at the gene promoter. The Dnmt3A-Dnmt3L canonically targets CpG dinucleotides. Examples of genes that can be targeted include, but are not limited to, CXCR4, CD4, CD8, CD45, PD-1, CLTA-4, TGFBR, TCRa, TCRb, B2M.

[0244] Example 3

[0245] Cells were harvested that lost expression of ITGB1 (CD29), CD81 and CD151 thirty-six days post-transfection and analyzed their RNA expression profiles were analyzed. As shown in FIGS. 3A-3C, successful knockdown of the targeted genes was detected compared to the non-targeting sgRNA control. FIGS. 3D-3F are volcano plots show that the targeted gene is the only significant gene knocked down for each experiment, signifying high specificity of gene

silencing. FIGS. 3G-3I are the quantification of transcript levels showing greater than 96% knockdown of the targeted gene.

[0246] Example 4

[0247] The all-in-one protein can be transfected and expressed in HeLa (cervical), U2OS (bone) and human induced pluripotent stem cells (iPSC). Flow cytometry plots in FIGS. 4A-4F show BFP expression, which is fused to the protein. Three endogenous genes in HeLa and U2OS cells (i.e., CD29, CD81, and CD151) were targeted. As shown in FIG. 4G, stable silencing, measured at 18 days post-transfection, was detected. Gene silencing in AML12 mouse hepatocyte cell lines was detected when targeting Pcsk9, Npc1, Spcs1 and Cd81. Silencing was detected by qPCR, measured 14 days post transfection, as shown in FIG. 4H. The sgRNA sequences used in this experiment are summarized in Table 3

[0248] Table 3

Name	Targeted sequence (5' to 3')	sgRNA sequence (5' to 3')
Pcsk9 sgRNA-1	TCCGGAAACGCATTCCTCT (SEQ ID NO:43)	AGAGGAATGCGTTTCCGGA (SEQ ID NO:44)
Pcsk9 sgRNA-2	ACCGGCAGCCTGCGCGTCC (SEQ ID NO: 61)	GGACGCGCAGGCTGCCGGT (SEQ ID NO: 62)
Pcsk9 sgRNA-3	CGATGGGCACCCACTGCTC (SEQ ID NO: 63)	GAGCAGTGGGTGCCCATCG (SEQ ID NO: 64)
Pcsk9 sgRNA-4	CCTTCACGTGGACGCGCAG (SEQ ID NO: 65)	CTGCGCGTCCACGTGAAGG (SEQ ID NO: 66)
Pcsk9 sgRNA-5	CGTGAAGGTGGAAGCCTTC (SEQ ID NO: 67)	GAAGGCTTCCACCTTCACG (SEQ ID NO: 68)
Npc1 sgRNA-1	CTCCTTGGTCAGGCGCCGG (SEQ ID NO: 69)	CCGGCGCCTGACCAAGGAG (SEQ ID NO: 70)
Npc1 sgRNA-2	TGGTCAGGCGCCGGTTCCG (SEQ ID NO: 71)	CGGAACCGGCGCCTGACCA (SEQ ID NO: 72)
Npc1 sgRNA-3	TAGAGGTCGCCTTCTCCTC (SEQ ID NO: 73)	GAGGAGAAGGCGACCTCTA (SEQ ID NO: 74)
Npc1 sgRNA-4	CGACGCTCGGGTCGCGGTG (SEQ ID NO:75)	CACCGCGACCCGAGCGTCG (SEQ ID NO:76)
Npc1 sgRNA-5	ATGCTGTCGCCGCGCGGGG (SEQ ID NO:77)	CCCCGCGCGGCGACAGCAT (SEQ ID NO:78)
Spcs1 sgRNA-1	CTCACCCTCACCGGAGCCA (SEQ ID NO:79)	TGGCTCCGGTGAGGGTGAG (SEQ ID NO:80)
Spcs1 sgRNA-2	CCGCAAACTTTACTCCTTA (SEQ ID NO:81)	TAAGGAGTAAAGTTTGCGG (SEQ ID NO:82)
Spcs1 sgRNA-3	CTCGGAGACATCCGCTTCC (SEQ ID NO: 60)	GGAAGCGGATGTCTCCGAG (SEQ ID NO: 60)
Spcs1 sgRNA-4	CTCCTAAGATTGGCTTCAC (SEQ ID NO:83)	GTGAAGCCAATCTTAGGAG (SEQ ID NO:84)
Spcs1 sgRNA-5	CCGGAGCCACTCCTAAGAT (SEQ ID NO:85)	ATCTTAGGAGTGGCTCCGG (SEQ ID NO:86)

Cd81 sgRNA-1	TTCTCTACCCTACGTCTCA (SEQ ID NO:87)	TGAGACGTAGGGTAGAGAA (SEQ ID NO:88)
Cd81 sgRNA-2	TACGTCTCATTCTCCGCAA (SEQ ID NO:89)	TTGCGGAGAATGAGACGTA (SEQ ID NO:90)
Cd81 sgRNA-3	GCTAGGCCTCCAGCCCTTC (SEQ ID NO:91)	GAAGGGCTGGAGGCCTAGC (SEQ ID NO:92)
Cd81 sgRNA-4	ACAGGTGGCGCCGCAACTT (SEQ ID NO:93)	AAGTTGCGGCGCCACCTGT (SEQ ID NO:94)
Cd81 sgRNA-5	AGCCGGAGGCGCGAGAGTC (SEQ ID NO:95)	GACTCTCGCGCCTCCGGCT (SEQ ID NO:96)

[0249] Example 5

[0250] FIG. 5 provides a schematic of the all-in-one protein constructs that were designed and tested for gene silencing. The initial design (p76, V1) of SEQ ID NO:1 was modified to encode XTEN linkers (e.g., 16 amino acids (SEQ ID NO: 31) or 80 amino acids (SEQ ID NO:32)) at either the N or C terminus of dCas9 (SEQ ID NO:29). All vectors contain HA tags (SEQ ID NO:24) at the C-terminus of dCas9. In aspects, CAG promoter is used since it provides good expression, for example, in constructs p76, and p90-102, p112 (V2). With reference to FIG. 5, the protein constructs of p90 to p102 correspond to SEQ ID NOS:2-14, respectively, and protein construct p112 corresponds to SEQ ID NO:15.

[0251] Example 6

[0252] The protein constructs shown in FIG. 5 were tested for silencing of the CLTA gene in HEK293T cells for 18 days post-transfection (FIGS. 6A-6B). Variable levels of gene silencing activities were detected, including a panel of variants with more durable gene silencing compared to the p76 (V1) design such as p99 (SEQ ID NO:11), p100 (SEQ ID NO:12), and p112 (SEQ ID NO:15). FIGS. 6A and 6B show tat the dCas9-KRAB and dCas9-Dnmt3A-Dnmt3L constructs showed transient and lower efficiency of long term silencing.

[0253] p76 (SEQ ID NO:1), p112 (SEQ ID NO:15) were tested for silencing the HIST2H2BE (H2B) endogenous gene and a synthetic Snrpn-GFP reporter gene stably expressed in HEK293T cells (FIGS. 6C-6D). Cells were followed for 50 days post-transfection. The p112 variant sustained gene silencing at a higher efficiency than the p76 (V1) design. The dCas9-Dnmt3A-Dnmt3L and dCas9-KRAB fusion proteins have transient and lower efficiency of long term silencing. FIG. 6E provides a plot of protein expression of p76 and p112 over the 50 day time course to turn off the HIST2H2BE (H2B) gene. Protein levels were measured by flow cytometry detection of BFP, which is co-expressed with the all-in-one protein.

[0254] Example 7

[0255] Western blot analysis was performed with the all-in-one protein variants p76, p90-p102 using an antibody against *Steptococcus pyogenes* Cas9. With reference to FIG. 7A, the top band represents full-length protein and smaller-sized bands represent proteolysis of the all-in-one protein. Variants that show little proteolysis, such as p99 (SEQ ID NO:11), p100 (SEQ ID NO:12), and p102 (SEQ ID NO:14), exhibited higher efficiency of gene silencing. Variants with high levels of proteolysis, such as p96 (SEQ ID NO:8) and p97 (SEQ ID NO:9), led to lower efficiency of sustained gene silencing.

[0256] Western blot analysis was performed with the all-in-one protein variants to detect free Dnmt3A that is cleaved from the fusion protein. As shown in FIG. 7B, variants that had little or no detectable free Dnmt3, such as p92 (SEQ ID NO:4), p100 (SEQ ID NO:12), p101 (SEQ ID NO:13), and p102 (SEQ ID NO:14), had higher efficiency of sustained gene silencing compared to variants with detectable cleaved Dnmt3A, i.e., p76 (SEQ ID NO:1), p91 (SEQ ID NO:3), p96 (SEQ ID NO:8), p98 (SEQ ID NO:10).

[0257] Example 8

[0258] A pooled screen was assayed, as shown in FIG. 8A, to determine the optimal sgRNAs that leads to long term gene silencing. Four HEK293T cell lines were used, each with a different gene with a GFP tag (CLTA, VIM, HIST2H2BE (H2B), and RAB11A). Tiling libraries consisting of sgRNAs that span +/- 2.5 kb from the transcription start site (TSS) of each gene were stably expressed in cells by lentiviral delivery, followed by transient expression plasmid DNA expressing the all-in-one protein. Four weeks post-transfection, cells that maintained gene silencing were sorted to determine the sgRNA identity. FIGS. 8B-8E are flow cytometry histograms showing the percent of cells undergoing gene silencing four weeks post-transfection.

[0259] FIGS. 9A-9D are maps of sgRNA functionality across the transcription start site of the targeted gene (CLTA, H2B, RAB11, VIM). The transcription start site (TSS) and CpG island are annotated above each plot. Each dot represents one sgRNA and its efficacy in long term gene silencing is plotted as the log2 fold change in sgRNA abundance. Nucleosome occupancy (bottom plot) is plotted from MNase signal.

[0260] Example 9

[0261] FIG. 10A shows the workflow of a pooled screen in HEK293T cells to determine optimal sgRNA targeting positions for the all-in-one protein, adapted from a previous ricin tiling screen in K562 cells to determine optimal sgRNAs for dCas9-KRAB (Gilbert, Horlbeck et al., Cell 2014). The sgRNAs are first stably expressed in HEK293T cells by lentiviral delivery, followed by transient transfection of a plasmid encoding the all-in-one protein (Day 0). Cells

expressing the all-in-one protein are sorted (Day 2) and allowed to grow for three more days. Cells are split on Day 5, from which half are harvested as an initial time point, and the other half are passaged for ten more days (Day 15) for a final time point. The growth phenotype (γ) is calculated as the log2 sgRNA enrichment divided by the number of cell doublings between T(initial) and T(final).

[0262] FIGS. 10B-10E are representative plots showing growth phenotypes for four genes (ARL1, EIF6, SMC3, HEATR1) from existing dCas9-KRAB/CRISPRi datasets in K562 cells (Gilbert, Horlbeck et al., 2014) and with the all-in-one protein (bottom plot). Each dot represents an sgRNA. The TSS and annotated CpG island are shown for each gene. The functional sgRNAs using the all-in-one protein spans a wider range than the functional sgRNAs, signifying a broader range of effective targeting.

[0263] Example 10

[0264] FIGS. 11A-11B provide a comparison of growth phenotypes and nucleosome positioning (from MNase signal) for VPS53 and VPS54 and show the location of functional sgRNAs at nucleosome-depleted regions. Furthermore, the range of functional sgRNAs is broader when using the all-in-one protein compared to dCas9-KRAB/CRISPRi.

[0265] Example 11

[0266] The in vitro transcription of two all-in-one variants (p102 (SEQ ID NO:14) and p112 (SEQ ID NO:15)) show full length synthesis of each design (FIGS. 12A). FIG. 12B provides a flow cytometry plot showing expression of p102 and p112 one day post-transfection of mRNA into HEK293T cells. FIG. 12C shows the time course of CLTA endogenous gene silencing in HEK293T cells after transfecting mRNA expressing the p102 and p112 all-in-one variants.

[0267] Example 12

[0268] FIG. 13A provides flow cytometry plots showing induced expression of the all-in-one protein by addition of doxycycline in K562 cells that stably encode the all-in-one protein under a doxycycline-inducible promoter. Protein expression was followed for four days after doxycycline induction. The dotted lines in the panels in FIG. 13A represent the baseline median BFP fluorescence without doxycycline administration. Western blots of cells were performed to detect expression of the all-in-one protein before and after doxycycline treatment (FIG. 13B). The presence of the all-in-one protein is not detectable by 96 hours post-induction. Flow cytometry plots of CD81 and CD151 knockdown 14 days post-doxycycline treatment of K562 cells are shown in FIGS. 13C-13F. The percent of cells with the targeted gene knocked down is shown. There is no detectable expression of the all-in-one protein, as no cells are present in the

BFP+ quadrants. Quantification of CD81 and CD151 knockdown 14 days post-doxycycline treatment or without doxycycline treatment is shown in FIG. 13G.

[0269] References

[0270] Ecco et al, Development 144, 2017. Lambert et al, Cell 172, 2018. Siddique et al., J. Mol. Biol., 425, 2013. Stepper et al, Nucleic Acids Res., 45, 2017. Shmakov et al., Nat. Rev. Microbiol. 15, 2017. Cebrian-Serrano et al, Mamm. Genome 7-8, 2017. Pulecio et al., Cell Stem Cell 21, 2017.

[0271] Informal Sequence Listing

[0272] SEQ ID NO:1 (p76 (all-in-one protein sequence, version 1): 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG*
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

*LGGD*SRADypydvpdyaSGS*pkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKVYPPVPAEKR KPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVT QKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDR PFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLA STVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEME RVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV***SSGNSNANSRGPS *FSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSG GGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQ YALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFS QNSLPL**SRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegegkpyegtqt mrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnftsngpv mqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyveqhevav arycdlpsklghkln*

[0273] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPGGSGGGS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS

*SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGD*SRADypydvpdya<u>SGS</u>*pkkkrkv*<u>SPG</u>SGSETPGTSESATPES***NHDQEFDPPKVYPPVPAE KRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRS VTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGD DRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRP LASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTE MERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV***<u>SSGNSNANSR</u>

<u>*GPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGS GSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHR ILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNA MRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYF KYFSQNSLPL**</u>SRAD*pkkkrkv*<u>GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegegk pyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvn ftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyve qhevavarycdlpsklghkln</u>*

[0274] SEQ ID NO:3 (p91 (KRAB-dCas9-Dnmt3A-Dnmt3L-<u>P2A-P2A</u>-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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEP<u>GGSGGGS</u>*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI*

LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
*LGGD*SRADypydvpdya*SGSpkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKVYPPVPAEKR***
KPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVT
QKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDR
PFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLA
STVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEME
RVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNSNANSRGPS*
*FSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSG**
GGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQ
YALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRV
WSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFS
QNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selikenmhmklym
egtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatq
dtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvy
yvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0275] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG*
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ

LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
*LGGD*SRADypydvpdyaSGS*pkkkrkv*SPGSGSETPGTSESATPES***NHDQEFDPPKVYPPVPAE***
KRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRS
VTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGD
DRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRP
LASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTE
MERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSR
*GPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGS**
GSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHR
ILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNA
MRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYF
KYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selikenmh
mklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedg
gvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlk
mpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0276] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG*

NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
*LGGD*SRADypydvpdyaSGS*pkkkrkv*SPGGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESG
PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE***NHDQEFDPPKVYPPVP***
AEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDV
RSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEG
DDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNR
PLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT
EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANS
*RGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESG**
SGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFH
RILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN
AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREY
FKYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegeg
kpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgv
nftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyv
eqhevavarycdlpsklghkln*

[0277] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGSGGGS*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG*
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLF
LAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
*LGGD*SRADypydvpdyaSGS*pkkkrkv*SPGGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESG
PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE***NHDQEFDPPKVYPPVP***
AEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDV
RSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEG
DDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNR
PLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCT
EMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANS
*RGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESG**
SGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFH
RILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQN
AMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREY
FKYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selikenm
hmklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyed
ggvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknl
kmpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0278] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPSGSETPGTSESATPES*MDKKYSIGLAIGTNSVGWAVITDEYKVPS*
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
*TRIDLSQLGGD*SRADypydvpdyaSGS*pkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKVYPP***
VPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVG
DVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPK
EGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGM
NRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILW
CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNA
*NSRGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLE**
SGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQ
FHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDY
QNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLR
EYFKYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctse
gegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvki
rgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeanne
tyveqhevavarycdlpsklghkln*

[0279] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPSGSETPGTSESATPES*MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGD*SRADypydvpdya*SGS*pkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKVYPP VPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVG DVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPK EGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGM NRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILW CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNSNA NSRGPSFSSGLVPLSLRGSH**MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLE SGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQ FHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDY QNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLR EYFKYFSQNSLPL**SRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selike nmhmklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervtt yedggvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpa knlkmpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0280] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSE
GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE*MDKKYSIGLAIGTNSVGWAVITDEYKV*
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA
KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE
HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
*LYETRIDLSQLGGD*SRADypydvpdyaSGS*pkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKV***
YPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM
YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDA
RPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKE
DILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSG
*NSNANSRGPSFSSGLVPLSLRGSH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSL**
GFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGW
YMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNC
LLPLREYFKYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdn

hhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqd gcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrl erikeannetyveqhevavarycdlpsklghkln*

[0281] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE*MDKKYSIGLAIGTNSVGWAVITDEYKV PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG LYETRIDLSQLGGDSRADypydvpdyaSGS*pkkkrkv*EASGSGRASPGIPGSTR***NHDQEFDPPKV YPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDA RPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKE DILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV**SSG NSNANSRGPSFSSGLVPLSLRGSH**MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSL GFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGW**

YMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNC LLPLREYFKYFSQNSLPLSRAD*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenp gp**selikenmhmklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpeg ftwervttyedggvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianiktt yrskkpaknlkmpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0282] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEPSGSETPGTSESATPES*MDKKYSIGLAIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGD*SRADypydvpdyaSGS*pkkkrkv*SPG***ggpssgapppsggspagsptsteegtsesatpesgpgtste psegsapgspagsptsteegtstepsegsapgtstepseNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIAT GLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLV IGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMG VSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLE***

HGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVS
*NMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNSNANSRGPSFSSGLVPLSLRGSH
MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVT
NVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPF
FWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKH
APLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRAD*pk*
*kkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpfafdi
latsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweaftetlyp
adgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkln*

[0283] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEPSGSETPGTSESATPES*MDKKYSIGLAIGTNSVGWAVITDEYKVPS*
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLA
LAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE
*TRIDLSQLGGD*SRADypydvpdyaSGS*pkkkrkv*SPG***ggpssgapppsggspagsptsteegtsesatpesgpgtste***

psegsapgspagsptsteegtstepsegsapgtstepseNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIAT GLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLV IGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMG VSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLE HGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVS NMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSH**MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVT NVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPF FWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKH APLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL**SRAD***pk kkrkv***GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegeg kpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgv nftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyv eqhevavarycdlpsklghkln*

[0284] 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))

**DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY QLTKPDVILRLEKGEEP*ggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsapgspagsptsteegt stepsegsapgtstepse*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK

KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
*TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*SRADypydvp
dya*SGSpkkkrkv*SPGSGSETPGTSESATPES***NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGI***
ATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFD
LVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVA
MGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQEC
LEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTD
VSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV*SSGNSNANSRGPSFSSGLVPLSLRG*
*SH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVED**
VTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQ
RPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKS
KHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRA
D*pkkkrkv*GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctsegegkpyegtqtmrikvveggplpf
afdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnftsngpvmqkktlgweafte
tlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeannetyveqhevavarycdlpsklghkl
n*

[0285] 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))

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGY
QLTKPDVILRLEKGEEP*ggpssgapppsggspagsptsteegtsesatpesgpgtstepsegsapgspagsptsteegt*
stepsegsapgtstepse*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL*
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYP
FLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT
LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP

SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR
PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP
EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
*TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*SRADypydvp
dyaSGS*pkkkrkv*SPGSGSETPGTSESATPES***NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGI***
ATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFD
LVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVA
MGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQEC
LEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTD
VSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRG
*SH***MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVED**
VTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQ
RPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKS
KHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPLSRA
D*pkkkrkv*GSG**atnfsllkqagdveenpgp**GSG**atnfsllkqagdveenpgp**selikenmhmklymegtvdnhhfkctse
gegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvki
rgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerikeanne
tyveqhevavarycdlpsklghkln*

[0286] SEQ ID NO:15 (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))

NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSIT
VGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGR
LFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ
HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFA
PLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSH**MGPMEIYKTVSAWKRQPVRVLSL**
FRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPL
GSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQ
TEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAP
KVDLLVKNCLLPLREYFKYFSQNSLPL*ggpssgapppsggspagsptsteegtsesatpesgpgtstepsegs*
apgspagsptsteegtstepsegsapgtstepse*MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTD*
RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG

HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR
DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH
HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNF
FKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS
SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
*LGGD*Aypydvpdya<u>SLGSGS*pkkkrkv*ED*pkkkrkv*DGIGSGSNGSSGSselikenmhmklymegtvdnhhf</u>
<u>kctsegegkpyegtqtmrikvveggplpfafdilatsflygsktfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcli</u>
<u>ynvkirgvnftsngpvmqkktlgweaftetlypadgglegrndmalklvggshlianikttyrskkpaknlkmpgvyyvdyrlerik</u>
<u>eannetyveqhevavarycdlpsklghklnGGGGGMDAKSLTAWS</u>**RTLVTFKDVFVDFTREEWKL**
LDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP*

[0287] SEQ ID NO:16 (KRAB; from Gilbert et al., Cell, 2013, 2014)

DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLT
KPDVILRLEKGEEP

[0288] SEQ ID NO:17 (Linker)

GGSGGGS

[0289] SEQ ID NO:18 (Linker)

SGS

[0290] SEQ ID NO:19 (Linker)

EASGSGRASPGIPGSTR

[0291] SEQ ID NO:20 (Linker)

SRAD

[0292] SEQ ID NO:21 (Linker)

GSG

[0293] SEQ ID NO:22 (Linker

SPG

[0294] SEQ ID NO:23 (dCas9)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

[0295] SEQ ID NO:24 (HA tag)

YPYDVPDYA

[0296] SEQ ID NO:25 (SV40 NLS)

PKKKRKV

[0297] SEQ ID NO:26 (Dnmt3A; residues 612-912; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITV
GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL
FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAH
RARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHF

PVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK
EYFACV

[0298] SEQ ID NO:27 (27 amino acid linker; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

SSGNSNANSRGPSFSSGLVPLSLRGSH

[0299] SEQ ID NO:28 (Dnmt3L; from Siddique et al., JMB, 2013; Stepper et al., NAR, 2016)

MGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVV
RRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMD
NLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEY
LQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL

[0300] SEQ ID NO:29 (P2A peptide cleave sequence)

ATNFSLLKQAGDVEENPGP

[0301] SEQ ID NO:30 (BFP)

SELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILA
TSFLYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNV
KIRGVNFTSNGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANIKTT
YRSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQHEVAVARYCDLPSKLGHKLN*

[0302] SEQ ID NO:31 (XTEN16 (16 amino acid sequence))

SGSETPGTSESATPES

[0303] SEQ ID NO:32 (XTEN80 (80 amino acid sequence))

GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT
STEPSEGSAPGTSTEPSE

[0304] SEQ ID NO:33 (Dnmt3A-Dnmt3L domain)

NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITV
GMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRL
FFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAH
RARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHF
PVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLK
EYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFRN
IDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDR
CPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPL
REYFKYFSQNSLPL

[0305] SEQ ID NO:34 (ddAsCfp1)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY
ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDA
INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNV
FSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSF
PFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL
FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISH
KKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGK
ELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDES
NEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN
KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMI
PKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ
KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIA
EKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQA
ELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEA
RALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETP
IIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTI
KDLKQGYLSQVIHEIVDLMIHYQAVVVLANLNFGFKSKRTGIAEKAVYQQFEKMLIDK
LNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVD
PFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILP
KLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWP
MDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

[0306] SEQ ID NO:35 (ddLbCfp1)

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYY
LSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLF
KKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLT
RYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIG
GFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVL
EVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKW
NAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQK
VDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKET
NRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKD
KETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPK
VFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAY
DFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDK
SHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPD
NPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIA
RGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTS
IENIKELKAGYISQVVHKICELVEKYDAVIALADLNSGFKNSRVKVEKQVYQKFEKMLI
DKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGF
VNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYG
NRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFM
ALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNI
ARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH

[0307] SEQ ID NO:36 (ddFnCfp1)

MYPYDVPDYASGSGMSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAK
DYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTI
KKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEAL
EIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPE

AINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGK
FVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDV
VTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFD
DYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDL
LDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKP
YSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKE
NKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGY
EKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFE
NISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLN
GEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGN
DRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVV
FEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFE
TFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKG
YFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKD
YSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNR
NN

CLAIMS

What is claimed is:

1. A fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Krüppel associated box domain, and a DNA methyltransferase domain.

2. The fusion protein of claim 1, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9, ddCpf1, a nuclease-deficient Cas9 variant, or a nuclease-deficient Class II CRISPR endonuclease.

3. The fusion protein of claim 1, wherein said DNA methyltransferase domain is a Dnmt3A-3L domain.

4. The fusion protein of claim 1, 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.

5. The fusion protein of claim 4, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and said DNA methyltransferase domain is a Dnmt3A-3L domain.

6. The fusion protein of claim 5, wherein said dCas9 is covalently linked to said Dnmt3A-3L domain via a peptide linker and wherein said Dnmt3A-3L domain is covalently linked to said Krüppel associated box domain via a peptide linker.

7. The fusion protein of claim 6, wherein said peptide linker is a XTEN linker.

8. The fusion protein of claim 1, 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.

9. The fusion protein of claim 8, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is dCas9 and said DNA methyltransferase domain is a Dnmt3A-3L domain.

10. The fusion protein of claim 9, wherein said dCas9 is covalently linked to said Dnmt3A-3L domain via a peptide linker and wherein said Krüppel associated box domain is covalently linked to said dCas9 via a peptide linker.

11. The fusion protein of claim 10, wherein said peptide linker is a XTEN

linker.

12. The fusion protein of claim 1, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to said Krüppel associated box domain via a peptide linker.

13. The fusion protein of claim 1, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is covalently linked to said DNA methyltransferase domain via a peptide linker.

14. The fusion protein of claim 1, wherein said Krüppel associated box domain is covalently linked to said DNA methyltransferase domain via a peptide linker.

15. The fusion protein of claim 12, wherein said peptide linker is a XTEN linker.

16. The fusion protein of claim 15, wherein said XTEN linker comprises about 16 to 80 amino acid residues.

17. The fusion protein of claim 1, further comprising a nuclear localization signal peptide.

18. The fusion protein of claim 1, wherein said fusion protein comprises the amino acid sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 12, 13, 14, or 15.

19. A nucleic acid sequence encoding said fusion protein of claim 1.

20. The nucleic acid sequence of claim 19, wherein the nucleic acid sequence is messenger RNA.

21. A complex comprising: (i) a fusion protein of claim 1; and (ii) a polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for said nuclease-deficient RNA-guided DNA endonuclease enzyme, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is bound to said polynucleotide via said binding sequence.

22. The complex of claim 21, wherein said target polynucleotide sequence is part of a gene.

23. The complex of claim 21, wherein said target polynucleotide sequence is part of a transcriptional regulatory sequence.

24. The complex of claim 21, wherein said target polynucleotide sequence is

part of a promoter, enhancer, or silencer.

25. The complex of claim 21, wherein said target polynucleotide sequence is within about 3000 bp flanking a transcription start site.

26. A vector comprising said nucleic acid sequence of claim 19.

27. The vector of claim 26, further comprising a polynucleotide, wherein said polynucleotide comprises: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for said nuclease-deficient RNA-guided DNA endonuclease enzyme.

28. A cell comprising the fusion protein of claim 1; the nucleic acid of claim 19; the complex of claim 21, or the vector of claim 26.

29. The cell of claim 28, wherein said cell is a eukaryotic cell.

30. The cell of claim 28, wherein said cell is a mammalian cell.

31. A method of silencing a target nucleic acid sequence in a cell, comprising: (i) delivering a first polynucleotide encoding a fusion protein of claim 1 to a cell containing said target nucleic acid; and (ii) delivering to said cell a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to said target nucleic acid sequence; and (b) a binding sequence for said nuclease-deficient RNA-guide DNA endonuclease enzyme.

32. The method of claim 31, wherein said fusion protein silences said target nucleic acid sequence in said cell by methylating a chromatin containing said target nucleic acid sequence and/or by introducing repressive chromatin marks to a chromatin containing said target nucleic acid sequence.

33. The method of claim 31, wherein said first polynucleotide is contained within a first vector.

34. The method of claim 31, wherein said first polynucleotide is contained within a second vector.

35. The method of claim 34, wherein said first vector and the second vector are the same.

36. The method of claim 34, wherein said first vector is different from said second vector.

37. The method of claim 31, wherein said cell is a mammalian cell.

38. The method of claim 31, wherein said method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

39. A method of silencing a target nucleic acid sequence in a cell, the method comprising delivering the complex of claim 21 to a cell containing said target nucleic acid.

40. The method of claim 39, wherein said complex silences said target nucleic acid sequence in said cell by methylating a chromatin containing said target nucleic acid sequence and/or by introducing repressive chromatin marks to a chromatin containing said target nucleic acid sequence.

41. The method of claim 39, wherein said cell is a mammalian cell.

42. The method of claim 39, wherein said method has a specificity that is 2-fold higher than a specificity to a non-target nucleic acid sequence.

FIG. 1A

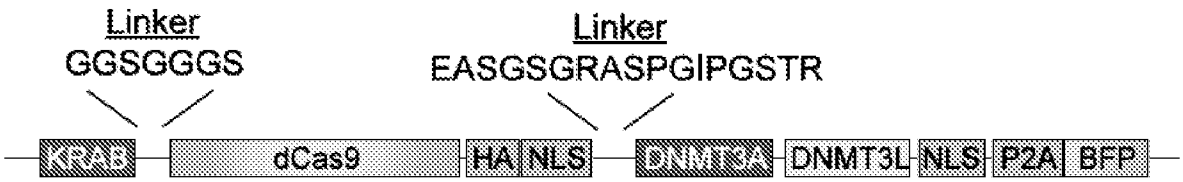


FIG. 1B

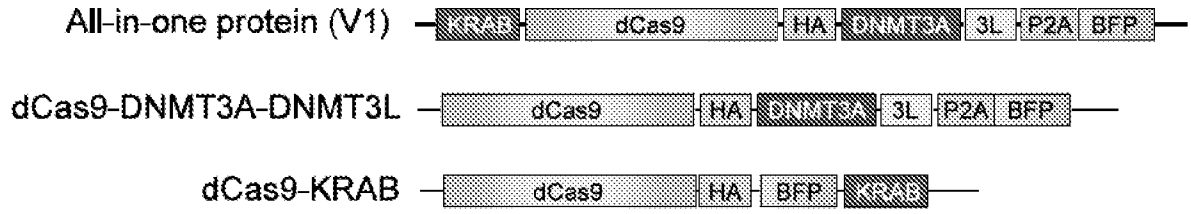


FIG. 1C

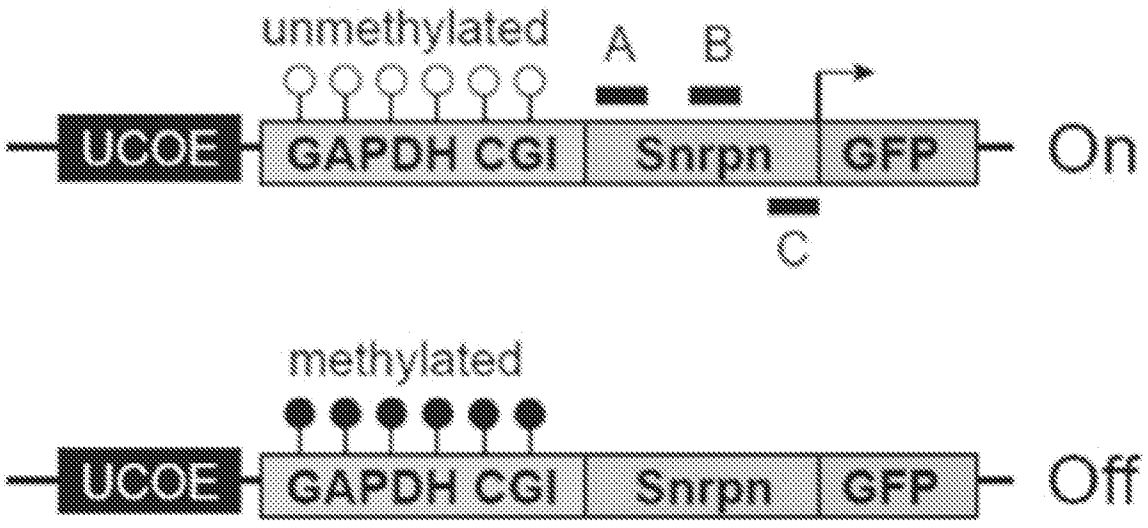


FIG. 1D

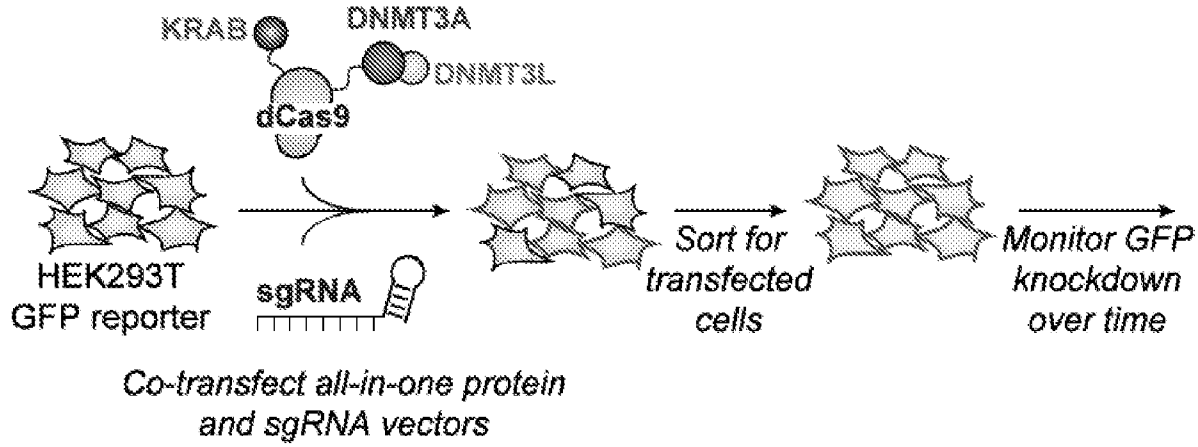

KRAB
DNMT3A
DNMT3L
dCas9
HEK293T
GFP reporter
sgRNA
Co-transfect all-in-one protein
and sgRNA vectors
Sort for
transfected
cells
Monitor GFP
knockdown
over time

FIG. 1E

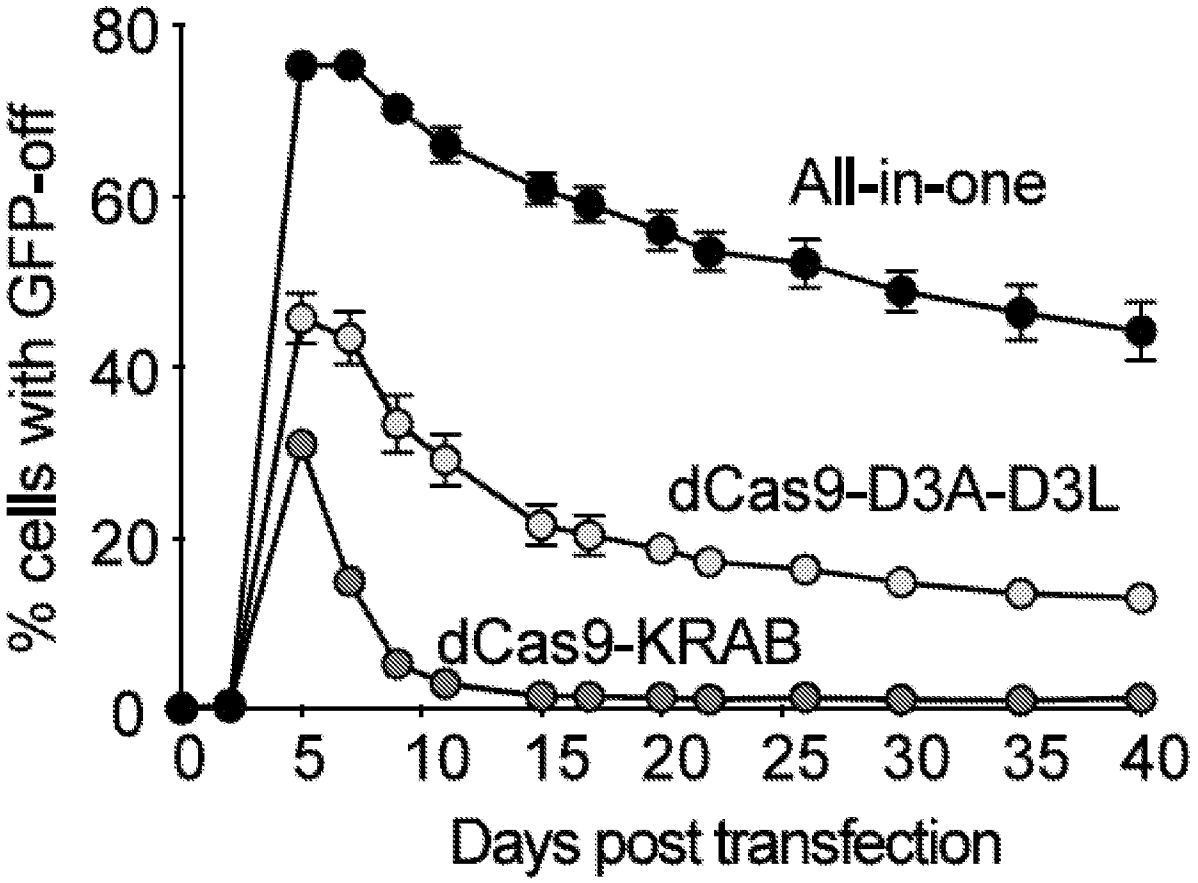

% cells with GFP-off
80
60
40
20
0
0
5
10
15
20
25
30
35
40
Days post transfection
All-in-one
dCas9-D3A-D3L
dCas9-KRAB

FIG. 1F

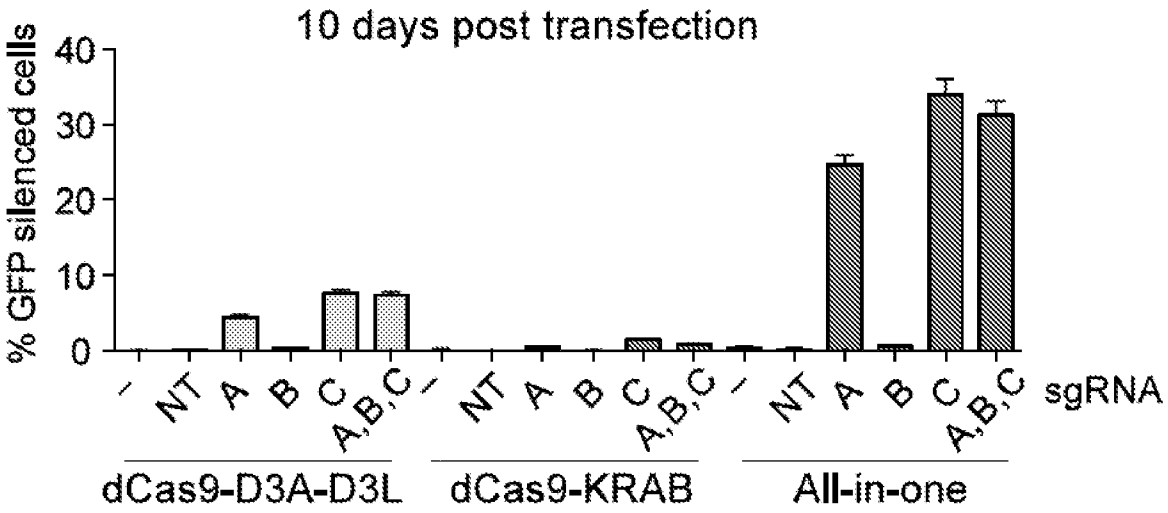


FIG. 2A FIG. 2B FIG. 2C

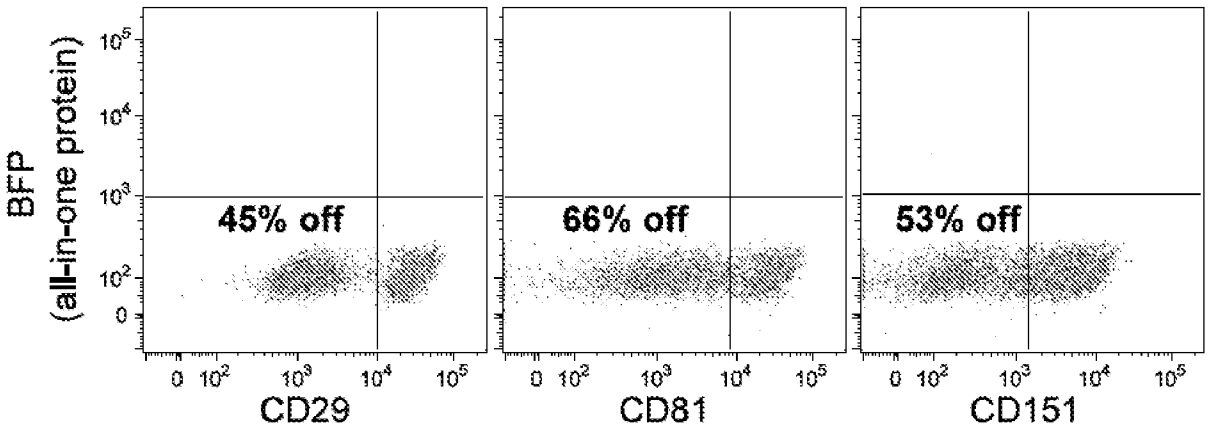


FIG. 2D

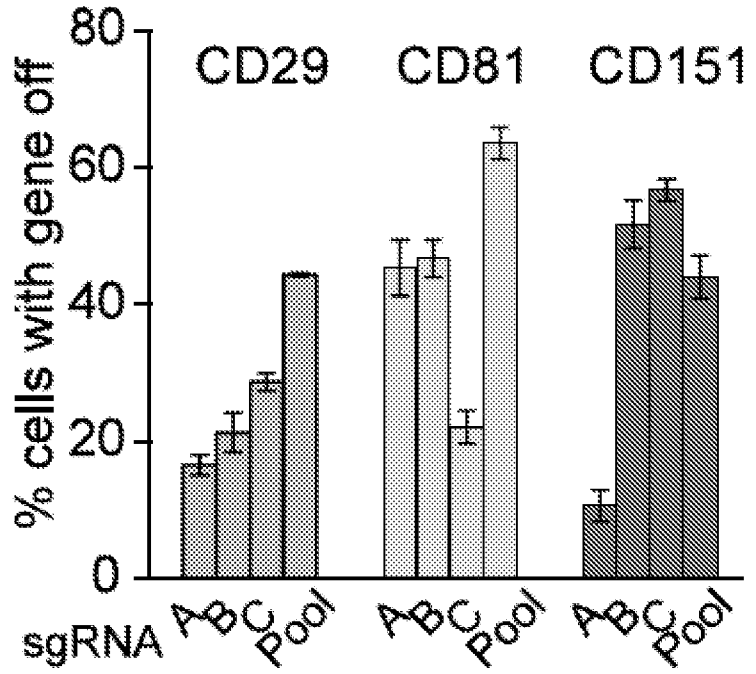

CD29
CD81
CD151
% cells with gene off
80
60
40
20
0
sgRNA
A
B
C
Pool

FIG. 2E

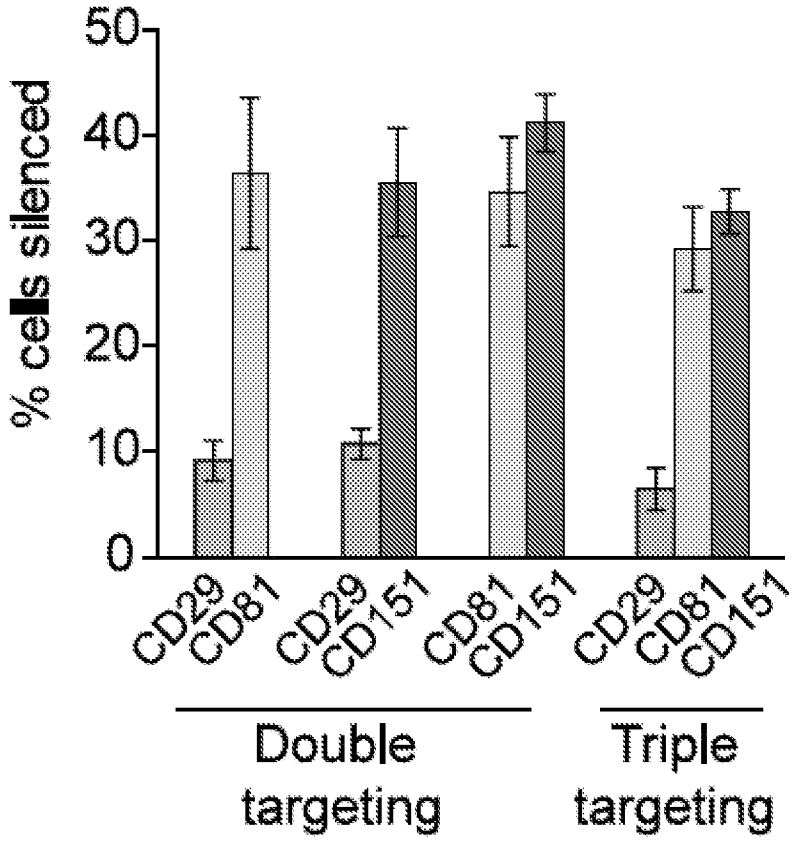

% cells silenced
50
40
30
20
10
0
CD29
CD81
CD29
CD151
CD81
CD151
CD29
CD81
CD151
Double targeting
Triple targeting

FIG. 2F

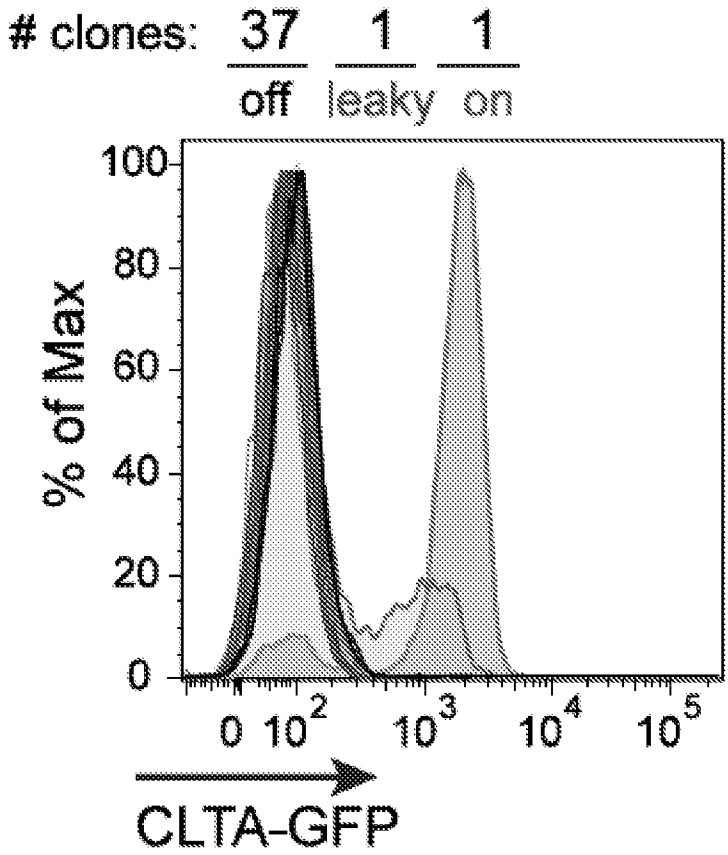


FIG. 3A

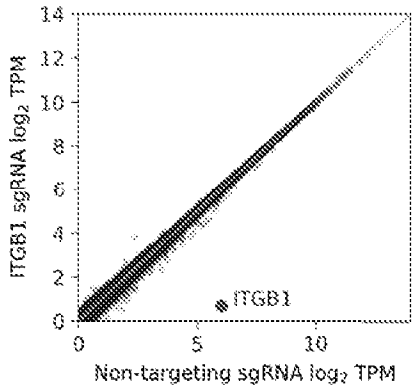


FIG. 3B

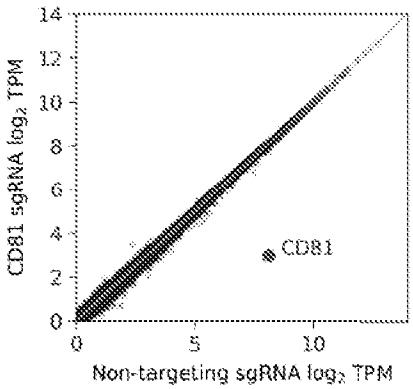


FIG. 3C

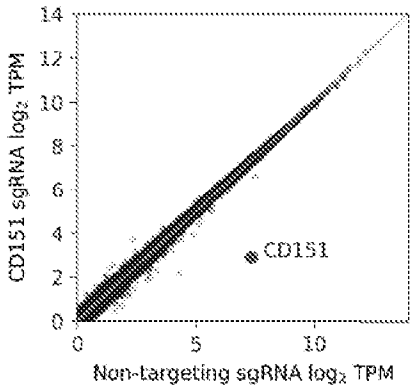


FIG. 3D

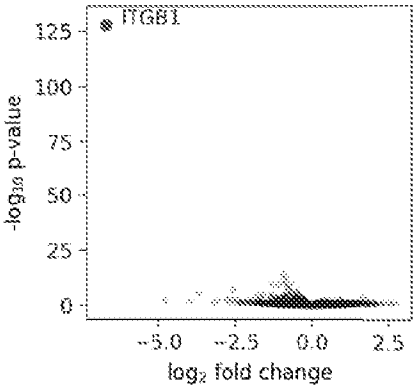


FIG. 3E

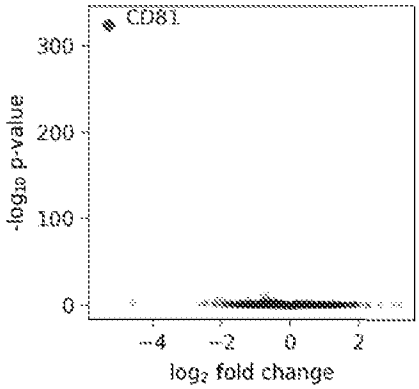


FIG. 3F

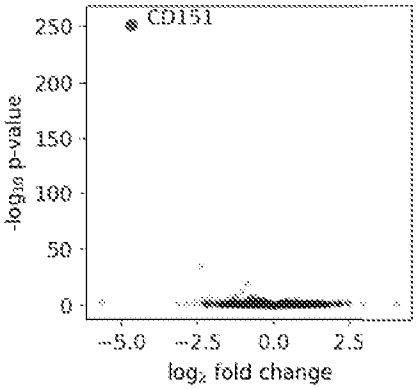


FIG. 3G

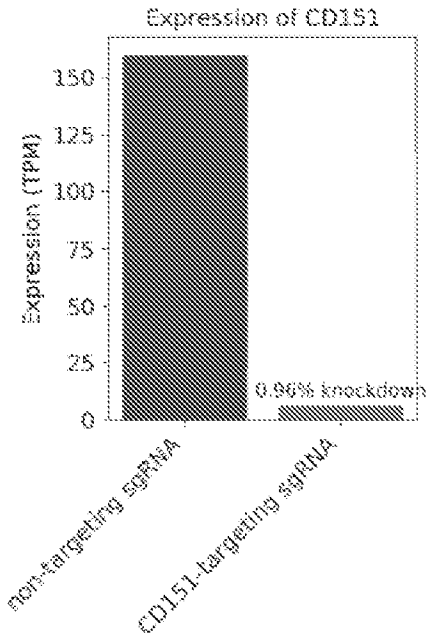


FIG. 3H

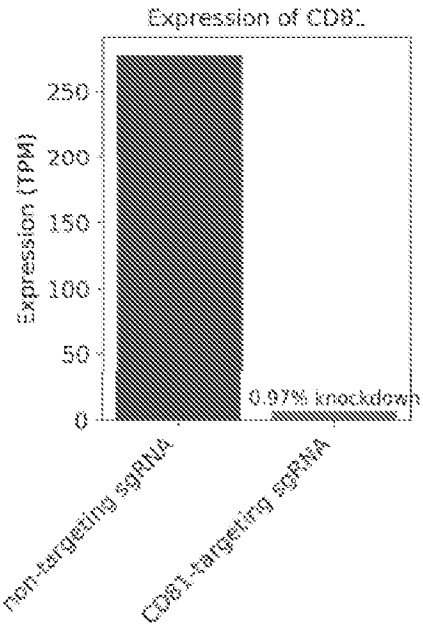


FIG. 3I

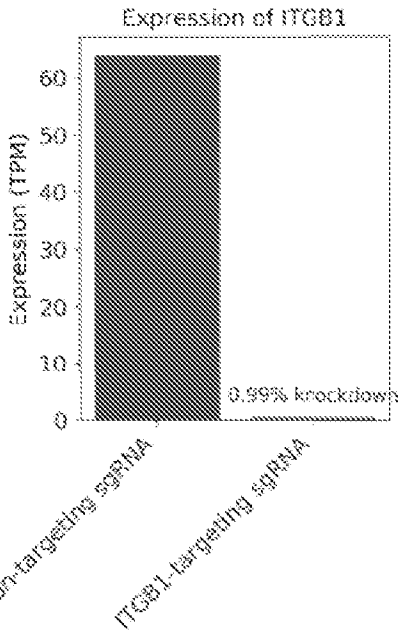


FIG. 4A

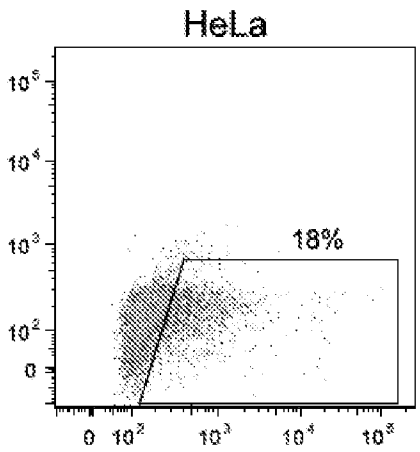


FIG. 4B

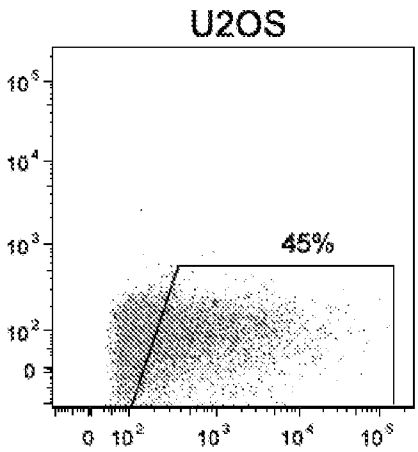


FIG. 4C

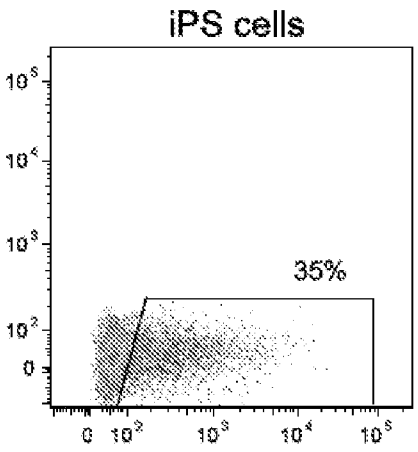


FIG. 4D

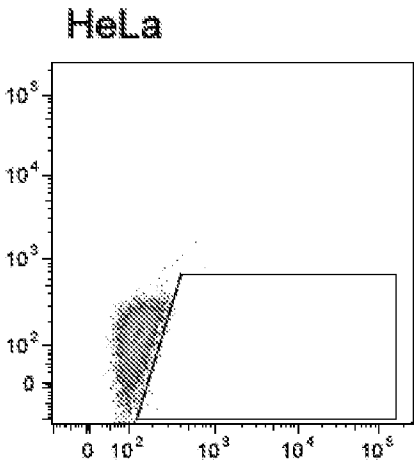


FIG. 4E

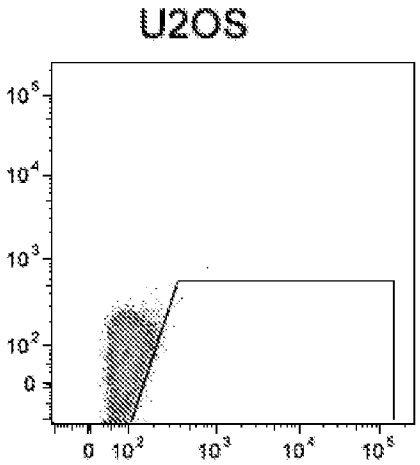


FIG. 4F

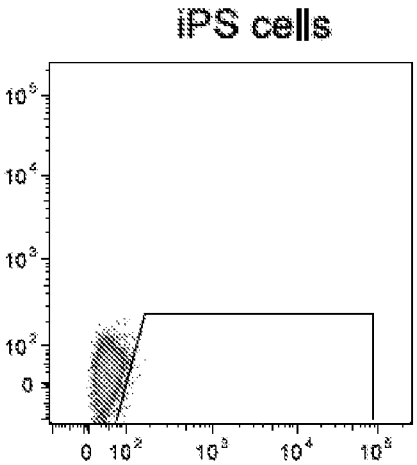


FIG. 4G

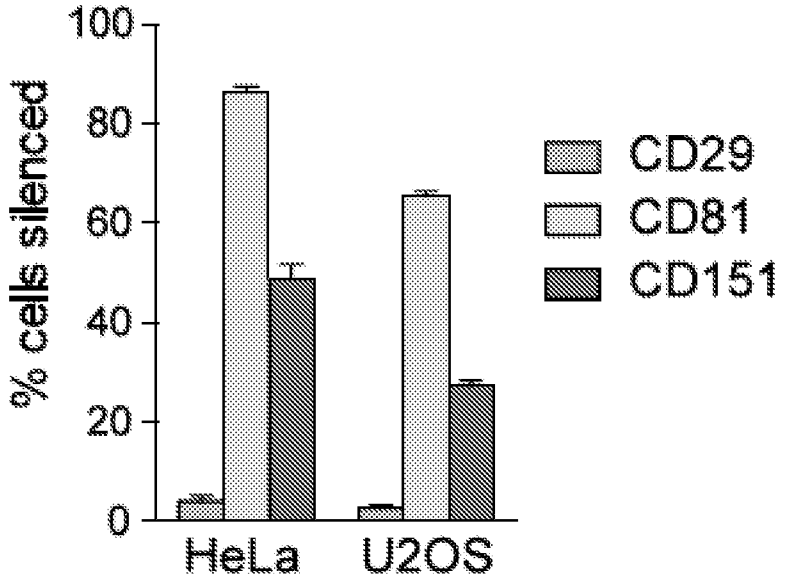

100
80
60
40
20
0
% cells silenced
CD29
CD81
CD151
HeLa
U2OS

FIG. 4H

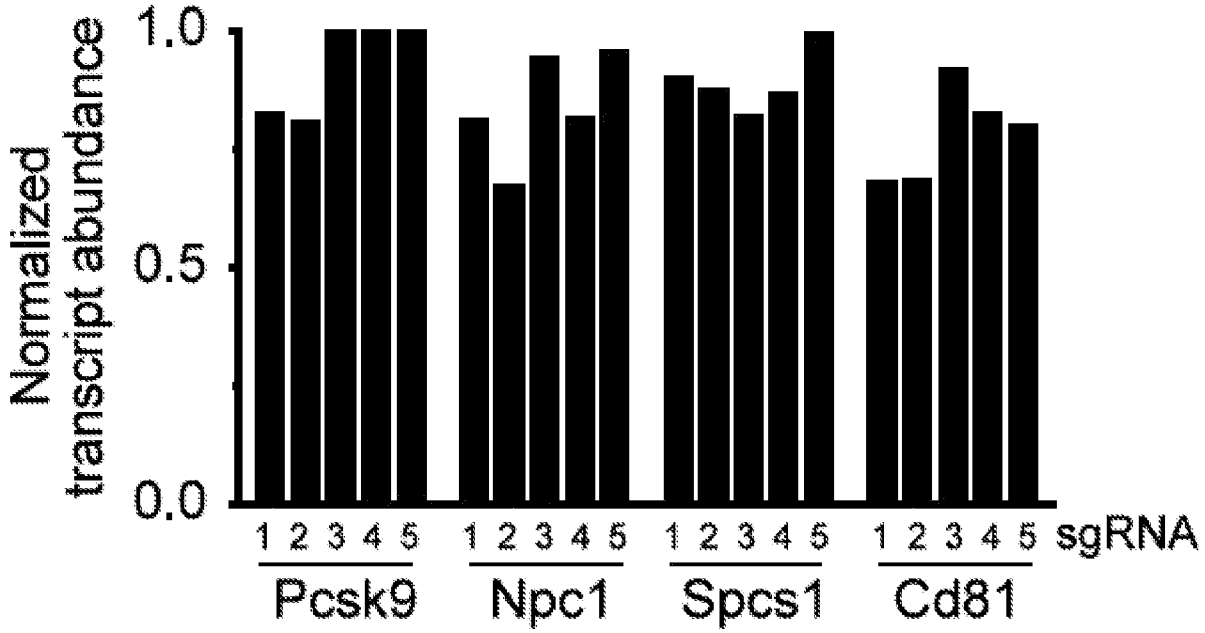

1.0
0.5
0.0
Normalized transcript abundance
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
sgRNA
Pcsk9
Npc1
Spcs1
Cd81

FIG. 5

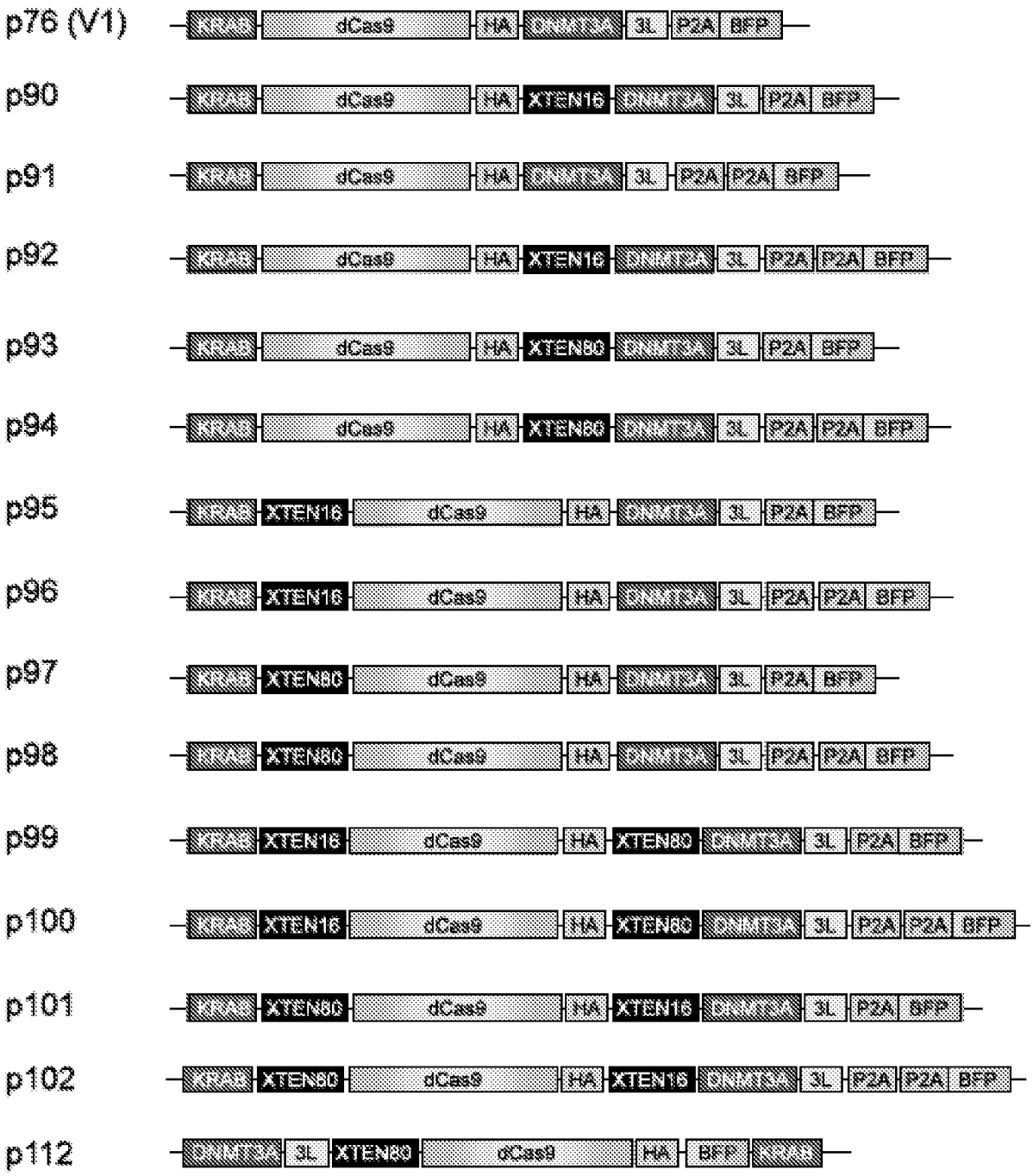

p76 (V1)
KRAB
dCas9
HA
DNMT3A
3L
P2A
BFP
p90
KRAB
dCas9
HA
XTEN16
DNMT3A
3L
P2A
BFP
p91
KRAB
dCas9
HA
DNMT3A
3L
P2A
P2A
BFP
p92
KRAB
dCas9
HA
XTEN16
DNMT3A
3L
P2A
P2A
BFP
p93
KRAB
dCas9
HA
XTEN80
DNMT3A
3L
P2A
BFP
p94
KRAB
dCas9
HA
XTEN80
DNMT3A
3L
P2A
P2A
BFP
p95
KRAB
XTEN16
dCas9
HA
DNMT3A
3L
P2A
BFP
p96
KRAB
XTEN16
dCas9
HA
DNMT3A
3L
P2A
P2A
BFP
p97
KRAB
XTEN80
dCas9
HA
DNMT3A
3L
P2A
BFP
p98
KRAB
XTEN80
dCas9
HA
DNMT3A
3L
P2A
P2A
BFP
p99
KRAB
XTEN16
dCas9
HA
XTEN80
DNMT3A
3L
P2A
BFP
p100
KRAB
XTEN16
dCas9
HA
XTEN80
DNMT3A
3L
P2A
P2A
BFP
p101
KRAB
XTEN80
dCas9
HA
XTEN16
DNMT3A
3L
P2A
BFP
p102
KRAB
XTEN80
dCas9
HA
XTEN16
DNMT3A
3L
P2A
P2A
BFP
p112
DNMT3A
3L
XTEN80
dCas9
HA
BFP
KRAB

FIG. 6A

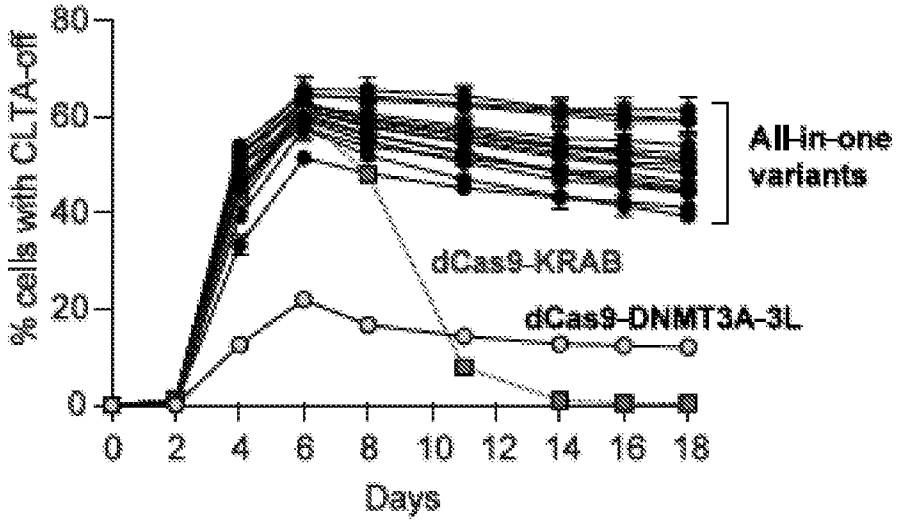

% cells with CLTA-off
80
60
40
20
0
0 2 4 6 8 10 12 14 16 18
Days
All-in-one variants
dCas9-KRAB
dCas9-DNMT3A-3L

FIG. 6B

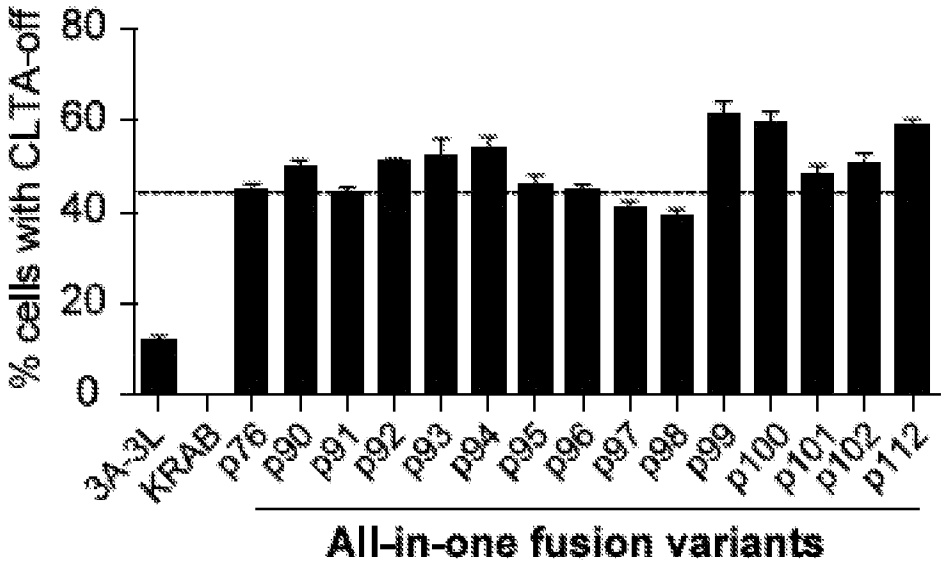

% cells with CLTA-off
80
60
40
20
0
3A-3L
KRAB
p76
p90
p91
p92
p93
p94
p95
p96
p97
p98
p99
p100
p101
p102
p112
All-in-one fusion variants

FIG. 6C

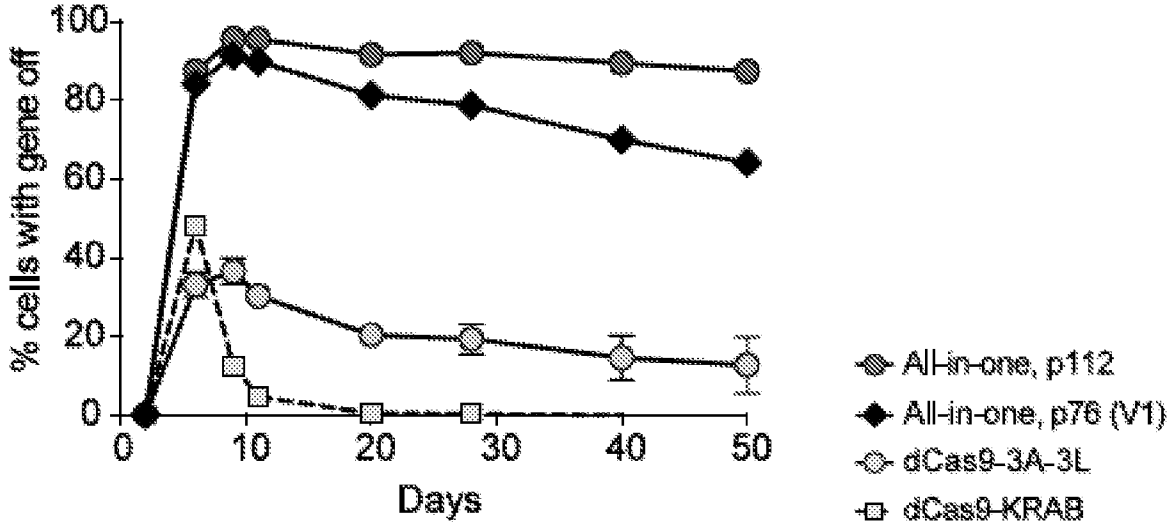

% cells with gene off
100
80
60
40
20
0
0
10
20
30
40
50
Days
All-in-one, p112
All-in-one, p76 (V1)
dCas9-3A-3L
dCas9-KRAB

FIG. 6D

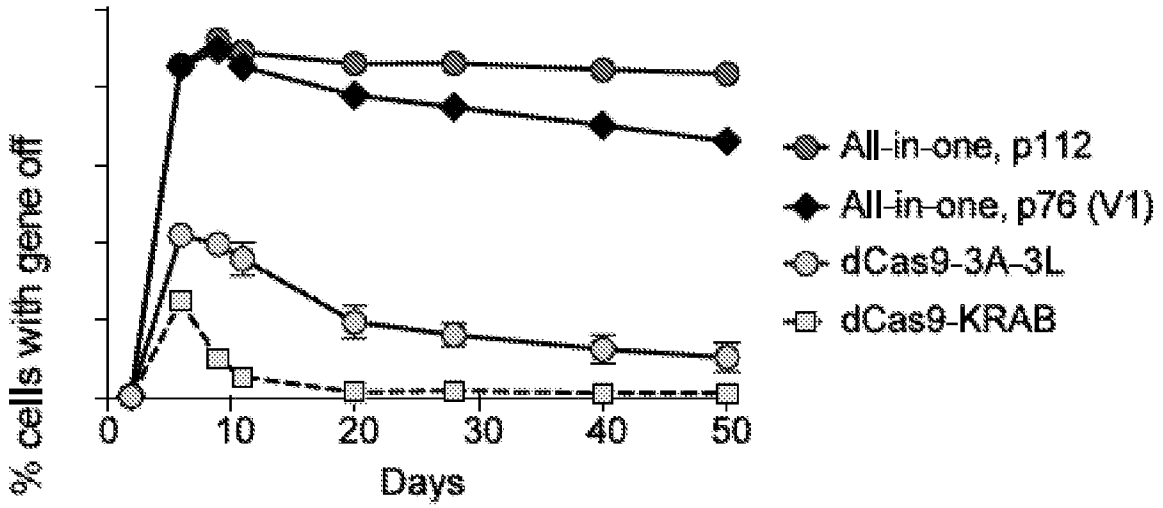

% cells with gene off
0
10
20
30
40
50
Days
All-in-one, p112
All-in-one, p76 (V1)
dCas9-3A-3L
dCas9-KRAB

FIG. 6E

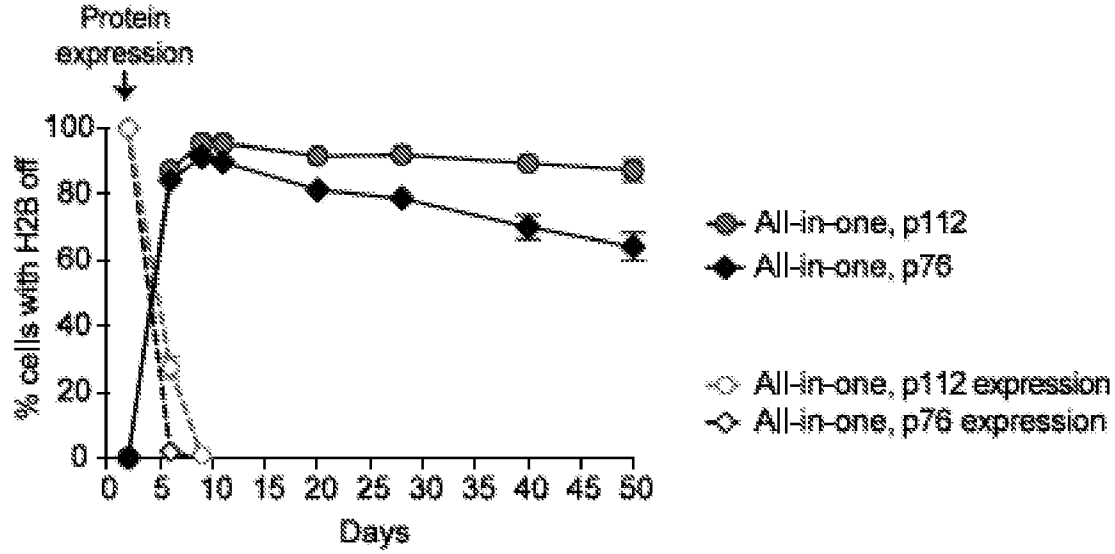


FIG. 7A

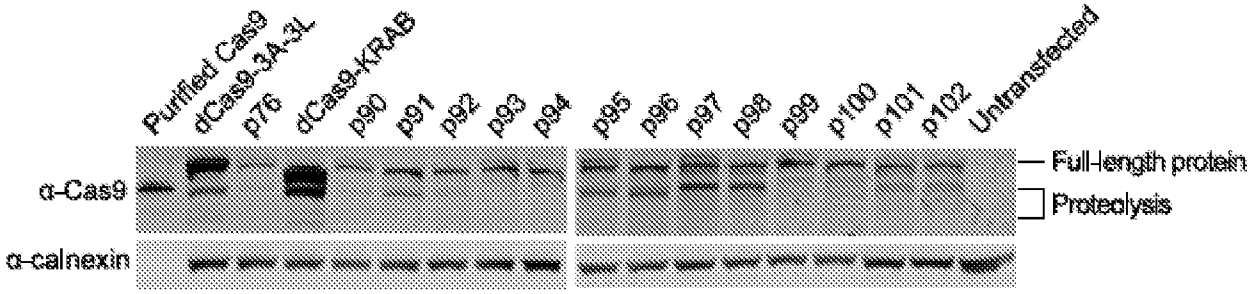


FIG. 7B



FIG. 8A

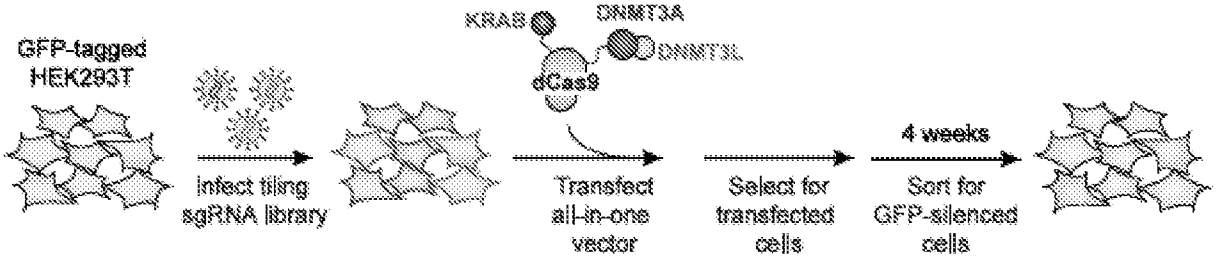


FIG. 8B FIG. 8C FIG. 8D FIG. 8E

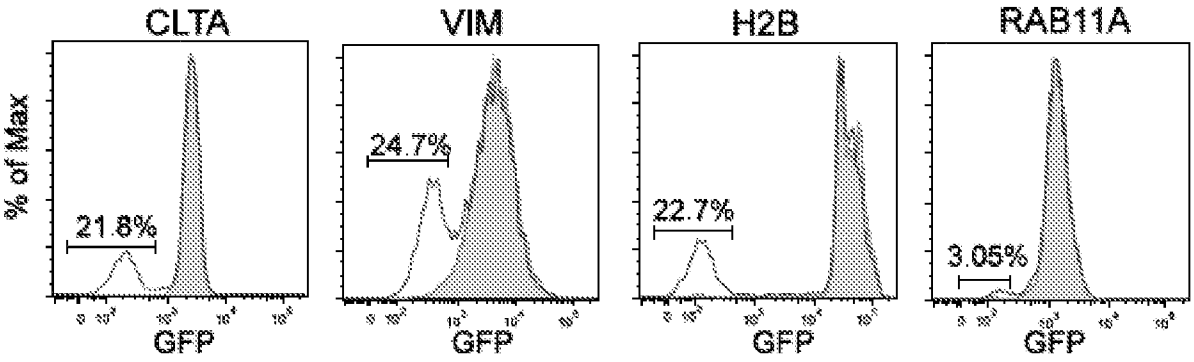


FIG. 9A FIG. 9B

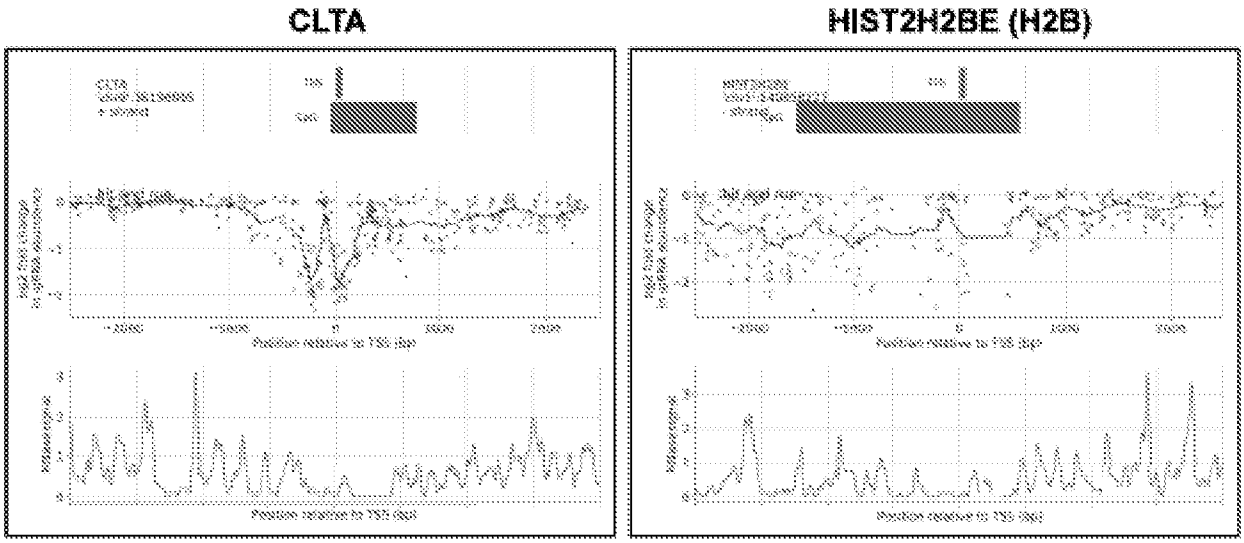


FIG. 9C

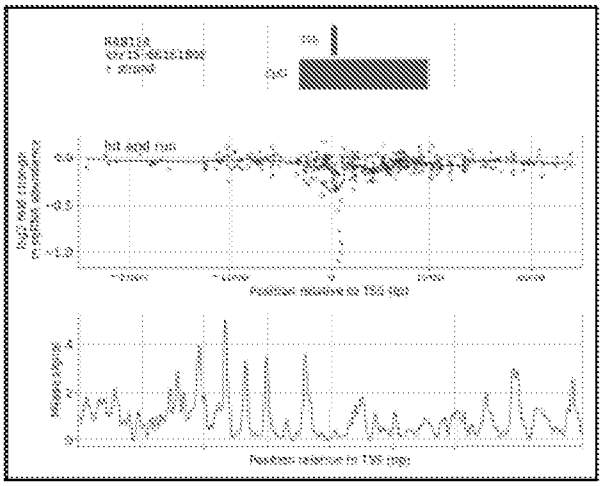


FIG. 9D

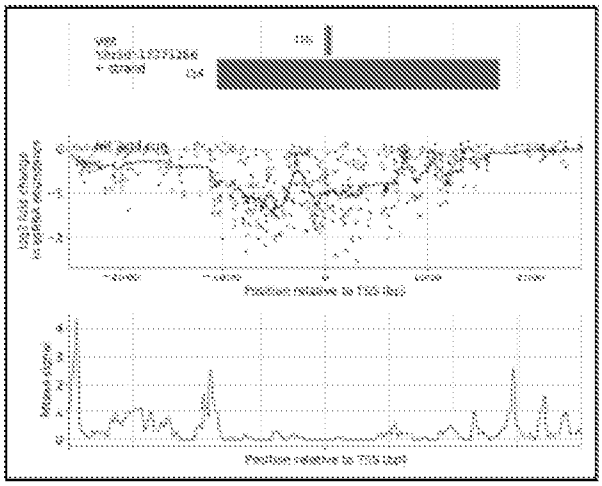


FIG. 10A

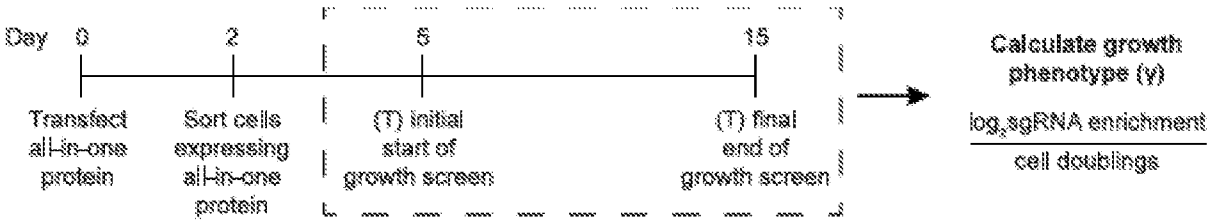


FIG. 10B

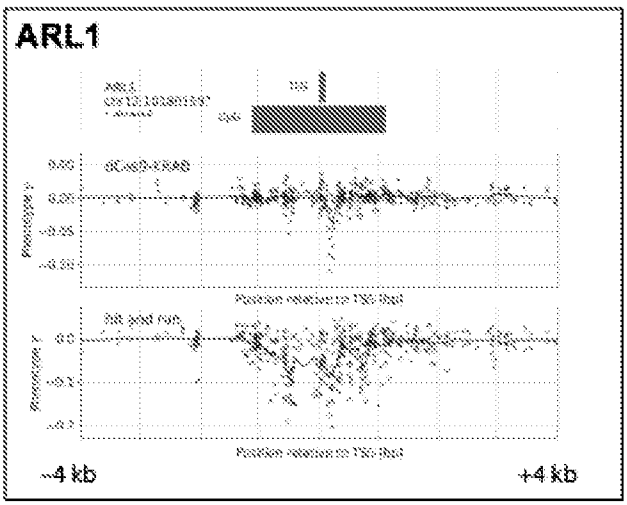


FIG. 10C

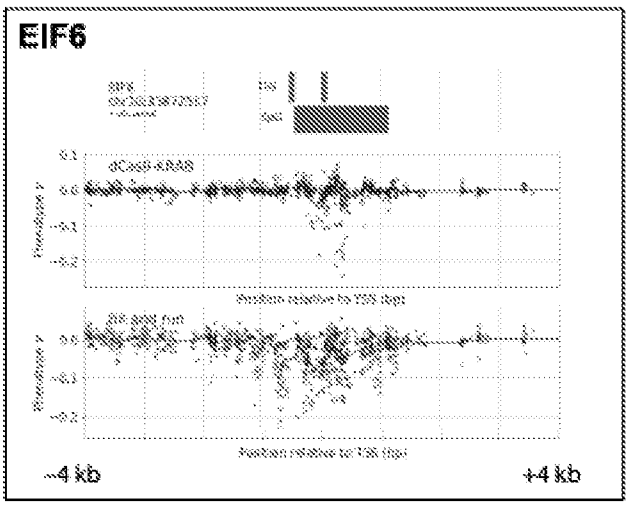


FIG. 10D

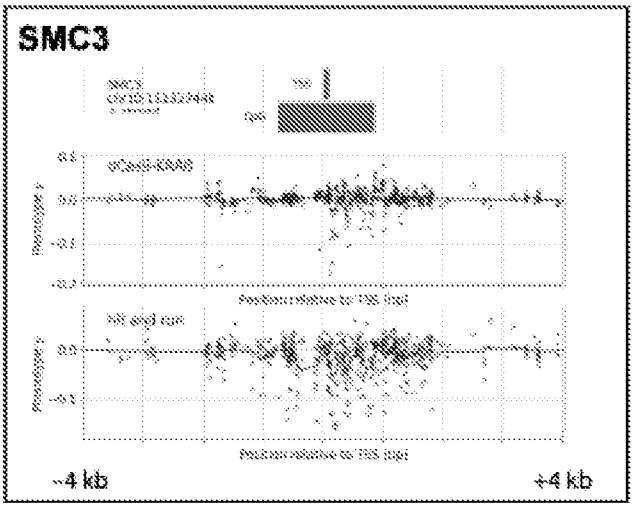


FIG. 10E

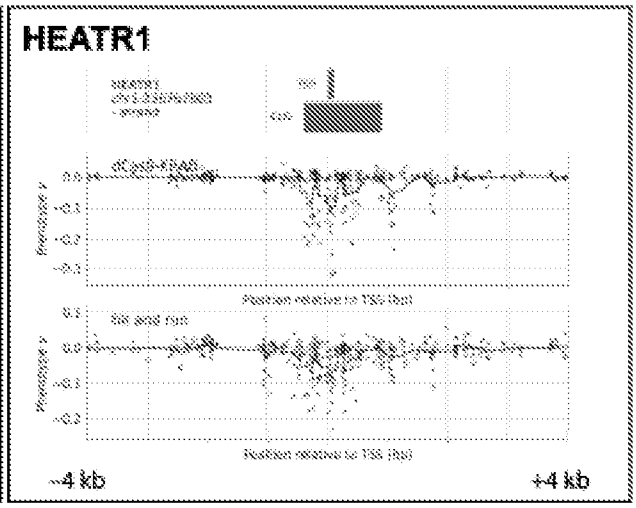


FIG. 11A

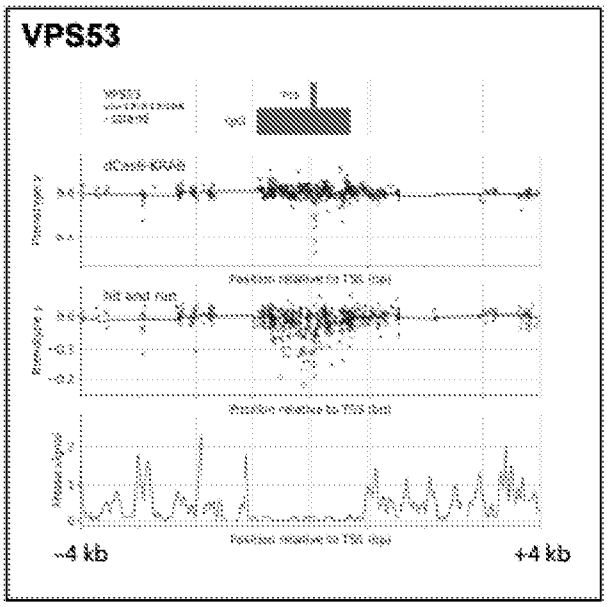


FIG. 11B

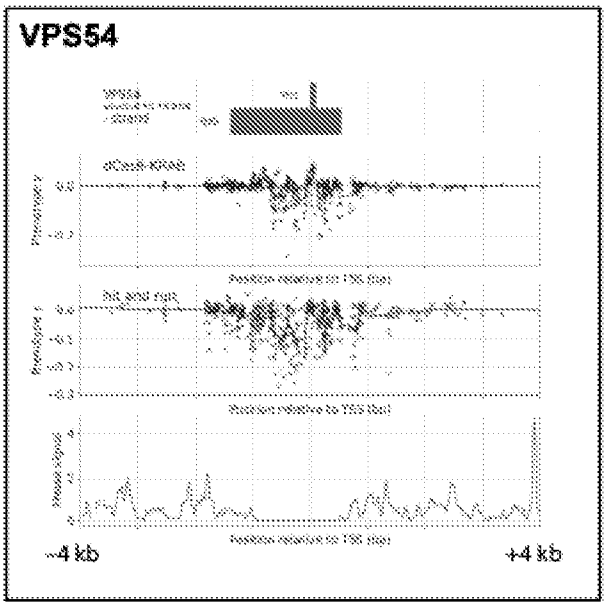


FIG. 12A

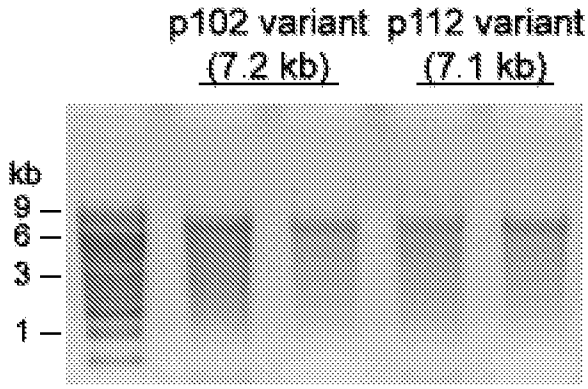


FIG. 12B

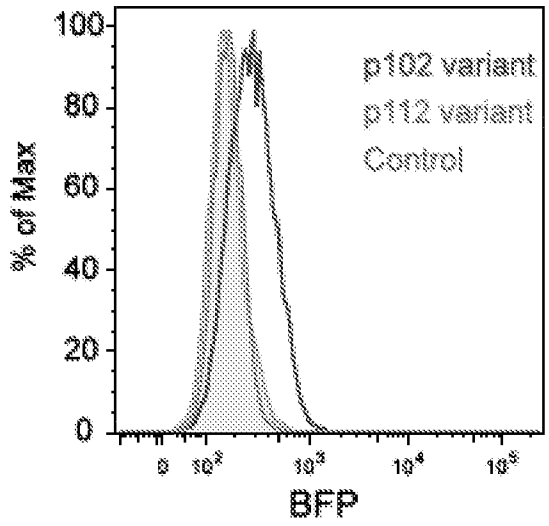


FIG. 12C

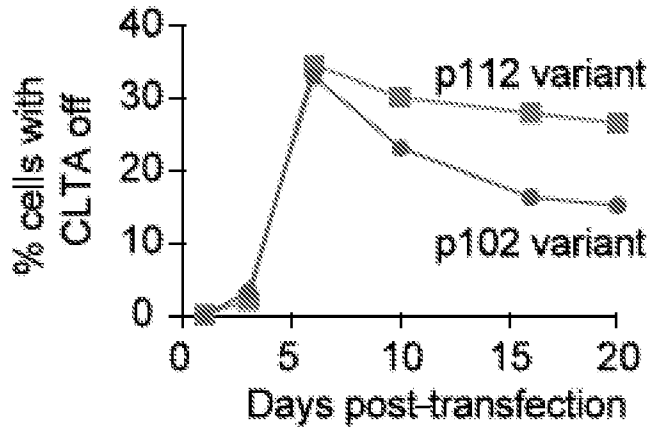


FIG. 13A

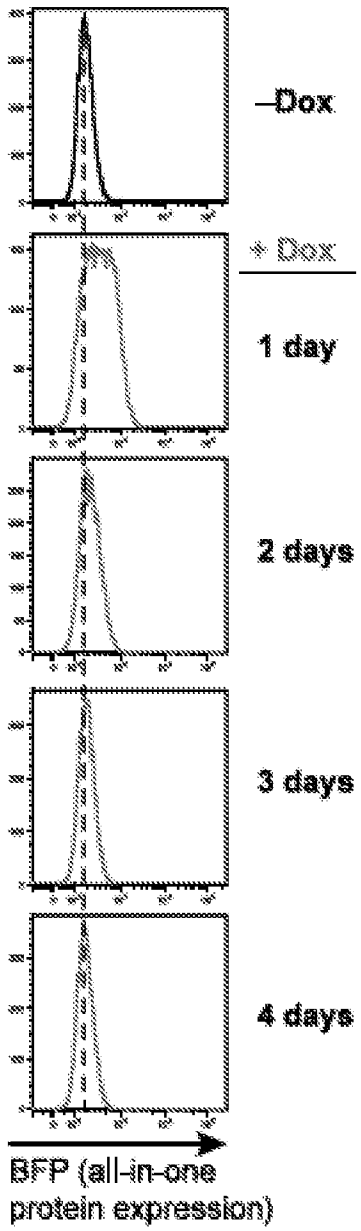

–Dox
+ Dox
1 day
2 days
3 days
4 days
BFP (all-in-one protein expression)

FIG. 13B

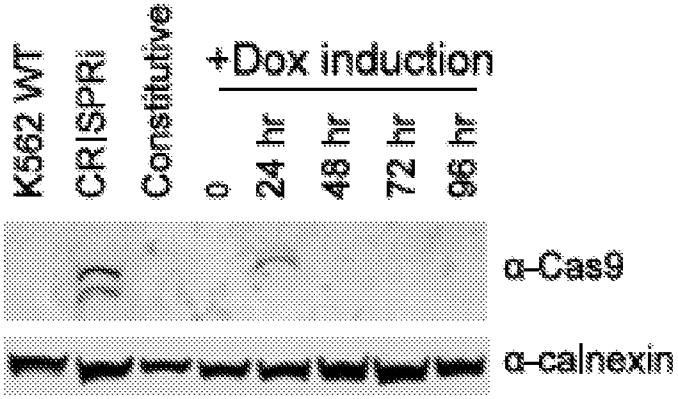


FIG. 13C

FIG. 13D

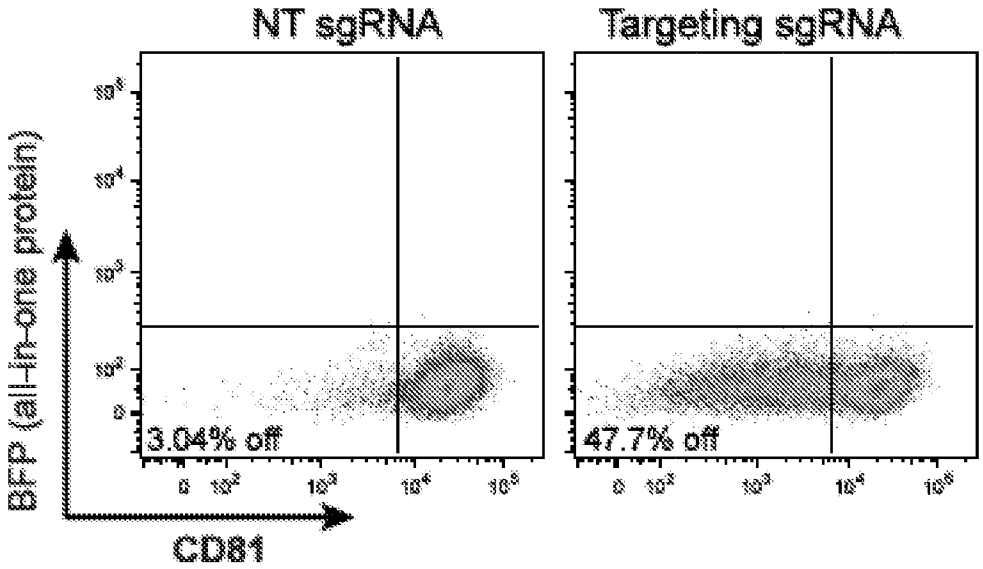


FIG. 13E

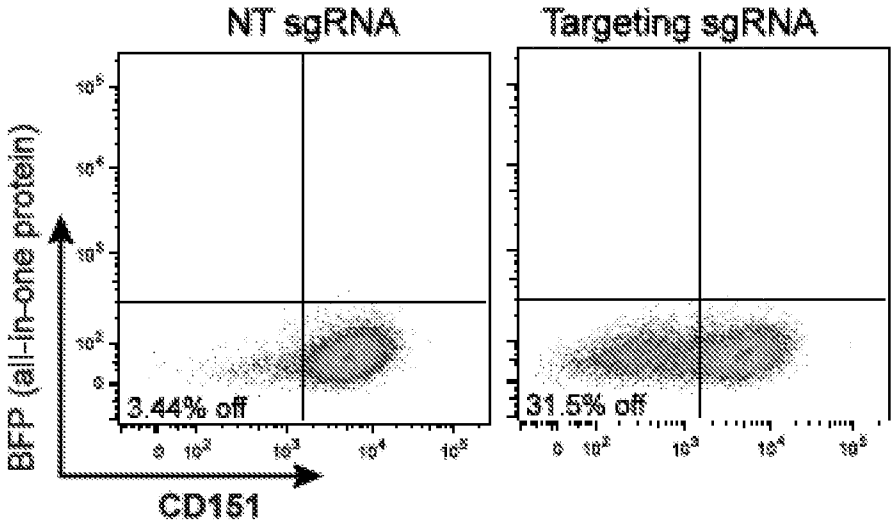
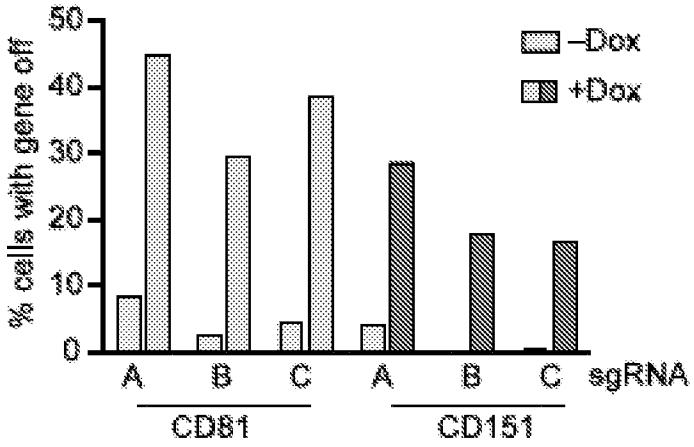


FIG. 13F

FIG. 13G



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/28377

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12Q 1/68; C12N 9/22, 15/63, 15/117, 15/85 (2019.01)

CPC - C12Q 1/68; C12N 9/22, 15/63, 15/117, 15/85

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2018/053035 A1 (THE JACKSON LABORATORY) 22 March 2018; paragraphs [0024], [0040], [0047], [0111]-[0112], [0132], [0175]-[0178], [0183], [0291]-[0292], [0299], [0307], [0323]; figures 1D, 4C; claims 1-6, 26-34	1-6, 8-10, 12-14, 17, 19-27, 28/1, 28/19, 28/21, 28/26, 29/28/1, 29/28/19, 29/28/21, 29/28/26, 30/28/1, 30/28/19, 30/28/21, 30/28/26, 31-42 --- 7, 11, 15-16
Y	(GUILINGER, JP et al.) Fusion of Catalytically Inactive Cas9 to FokI Nuclease Improves the Specificity of Genome Modification. Nature Biotechnology. 25 April 2014; Vol. 32, No. 6; pages 1-17; page 4, paragraph 1; DOI: 10.1038/nbt.2909	7, 11, 15-16
A	WO 2018/031762 A1 (DUKE UNIVERSITY) 15 February 2018; paragraph [0014]	18
A	(LIN, L et al.) Genome-Wide Determination of On-Target and Off-Target Characteristics for RNA-Guided DNA Methylation by Dcas9 Methyltransferases. Gigascience. 01 March 2018; Vol. 7, No. 3; pages 1-19; abstract; page 5, column 2, paragraph 2; DOI: 10.1093/gigascience/giy011	18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 August 2019 (26.08.2019)	10 SEP 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US19/28377

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

-***- Please See within the Next Supplemental Page-***-

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Groups I+, Claims 1-42 and SEQ ID NO: 1 (fusion protein)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US19/28377

-***-Continued from Box III Observations where unity of invention is lacking -***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I+, Claims 1-42 and SEQ ID NO: 1 (fusion protein) are directed toward a fusion protein, nucleic acid encoding the protein, vectors comprising the nucleic acid, cells comprising the nucleic acid or vector or protein or a complex comprising the protein; and methods associated therewith.

The fusion protein, nucleic acid, vector, cells, complexes, and methods will be searched to the extent they encompass a fusion protein encompassing SEQ ID NO: 1 (first exemplary fusion protein). Applicant is invited to elect additional fusion protein(s), with specified SEQ ID NO: for each, such that the sequence of each elected species is fully specified (i.e. no optional or variable residues or substituents), and available as an option within at least one searchable claim, to be searched. Additional fusion protein sequence(s) will be searched upon the payment of additional fees. It is believed that claims 1-17, 18 (in-part), and 19-42 encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass SEQ ID NO: 1 (fusion protein). Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be SEQ ID NO: 2 (fusion protein).

Groups I+ share the technical features including: a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Kruppel associated box domain, and a DNA methyltransferase domain; a nucleic acid sequence encoding the fusion protein; a vector comprising the nucleic acid sequence; a cell comprising the fusion protein, nucleic acid or vector; a complex comprising: (i) the fusion protein; and (ii) a polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for said nuclease-deficient RNA-guided DNA endonuclease enzyme, wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is bound to said polynucleotide via said binding sequence; a cell comprising the complex; and a method of silencing a target nucleic acid sequence in a cell, the method comprising delivering the complex to a cell containing said target nucleic acid; a method of silencing a target nucleic acid sequence in a cell, comprising: (i) delivering a first polynucleotide encoding the fusion protein to a cell containing said target nucleic acid; and (ii) delivering to said cell a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to said target nucleic acid sequence; and (b) a binding sequence for said nuclease-deficient RNA-guide DNA endonuclease enzyme.

However, these shared technical features are previously disclosed by the article 'Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing' by Amabile et al. (hereinafter 'Amabile') in view of US 2015/0232883 A1 to The Broad Institute et al. (hereinafter 'Broad').

Amabile discloses a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, and a Kruppel associated box domain (a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, and a Kruppel associated box domain; page 225, second column, first paragraph), and a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme and a DNA methyltransferase domain (a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme and a DNA methyltransferase domain; page 225, second column, first paragraph); a nucleic acid sequence encoding the fusion protein (pcDNA.CMV.T7.dCas9:KRAB, pcDNA.CMV.T7.dCas9:DNMT3A and pcDNA.CMV.T7.dCas9:DNMT3L (a nucleic acid sequence encoding the fusion protein); page e2); a vector comprising the nucleic acid sequence (pcDNA.CMV.T7.dCas9:KRAB, pcDNA.CMV.T7.dCas9:DNMT3A and pcDNA.CMV.T7.dCas9:DNMT3L (a vector comprising the nucleic acid sequence); page e2); a cell comprising the fusion protein, nucleic acid or vector (K-562 cells transfected with and expressing the fusion proteins (a cell comprising the fusion protein, nucleic acid or vector); page 225, second column, first paragraph); a complex comprising: (i) the fusion protein; and (ii) a polynucleotide (gene silencing effected by a combination of the fusion protein and sgRNAs targeting a B2M gene promoter) comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for said nuclease-deficient RNA-guided DNA endonuclease enzyme (comprising an sgRNA targeted to the promoter of the B2M gene (comprising: (a) a DNA-targeting sequence that is complementary to a target polynucleotide sequence; and (b) a binding sequence for said nuclease-deficient RNA-guided DNA endonuclease enzyme); Figure 4b, page 225, second column, first paragraph), wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is bound to said polynucleotide via said binding sequence (wherein said nuclease-deficient RNA-guided DNA endonuclease enzyme is bound to said promoter region (polynucleotide) polynucleotide via said sgRNA (binding sequence); Figure 4b, page 225, second column, first paragraph); a cell comprising the complex (a K-562 cell transfected with the expression plasmid for the fusion protein, the sgRNA, and the target polynucleotide (cell comprising the complex); Figure 4A, page 225, second column, first paragraph); and a method of silencing a target nucleic acid sequence in a cell (a method of silencing a target nucleic acid sequence in a cell; abstract, page 225, second column, first paragraph), the method comprising delivering the complex to a cell containing said target nucleic acid (the method comprising delivering the expression plasmid and sgRNA (complex) to a cell containing said target nucleic acid; Figure 4C, page 225, second column, first paragraph); and a method of silencing a target nucleic acid sequence in a cell (a method of silencing a target nucleic acid sequence in a cell; abstract, page 225, second column, first paragraph), comprising: (i) delivering a first polynucleotide encoding the fusion protein to a cell containing said target nucleic acid (comprising: (i) delivering a plasmid (first polynucleotide) encoding the fusion protein to a K-529 cell containing said target nucleic acid; page 225, second column, first paragraph); and (ii) delivering to said cell a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to said target nucleic acid sequence; and (b) a binding sequence for said nuclease-deficient RNA-guide DNA endonuclease enzyme (delivering to said cell an sgRNA for the target polynucleotide (a second polynucleotide comprising: (a) a DNA-targeting sequence that is complementary to said target nucleic acid sequence; and (b) a binding sequence for said nuclease-deficient RNA-guide DNA endonuclease enzyme); Figure 4B, page 225, second column, first paragraph). Ambile further discloses wherein the combination of the fusions including each of the Kruppel domain, the DNMT3a and the DNMT3L domains were required for efficient silencing (wherein the combination of the fusions including each of the Kruppel domain, the DNMT3a and the DNMT3L domains were required for efficient silencing; page 225, second column, first paragraph; Figures S4C and S4D).

-***-Continued Within the Next Supplemental Box-***-

-***-Continued from Previous Supplemental Page:

Amabile does not disclose a fusion protein comprising a nuclease-deficient RNA-guided DNA endonuclease enzyme, a Kruppel associated box domain, and a DNA methyltransferase domain.

Broad discloses delivery particle formulations or systems comprising one or more components of a CRISPR-Cas system (delivery particle formulations or systems comprising one or more components of a CRISPR-Cas system; abstract), including an enzymatically inactive Cas polypeptide fused to more than one heterologous domain (including an enzymatically inactive Cas polypeptide fused to more than one heterologous domain; paragraphs [0442], [0643]), fused with one or more heterologous protein domains, including 2 or 3 or more domains in addition to the CRISPR enzyme (fused with one or more heterologous protein domains, including 2 or 3 or more domains in addition to the CRISPR enzyme; paragraph [0442]); wherein the domains include a domain with methyltransferase activity (wherein the domains include a domain with methyltransferase activity; paragraph [0442]), or transcription repression activity (or transcription repression activity; paragraph [0442]), such as a Kruppel domain (such as a KRAB (Kruppel) domain; paragraph [0643]).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified the disclosure of Amabile to have provided the fusion protein as a fusion including all 3 required domains, as disclosed by Broad, in order to better enable targeting of all of the polypeptides to the desired site at the same time in order to enhance the effectiveness of the silencing.

Since none of the special technical features of the Groups I+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by a combination of the Amabile and Broad references, unity of invention is lacking.