USING NUCLEOSOME INTERACTING PROTEIN DOMAINS TO ENHANCE TARGETED GENOME MODIFICATION

FIELD

[0001] The present disclosure relates to compositions and methods for increasing the efficiency of targeted genome modification, targeted transcriptional regulation, or targeted epigenetic modification.

BACKGROUND

[0002] Programmable endonucleases have increasingly become an important tool for targeted genome engineering or modification in eukaryotes. Recently, RNA-guided clustered regularly interspersed short palindromic repeats (CRISPR) systems have emerged as a new generation of genome modification tools. These new programmable endonucleases provided unprecedented simplicity and versatility as compared to previous generations of nucleases such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). However, chromatin barriers in eukaryotic cells can hinder target access and cleavage by the prokaryote-derived CRISPR systems (Hinz et al, Biochemistry, 2015, 54:7063-66; Horlbeck et al., eLife, 2016, 5:e12677).

[0003] In fact, no or low editing activity on certain mammalian genomic sites has been observed when using *Streptococcus pyogenes* Cas9 (SpCas9), which is considered the most active CRISPR nuclease to date. Moreover, many of the CRISPR nucleases that have been characterized thus far exhibit no activity in mammalian cells, even though they are active in bacteria or on purified DNA substrates. Therefore, there is a need to improve the ability of CRISPR nuclease systems and other programmable DNA modification proteins to overcome chromatin hindrance to increase the efficiency of targeted genome or epigenetic modification in eukaryotes.

SUMMARY

[0004] Among the various aspects of the present disclosure is the provision of fusion proteins, each fusion protein comprising at least one nucleosome interacting protein domain linked to a programmable DNA modification protein.

[0005] The at least one nucleosome interacting protein domain can be a DNA binding domain from a high mobility group (HMG) box (HMGB) protein chosen from HMGB1, HMGB2, or HMGB3; a HMG nucleosome-binding (HMGN) protein chosen from HMGN1, HMGN2, HMGN3a, HMGN3b, HMGN4, or HMGN5; a central globular domain from a histone H1 variant; a DNA binding domain from a chromatin remodeling complex protein chosen from switch/sucrose non-fermentable (SWI/SNF) complex, imitation switch (ISWI) complex, chromodomain-helicase-DNA binding (CHD) complex, nucleosome remodeling and deacetylase (NuRD) complex, INO80 complex, SWR1 complex, RSC complex, or combination thereof. In some embodiments, the at least one nucleosome interacting protein domain can be HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular domain, ISWI protein DNA binding domain, CHD1 protein DNA binding domain, or combination thereof.

[0006] In some embodiments, the programmable DNA modification protein has nuclease activity, and the programmable DNA modification protein can be a clustered regularly interspersed short palindromic repeats (CRISPR) nuclease or nickase, a zinc finger nuclease (ZFN), a transcription activator-like effector nuclease (TALEN), a meganuclease, or a chimeric protein comprising a programmable DNA binding domain linked to a nuclease domain.

[0007] In other embodiments, the programmable DNA modification protein has non-nuclease activity, and the programmable DNA modification protein can be a chimeric protein comprising a programmable DNA binding domain linked to a non-nuclease domain. The programmable DNA binding domain of the chimeric protein can be a CRISPR protein modified to lack all nuclease activity, a zinc finger protein, or a transcription activator-like effector, and the non-nuclease domain of the chimeric protein can have acetyltransferase activity, deacetylase activity, methyltransferase activity,

demethylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, demyristoylation activity, citrullination activity, helicase activity, amination activity, deamination activity, alkylation activity, dealkylation activity, oxidation activity, transcriptional activation activity, or transcriptional repressor activity. In certain embodiments, the non-nuclease domain of the chimeric protein can have cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0008] The at least one nucleosome interacting protein domain can be linked to programmable DNA modification protein directly via a chemical bond, indirectly via a linker, or a combination thereof. The at least one nucleosome interacting protein domain can be linked to the N-terminus, C-terminus, and/or an internal location of the programmable DNA modification protein. In some embodiments, the fusion protein comprises at least two nucleosome interacting protein domains linked to the programmable DNA modification protein.

[0009] The fusion proteins disclosed herein can further comprise at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof.

[0010] Another aspect of the present disclosure encompasses a fusion protein comprising a CRISPR protein linked to at least one nucleosome interacting protein domain.

[0011] In general, the CRISPR protein of said fusion protein can be a type II CRISPR/Cas9 protein or a type V CRISPR/Cpf1 protein. In certain embodiments, the CRISPR protein can be *Streptococcus pyogenes* Cas9 (SpCas9), *Streptococcus thermophilus* Cas9 (StCas9), *Streptococcus pasteurianus* (SpaCas9), *Campylobacter jejuni* Cas9 (CjCas9), *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9), *Neisseria cinerea* Cas9 (NcCas9), *Neisseria meningitis* Cas9 (NmCas9), *Francisella novicida* Cpf1 (FnCpf1), *Acidaminococcus sp.* Cpf1 (AsCpf1), or *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1).

[0012] In some embodiments, the CRISPR protein has nuclease or nickase activity. For example, the CRISPR protein can be a type II CRISPR/Cas9 nuclease or nickase, or a type V CRISPR/Cpf1 nuclease or nickase. In other embodiments, the CRISPR protein has non-nuclease activity. In such iterations, the CRISPR protein can be a type II CRISPR/Cas9 protein modified to lack all nuclease activity and linked to a non-nuclease domain, or a type V CRISPR/Cpf1 protein modified to lack all nuclease activity and linked to a non-nuclease domain, wherein the non-nuclease domain can have cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0013] The at least one nucleosome interacting protein domain of the CRISPR fusion protein can be a high mobility group (HMG) box (HMGB) DNA binding domain, a HMG nucleosome-binding (HMGN) protein, a central globular domain from a histone H1 variant, a DNA binding domain from a chromatin remodeling complex protein, or a combination thereof. In certain embodiments, the at least one nucleosome interacting protein domain of the CRISPR fusion protein can be HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular domain, imitation switch (ISWI) protein DNA binding domain, chromodomain-helicase-DNA protein 1 (CHD1) DNA binding domain, or a combination thereof.

[0014] The at least one nucleosome interacting protein domain can be linked to the CRISPR protein directly via a chemical bond, indirectly via a linker, or a combination thereof. The at least one nucleosome interacting protein domain can be linked to the N-terminus, C-terminus, and/or an internal location of the CRISPR protein. In some embodiments, the fusion protein comprises at least two nucleosome interacting protein domains linked to the CRISPR protein.

[0015] The CRISPR fusion proteins disclosed herein can further comprise at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof.

[0016] In certain embodiments, the CRISPR fusion protein can have an amino acid sequence having at least about 90% sequence identity with SEQ ID NO:61,

SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

[0017] In other embodiments, the CRISPR fusion protein can have an amino acid sequence as set forth in SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

[0018] Another aspect of the present disclosure encompasses protein-RNA complexes comprising at least one of the CRISPR-containing fusion proteins disclosed herein and at least one guide RNA.

[0019] A further aspect of the present disclosure provides nucleic acids encoding any of the fusion proteins disclosed herein. The nucleic acids can be codon optimized for translation in a eukaryotic cell. In some embodiments, the nucleic acids can be part of a vector such as, for example, a viral vector, a plasmid vector, or a self-replicating RNA.

[0020] Still another aspect of the present disclosure provides methods for increasing efficiency of targeted genome or epigenetic modification in a eukaryotic cell. The methods involve introducing into a eukaryotic cell (a) at least one fusion protein as disclosed herein or nucleic acid encoding said fusion protein(s), wherein the at least one nucleosome interacting protein domain of the at least one fusion protein alters nucleosomal or chromatin structure such that the at least one fusion protein has increased access to the target chromosomal sequence, thereby increasing efficiency of targeted genome or epigenetic modification.

[0021] In some iterations, the method comprises introducing into the eukaryotic cell (a) at least one CRISPR fusion protein as disclosed herein or nucleic acid encoding said CRISPR fusion protein(s), wherein the CRISPR protein (i) has nuclease or nickase activity or (ii) is modified to lack all nuclease activity and is linked to a non-nuclease domain; and (b) at least one guide RNA or nucleic acid encoding at

least one guide RNA; wherein the CRISPR protein of the at least one CRISPR fusion protein is targeted to a target chromosomal sequence and the at least one nucleosome interacting protein domain of the at least one CRISPR fusion protein alters nucleosomal or chromatin structure such that the at least one CRISPR fusion protein has increased access to the target chromosomal sequence, thereby increasing efficiency of targeted genome or epigenetic modification.

[0022] In certain embodiments, the methods can further comprise introducing into the eukaryotic cell at least one donor polynucleotide, the donor polynucleotide comprising at least one donor sequence.

[0023] The eukaryotic cells used in the methods disclosed herein can be mammalian cells. In some embodiments, the cells can be human cell. The cells can be *in vitro* or *in vivo*.

[0024] Other aspects and features of the disclosure are detailed below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] FIG. **1** presents the cleavage efficiency (as the percent of indels) of wild-type CjCas9 (CjeCas9), a fusion protein comprising CjCas9 linked to HMGN1 and HMGB1 box A (CjeCas9-HN1HB1), and a fusion protein comprising CjCas9 linked to HMGN1 and Histone H1 central globular domain (CjeCas9-HN1H1G) in the presence of wild-type sgRNA scaffold or modified sgRNA scaffold.

DETAILED DESCRIPTION

[0026] The present disclosure provides compositions and methods for increasing the accessibility of chromosomal DNA to programmable DNA modification proteins including CRISPR systems. In particular, the present disclosure provides fusion proteins comprising at least one nucleosome interacting protein domain linked to a programmable DNA modification protein. The nucleosome interacting protein domains alter or remodel nucleosomal and/or chromatin structure such that the programmable DNA modification protein has increased access to targeted

chromosomal sequences, thereby increasing the efficiency of targeted genome modification, targeted transcriptional regulation, or targeted epigenetic modification.

(I) Fusion Proteins

[0027] One aspect of the present disclosure provides fusion proteins, wherein each fusion protein comprises at least one nucleosome interacting protein domain linked to a programmable DNA modification protein. The programmable DNA modification protein can have nuclease activity (see section (I)(b)(i), below) or non-nuclease activity (see section (I)(b)(ii) below). Nucleosome interacting protein domains are described below in section (I)(a) and linkages between the domains are described below in section (I)(c).

(a) Nucleosome Interacting Protein Domains

[0028] Nucleosome interacting protein domains refer to chromosomal proteins or fragments thereof that interact with nucleosome and/or chromosomal proteins to facilitate nucleosome rearrangement and/or chromatin remodeling. In some embodiments, the nucleosome interacting protein domain can be derived from high mobility group (HMG) box (HMGB) proteins. In other embodiments, the nucleosome interacting protein domain can be HMG nucleosome-binding (HMGN) proteins or fragments thereof. In further embodiments, the nucleosome interacting protein domain can be derived from linker histone H1 variants. In still other embodiments, the nucleosome interacting protein domain can be derived from chromatin remodeling complex proteins.

(i) HMGB Proteins

[0029] In some embodiments, the at least one nucleosome interacting protein domain can be derived from an HMGB protein. HMGB proteins interact with nucleosomes and other chromosomal proteins to regulate chromatin structure and function. Suitable HMGB proteins include mammalian HMGB1, mammalian HMGB2, and mammalian HMGB3. For example, the nucleosome interacting protein domain can

be derived from a human HNGB1 (RefSeqGene, U51677), human HMGB2 (RefSeqGene, M83665), or human HMGB3 (RefSeqGene, NM_005342). In other embodiments, the nucleosome interacting protein domain can be derived from an HMGB protein or HMGB-like protein from other vertebrates, invertebrates (*e.g.*, Drosophila DSP1), plants, yeast, or other single cell eukaryotes.

[0030] In specific embodiments, the at least one nucleosome interacting protein domain can be a fragment of an HMGB protein. In particular, the fragment of the HMGB protein is a DNA-binding domain. HMGB proteins typically contain two DNA-binding domains, which are called box A and box B. In some embodiments, the nucleosome interacting domain can be a box A domain or a box B domain from a HMGB protein. In specific embodiments, the nucleosome interacting domain can be a HMGB1 box A domain, a HMGB2 box A domain, or a HMGB3 box A domain.

(ii) HMGN Proteins

[0031] In other embodiments, the at least one nucleosome interacting protein domain can be a HMGN protein or fragment thereof. HMGN proteins are chromosomal proteins that modulate the structure and function of chromatin. Suitable mammalian HMGN proteins include HMGN1, HMGN2, HMGN3, HMGN3, HMGN4, and HMGN5. In various embodiments, the nucleosome interacting protein domain can be human HMGN1 (RefSeqGene, M21339), human HMGN2 (RefSeqGene, X13546), human HMGN3a or human HMGN3b (RefSeqGene, L40357), human HMGN4 (RefSeqGene, NM_030763), human HMGN5 (RefSeqGene, NM_016710), a fragment thereof, or a derivative thereof. In other embodiments, the nucleosome interacting protein domain can be a non-human HMGN protein, fragment, or derivative thereof. HMGN proteins are relatively small proteins. As such, the entire HMGN protein can be linked to the programmable DNA modification protein. In some embodiments, however, a fragment (e.g., the centrally-located nucleosome-binding domain) of a HMGN protein can be linked to the programmable DNA modification protein.

(iii) Histone H1 Variants

[0032] In still other embodiments, the at least one nucleosome interacting protein domain can be derived from a linker histone H1 variant. For example, nucleosome interacting protein domain can be a central globular domain from a histone H1 variant. Histone H1 variants bind to the linker DNA between nucleosomes and the central globular domain (of about 80 amino acids) binds to the linker DNA at the nucleosome entry and exit sites close to the nucleosome dyad. Histone H1 variants comprise a large family of related proteins with distinct specificity for tissues, developmental stages, and organisms in which they are expressed. For example, human and mouse contain 11 histone H1 variants, chicken has six variants (which are called histone H5), frog has five variants, nematode has eight variants, fruit fly species have from one to three variants, and tobacco has six variants. In some embodiments, the histone H1 variant can be a human variant as shown below.

Protein name*	Gene Symbol	UniProtKB Accession
Histone H1.0	H1F0	P07305
Histone H1.1	HIST1H1A	Q02539
Histone H1.2	HIST1H1C	P16403
Histone H1.3	HIST1H1D	P16402
Histone H1.4	HIST1H1E	P10412
Histone H1.5	HIST1H1B	P16401
Histone H1.6 (testis specific)	HIST1H1T	P22492
Histone H1.7 (testis specific)	H1FNT	Q75WM6
Histone H1.8 (oocyte specific)	H1FOO	Q8IZA3
Histone H1.9 (testis specific)	HILS1	P60008
Histone H1.10	H1FX	Q92522

* Talbert *et al.*, Epigenetics & Chromatin, 2012, 5:7.

(iv) Chromatin Remodeling Complex Proteins

[0033] In further embodiments, the at least one nucleosome interacting protein domain can be derived from a chromatin remodeling complex protein. For example, the nucleosome interacting protein domain can be DNA binding domain from a chromatin remodeling complex protein. Chromatin remodeling complexes are multi-subunit enzyme complexes with the capacity to remodel the structure of chromatin. These remodeling complexes use the energy of ATP hydrolysis to move, destabilize, eject, or restructure nucleosomes.

[0034] Examples of chromatin remodeling complexes include SWI/SNF (SWItch/Sucrose Non-Fermentable), ISWI (Imitation SWItch), CHD (Chromodomain-Helicase-DNA binding), Mi-2/NuRD (Nucleosome Remodeling and Deacetylase), INO80, SWR1, and RSC complexes. In various embodiments, the nucleosome interacting protein domain can be derived from an ATPase, a helicase, and/or a DNA binding protein in the chromatin remodeling complex. In some embodiments, the nucleosome interacting protein domain can be derived from the ATPase ISWI from the ISWI complex, the DNA-binding protein CHD1 from the CHD complex, the ATP-dependent helicase SMARCA4 or the ATPase Snf2 from the SWI/SNF complex, ATPase Mi-2α or ATPase Mi2-β of the Mi-1/NuRD complex, the RuvB-like AAA ATPase 1 or the RuvB-like AAA ATPase 2 from the INO80 complex, the ATPase Swr1 from the SWR1 complex, or the ATPase Rsc1 or ATPase Rcs2 from the RSC complex. In specific embodiments, the nucleosome interacting protein domain can be a DNA binding domain from ISWI protein or a DNA binding domain from CHD1 protein.

(b) Programmable DNA Modification Proteins

[0035] A programmable DNA modification protein is a protein targeted to bind a specific sequence in chromosomal DNA, where it modifies the DNA or a protein associated with the DNA at or near the targeted sequence. Thus, a programmable DNA

modification protein comprises a programmable DNA binding domain and a catalytically active modification domain.

[0036] The DNA binding domain of the programmable DNA modification proteins is programmable, meaning that it can be designed or engineered to recognize and bind different DNA sequences. In some embodiments, for example, DNA binding is mediated by interactions between the DNA modification protein and the target DNA. Thus, the DNA binding domain can be programed to bind a DNA sequence of interest by protein engineering. In other embodiments, for example, DNA binding is mediated by a guide RNA that interacts with the DNA modification protein and the target DNA. In such instances, the programmable DNA binding protein can be targeted to a DNA sequence of interest by designing the appropriate guide RNA.

[0037] A variety of modification domains can be included in the programmable DNA modification protein. In some embodiments, the modification domain has nuclease activity and can cleave one or both strands of a double-stranded DNA sequence. The DNA break can then be repaired by a cellular DNA repair process such as non-homologous end-joining (NHEJ) or homology-directed repair (HDR), such that the DNA sequence can be modified by a deletion, insertion, and/or substitution of at least one base pair. Examples of programmable DNA modification proteins having nuclease activity include, without limit, CRISPR nucleases (or nickases), zinc finger nucleases, transcription activator-like effector nucleases, meganucleases, and a programmable DNA binding domain linked to a nuclease domain. Programmable DNA modification proteins having nuclease activity are detailed below in section (I)(b)(i).

[0038] In other embodiments, the modification domain of the programmable DNA modification protein has non-nuclease activity (*e.g.*, epigenetic modification activity or transcriptional regulation activity) such that the programmable DNA modification protein modifies the structure and/or activity of the DNA and/or protein(s) associated with the DNA. Thus, the programmable DNA modification protein can comprise a programmable DNA binding domain linked to a non-nuclease domain linked. Such proteins are detailed below in section (I)(b)(ii).

[0039] The programmable DNA modification proteins can comprise wild-type or naturally-occurring DNA binding and/or modification domains, modified versions of naturally-occurring DNA binding and/or modification domains, synthetic or artificial DNA binding and/or modification domains, and combinations thereof.

(i) Programmable DNA Modification Proteins with Nuclease Activity

[0040] Examples of programmable DNA modification proteins having nuclease activity include, without limit, CRISPR nucleases, zinc finger nucleases, transcription activator-like effector nucleases, meganucleases, and programmable DNA binding domains linked nuclease domains.

[0041] *CRISPR Nucleases*. The CRISPR nuclease can be derived from a type I, type II (*i.e.*, Cas9), type III, type V (i.e., Cpf1), or type VI (i.e., Cas13) CRISPR protein, which are present in various bacteria and archaea. In further embodiments, the CRISPR nuclease can be derived from an archaeal CRISPR system, a CRISPR/CasX system, or a CRISPR/CasY system (Burstein *et al.*, Nature, 2017, 542(7640):237-241). In various embodiments, the CRISPR nuclease can be from *Streptococcus sp.* (*e.g., S. pyogenes, S. thermophilus, S. pasteurianus*), *Campylobacter sp.* (*e.g., Campylobacter jejuni*), *Francisella sp.* (*e.g., Francisella novicida*), *Acaryochloris sp., Acetohalobium sp., Acidaminococcus sp., Acidithiobacillus sp., Alicyclobacillus sp., Allochromatium sp., Ammonifex sp., Anabaena sp., Arthrospira sp., Bacillus sp., Burkholderiales sp., Caldicelulosiruptor sp., Candidatus sp., Clostridium sp., Crocosphaera sp., Cyanothece sp., Exiguobacterium sp., Finegoldia sp., Ktedonobacter sp., Lachnospiraceae sp., Lactobacillus sp., Lyngbya sp., Marinobacter sp., Methanohalobium sp., Microscilla sp., Microcoleus sp., Microcystis sp., Natranaerobius sp., Neisseria sp., Nitrosococcus sp., Nocardiopsis sp., Nodularia sp., Nostoc sp., Oscillatoria sp., Polaromonas sp., Pelotomaculum sp., Pseudoalteromonas sp., Petrotoga sp., Prevotella sp., Staphylococcus sp., Streptomyces sp., Streptosporangium sp., Synechococcus sp., Thermosipho sp.*, or *Verrucomicrobia* sp.

[0042] The CRISPR nuclease can be a wild type or naturally-occurring protein. Alternatively, the CRISPR nuclease can be engineered to have improved specificity, altered PAM specificity, decreased off-target effects, increased stability, and the like.

[0043] In some embodiments, the CRISPR nuclease can be a type II CRISPR/Cas 9 protein. For example, the CRISPR nuclease can be *Streptococcus pyogenes* Cas9 (SpCas9), *Streptococcus thermophilus* Cas9 (StCas9), *Streptococcus pasteurianus* (SpaCas9), *Campylobacter jejuni* Cas9 (CjCas9), *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9), *Neisseria cinerea* Cas9 (NcCas9), or *Neisseria meningitis* Cas9 (NmCas9). In other embodiments, the CRISPR nuclease can be a type V CRISPR/Cpf1 protein, e.g., *Francisella novicida* Cpf1 (FnCpf1), *Acidaminococcus sp.* Cpf1 (AsCpf1), or *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1). In further embodiments, the CRISPR nuclease can be a type VI CRISPR/Cas13 protein, *e.g.*, *Leptotrichia wadei* Cas13a (LwaCas13a) or *Leptotrichia shahii* Cas13a (LshCas13a).

[0044] In general, the CRISPR nuclease comprises at least one nuclease domain having endonuclease activity. For example, a Cas9 nuclease comprises a HNH domain, which cleaves the guide RNA complementary strand, and a RuvC domain, which cleaves the non-complementary strand, a Cpf1 protein comprises a RuvC domain and a NUC domain, and a Cas13a nuclease comprises two HNEPN domains. In some embodiments, both nuclease domains are active and the CRISPR nuclease has double-stranded cleavage activity (i.e., cleaves both strands of a double-stranded nucleic acid sequence). In other embodiments, one of the nuclease domains is inactivated by one or more mutations and/or deletions, and the CRISPR variant is a nickase that cleaves one strand of a double-stranded nucleic acid sequence. For example, one or more mutations in the RuvC domain of Cas9 protein (*e.g.*, D10A, D8A, E762A, and/or D986A) results in an HNH nickase that nicks the guide RNA complementary strand; and one or more mutations in the HNH domain of Cas9 protein (*e.g.*, H840A, H559A, N854A, N856A, and/or N863A) results in a RuvC nickase that nicks the guide RNA non-

complementary strand. Comparable mutations can convert Cpf1 and Cas13a nucleases to nickases.

[0045] *Zinc Finger Nucleases*. In still other embodiments, the programmable DNA modification protein having nuclease activity can be a pair of zinc finger nucleases (ZFN). A ZFN comprises a DNA binding zinc finger region and a nuclease domain. The zinc finger region can comprise from about two to seven zinc fingers, for example, about four to six zinc fingers, wherein each zinc finger binds three consecutive base pairs. The zinc finger region can be engineered to recognize and bind to any DNA sequence. Zinc finger design tools or algorithms are available on the internet or from commercial sources. The zinc fingers can be linked together using suitable linker sequences.

[0046] A ZFN also comprises a nuclease domain, which can be obtained from any endonuclease or exonuclease. Non-limiting examples of endonucleases from which a nuclease domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. In some embodiments, the nuclease domain can be derived from a type II-S restriction endonuclease. Type II-S endonucleases cleave DNA at sites that are typically several base pairs away from the recognition/binding site and, as such, have separable binding and cleavage domains. These enzymes generally are monomers that transiently associate to form dimers to cleave each strand of DNA at staggered locations. Non-limiting examples of suitable type II-S endonucleases include BfiI, BpmI, BsaI, BsgI, BsmBI, BsmI, BspMI, FokI, MboII, and SapI. In some embodiments, the nuclease domain can be a FokI nuclease domain or a derivative thereof. The type II-S nuclease domain can be modified to facilitate dimerization of two different nuclease domains. For example, the cleavage domain of FokI can be modified by mutating certain amino acid residues. By way of non-limiting example, amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of FokI nuclease domains are targets for modification. In specific embodiments, the FokI nuclease domain can comprise a first FokI half- domain comprising Q486E, I499L, and/or N496D mutations, and a second FokI half-domain comprising E490K, I538K, and/or H537R mutations. In

some embodiments, the ZFN has double-stranded cleavage activity. In other embodiments, the ZFN has nickase activity (*i.e.*, one of the nuclease domains has been inactivated).

[0047] *Transcription Activator-like Effector Nucleases*. In alternate embodiments, the programmable DNA modification protein having nuclease activity can be a transcription activator-like effector nuclease (TALEN). TALENs comprise a DNA binding domain composed of highly conserved repeats derived from transcription activator-like effectors (TALEs) that is linked to a nuclease domain. TALEs are proteins secreted by plant pathogen *Xanthomonas* to alter transcription of genes in host plant cells. TALE repeat arrays can be engineered via modular protein design to target any DNA sequence of interest. The nuclease domain of TALENs can be any nuclease domain as described above in the subsection describing ZFNs. In specific embodiments, the nuclease domain is derived from FokI (Sanjana *et al.*, 2012, Nat Protoc, 7(1):171-192). The TALEN can have double-stranded cleavage activity or nickase activity.

[0048] *Meganucleases or Rare-Cutting Endonucleases*. In still other embodiments, the programmable DNA modification protein having nuclease activity can be a meganuclease or derivative thereof. Meganucleases are endodeoxyribonucleases characterized by long recognition sequences, *i.e.*, the recognition sequence generally ranges from about 12 base pairs to about 45 base pairs. As a consequence of this requirement, the recognition sequence generally occurs only once in any given genome. Among meganucleases, the family of homing endonucleases named LAGLIDADG has become a valuable tool for the study of genomes and genome engineering. In some embodiments, the meganuclease can be I-SceI, I-TevI, or variants thereof. A meganuclease can be targeted to a specific chromosomal sequence by modifying its recognition sequence using techniques well known to those skilled in the art. In alternate embodiments, the programmable DNA modification protein having nuclease activity can be a rare-cutting endonuclease or derivative thereof. Rare-cutting endonucleases are site-specific endonucleases whose recognition sequence occurs rarely in a genome, preferably only once in a genome. The rare-cutting endonuclease

may recognize a 7-nucleotide sequence, an 8-nucleotide sequence, or longer recognition sequence. Non-limiting examples of rare-cutting endonucleases include NotI, AscI, PacI, AsiSI, SbfI, and FseI.

[0049] *Programmable DNA Binding Domains Linked to Nuclease Domains.* In yet additional embodiments, the programmable DNA modification protein having nuclease activity can be a chimeric protein comprising a programmable DNA binding domain linked to a nuclease domain. The nuclease domain can be any of those described above in the subsection describing ZFNs (*e.g.*, the nuclease domain can be a FokI nuclease domain), a nuclease domain derived from a CRISPR nuclease (*e.g.*, RuvC or HNH nuclease domains of Cas9), or a nuclease domain derived from a meganuclease or rare-cutting endonuclease.

[0050] The programmable DNA binding domain of the chimeric protein can be any programmable DNA binding protein such as, *e.g.*, a zinc finger protein or a transcription activator-like effector. Alternatively, the programmable DNA binding domain can be a catalytically inactive (dead) CRISPR protein that was modified by deletion or mutation to lack all nuclease activity. For example, the catalytically inactive CRISPR protein can be a catalytically inactive (dead) Cas9 (dCas9) in which the RuvC domain comprises a D10A, D8A, E762A, and/or D986A mutation and the HNH domain comprises a H840A, H559A, N854A, N865A, and/or N863A mutation. Alternatively, the catalytically inactive CRISPR protein can be a catalytically inactive (dead) Cpf1 protein comprising comparable mutations in the nuclease domains. In still other embodiments, the programmable DNA binding domain can be a catalytically inactive meganuclease in which nuclease activity was eliminated by mutation and/or deletion, *e.g.*, the catalytically inactive meganuclease can comprise a C-terminal truncation.

(ii) Programmable DNA Modification Proteins with Non-Nuclease Activity

[0051] In alternate embodiments, the programmable DNA modification protein can be a chimeric protein comprising programmable DNA binding domain linked to a non-nuclease domain. The programmable DNA binding domain can be a zinc finger protein, a transcription activator-like effector, a catalytically inactive (dead)

CRISPR protein, or a catalytically inactive (dead) meganuclease. For example, the catalytically inactive CRISPR protein can be a catalytically inactive (dead) Cas9 (dCas9) in which the RuvC domain comprises a D10A, D8A, E762A, and/or D986A mutation and the HNH domain comprises a H840A, H559A, N854A, N865A, and/or N863A mutation. Alternatively, the catalytically inactive CRISPR protein can be a catalytically inactive (dead) Cpf1 protein comprising comparable mutations in the nuclease domains.

[0052] In some embodiments, the non-nuclease domain of the chimeric protein can be an epigenetic modification domain, which alters DNA or chromatin structure (and may or may not alter DNA sequence). Non-limiting examples of suitable epigenetic modification domains include those with DNA methyltransferase activity (*e.g.*, cytosine methyltransferase), DNA demethylase activity, DNA deamination (*e.g.*, cytosine deaminase, adenosine deaminase, guanine deaminase), DNA amination, DNA oxidation activity, DNA helicase activity, histone acetyltransferase (HAT) activity (*e.g.*, HAT domain derived from E1A binding protein p300), histone deacetylase activity, histone methyltransferase activity, histone demethylase activity, histone kinase activity, histone phosphatase activity, histone ubiquitin ligase activity, histone deubiquitinating activity, histone adenylation activity, histone deadenylation activity, histone SUMOylating activity, histone deSUMOylating activity, histone ribosylation activity, histone deribosylation activity, histone myristoylation activity, histone demyristoylation activity, histone citrullination activity, histone alkylation activity, histone dealkylation activity, or histone oxidation activity. In specific embodiments, the epigenetic modification domain can comprise cytidine deaminase activity, histone acetyltransferase activity, or DNA methyltransferase activity.

[0053] In other embodiments, the non-nuclease modification domain of the chimeric protein can be a transcriptional activation domain or transcriptional repressor domain. Suitable transcriptional activation domains include, without limit, herpes simplex virus VP16 domain, VP64 (which is a tetrameric derivative of VP16), VP160, NFκB p65 activation domains, p53 activation domains 1 and 2, CREB (cAMP response element binding protein) activation domains, E2A activation domains, activation domain from human heat-shock factor 1 (HSF1), or NFAT (nuclear factor of activated T-cells)

activation domains. Non-limiting examples of suitable transcriptional repressor domains include inducible cAMP early repressor (ICER) domains, Kruppel-associated box (KRAB) repressor domains, YY1 glycine rich repressor domains, Sp1-like repressors, E(spl) repressors, IκB repressor, or methyl-CpG binding protein 2 (MeCP2) repressor. Transcriptional activation or transcriptional repressor domains can be genetically fused to the DNA binding protein or bound via noncovalent protein-protein, protein-RNA, or protein-DNA interactions.

[0054] In particular embodiments, the non-nuclease domain of the chimeric protein can comprise cytidine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0055] In some embodiments, the chimeric protein having non-nuclease activity can further comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

(c) Linkages

[0056] The fusion proteins disclosed herein comprise at least one nucleosome interacting protein domain linked to a programmable DNA modification protein. The linkage between the at least one nucleosome interacting protein domain and the programmable DNA modification protein can be direct via a chemical bond, or the linkage can be indirect via a linker.

[0057] In some embodiments, the at least one nucleosome interacting protein domain can be linked directly to the programmable DNA modification protein by a covalent bond (*e.g.*, peptide bond, ester bond, and the like). Alternatively, the chemical bond can be non-covalent (*e.g.*, ionic, electrostatic, hydrogen, hydrophobic, Van der interactions, or π-effects).

[0058] In other embodiments, the at least one nucleosome interacting protein domain can be linked to the programmable DNA modification protein by a linker. A linker is a chemical group that connects one or more other chemical groups via at

least one covalent bond. Suitable linkers include amino acids, peptides, nucleotides, nucleic acids, organic linker molecules (*e.g.*, maleimide derivatives, N-ethoxybenzylimidazole, biphenyl-3,4′,5-tricarboxylic acid, p-aminobenzyloxycarbonyl, and the like), disulfide linkers, and polymer linkers (*e.g.*, PEG). The linker can include one or more spacing groups including, but not limited to alkylene, alkenylene, alkynylene, alkyl, alkenyl, alkynyl, alkoxy, aryl, heteroaryl, aralkyl, aralkenyl, aralkynyl and the like. The linker can be neutral, or carry a positive or negative charge. Additionally, the linker can be cleavable such that the linker's covalent bond that connects the linker to another chemical group can be broken or cleaved under certain conditions, including pH, temperature, salt concentration, light, a catalyst, or an enzyme.

[0059] In still other embodiments, the at least one nucleosome interacting protein domain can be linked to the programmable DNA modification protein by peptide linkers. The peptide linker can be a flexible amino acid linker (*e.g.*, comprising small, non-polar or polar amino acids). Non-limiting examples of flexible linkers include LEGGGS (SEQ ID NO:1), TGSG (SEQ ID NO:2), GGSGGGSG (SEQ ID NO:3), $(GGGGS)_{1-4}$ (SEQ ID NO:4), and $(Gly)_{6-8}$ (SEQ ID NO:5). Alternatively, the peptide linker can be a rigid amino acid linker. Such linkers include $(EAAAK)_{1-4}$ (SEQ ID NO:6), $A(EAAAK)_{2-5}A$ (SEQ ID NO:7), PAPAP (SEQ ID NO:8), and $(AP)_{6-8}$ (SEQ ID NO:9). Examples of suitable linkers are well known in the art and programs to design linkers are readily available (Crasto *et al.*, Protein Eng., 2000, 13(5):309-312).

[0060] The at least one nucleosome interacting protein domain can be linked to N-terminus, the C-terminus, and/or an internal location of the programmable DNA modification protein. In some embodiments, at least one nucleosome interacting protein domain can be linked to N-terminus of the programmable DNA modification protein. In other embodiments, the at least one nucleosome interacting protein domain can be linked to C-terminus of the programmable DNA modification protein. In still other embodiments, at least one nucleosome interacting protein domain can be linked to N-terminus and at least one nucleosome interacting protein domain can be linked to C-terminus of the programmable DNA modification protein.

[0061] In some embodiments, the fusion protein can comprise one nucleosome interacting protein domain. In other embodiments, the fusion protein can comprise two nucleosome interacting protein domains. In still other embodiments, the fusion protein can comprise three nucleosome interacting protein domains. In additional embodiments, the fusion protein can comprise four, five, or more than five nucleosome interacting protein domains. The one or more nucleosome interacting protein domains can be the same or they can be different.

[0062] In embodiments in which the fusion protein comprises two or more nucleosome interacting protein domains, the two or more nucleosome interacting domains can be linked to either end, both ends, and/or an internal location of the programmable DNA modification protein. The two or more nucleosome interacting protein domains can be the same or they can be different. For example, the complex can comprise at least two HMG DNA-binding domains, at least two HMGN proteins, at least one HMG DNA-binding domain and at least one HMGN protein, at least one HMG DNA-binding domain or HMGN protein and at least one central domain from a histone H1 variant, at least one HMG DNA-binding domain or HMGN protein and at least one domain from a chromatin remodeling complex protein, at least one HMG DNA-binding domain or HMGN protein, at least one histone H1 variant central domain, and at least one domain from a chromatin remodeling complex protein, and the like.

(d) Optional Nuclear Localization Signal, Cell-Penetrating Domain, and/or Marker Domain

[0063] The fusion proteins disclosed herein can further comprise at least one nuclear localization signal, cell-penetrating domain, and/or marker domain.

[0064] Non-limiting examples of nuclear localization signals include PKKKRKV (SEQ ID NO:10), PKKKRRV (SEQ ID NO:11), KRPAATKKAGQAKKKK (SEQ ID NO:12), YGRKKRRQRRR (SEQ ID NO:13), RKKRRQRRR (SEQ ID NO:14), PAAKRVKLD (SEQ ID NO:15), RQRRNELKRSP (SEQ ID NO:16), VSRKRPRP (SEQ ID NO:17), PPKKARED (SEQ ID NO:18), PQPKKKPL (SEQ ID NO:19), SALIKKKKKMAP (SEQ ID NO:20), PKQKKRK (SEQ ID NO:21), RKLKKKIKKL (SEQ ID

NO:22), REKKKFLKRR (SEQ ID NO:23), KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:24), RKCLQAGMNLEARKTKK (SEQ ID NO:25), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:26), and RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:27).

[0065] Examples of suitable cell-penetrating domains include, without limit, GRKKRRQRRRPPQPKKKRKV (SEQ ID NO:28), PLSSIFSRIGDPPKKKRKV (SEQ ID NO:29), GALFLGWLGAAGSTMGAPKKKRKV (SEQ ID NO:30), GALFLGFLGAAGSTMGAWSQPKKKRKV (SEQ ID NO:31), KETWWETWWTEWSQPKKKRKV (SEQ ID NO:32), YARAAARQARA (SEQ ID NO:33), THRLPRRRRRR (SEQ ID NO:34), GGRRARRRRRR (SEQ ID NO:35), RRQRRTSKLMKR (SEQ ID NO:36), GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:37), KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:38), and RQIKIWFQNRRMKWKK (SEQ ID NO:39).

[0066] Marker domains include fluorescent proteins and purification or epitope tags. Suitable fluorescent proteins include, without limit, green fluorescent proteins (*e.g.*, GFP, eGFP, GFP-2, tagGFP, turboGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (*e.g.*, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), blue fluorescent proteins (*e.g.*, BFP, EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (*e.g.,* ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (*e.g.*, mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRasberry, mStrawberry, Jred), and orange fluorescent proteins (*e.g.*, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato). Non-limiting examples of suitable purification or epitope tags include 6xHis, FLAG®, HA, GST, Myc, and the like.

[0067] The at least one nuclear localization signal, cell-penetrating domain, and/or marker domain can be located at the N-terminus, the C-terminus, and/or in an internal location of the fusion protein.

(e) Specific Fusion Proteins

[0068] In general, the at least one nucleosome interacting protein domain of the fusion protein is chosen from HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular domain, imitation switch (ISWI) protein DNA binding domain, chromodomain-helicase-DNA protein 1 (CHD1) DNA binding domain, or combinations thereof.

[0069] In specific embodiments, the programmable DNA modification protein of the fusion protein is a CRISPR protein. For example, the CRISPR protein can be *Streptococcus pyogenes* Cas9 (SpCas9), *Streptococcus thermophilus* Cas9 (StCas9), *Streptococcus pasteurianus* (SpaCas9), *Campylobacter jejuni* Cas9 (CjCas9), *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9), *Neisseria cinerea* Cas9 (NcCas9), *Neisseria meningitis* Cas9 (NmCas9), *Francisella novicida* Cpf1 (FnCpf1), *Acidaminococcus sp.* Cpf1 (AsCpf1), or *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1).

[0070] In some embodiments, the fusion protein has an amino acid sequence having at least about 80% sequence identity with any of SEQ ID NOS:61-79. In general, any amino acid substitution is conservative, *i.e.*, limited to exchanges within members of group 1: glycine, alanine, valine, leucine, and Isoleucine; group 2: serine, cysteine, threonine, and methionine; group 3: proline; group 4: phenylalanine, tyrosine, and tryptophan; and group 5: aspartate, glutamate, asparagine, and glutamine. In various embodiments, the amino acid sequence of fusion protein has at least about 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98, or 99% sequence identity with any of SEQ ID NOS:61-79. In some embodiments, the fusion protein has an amino acid sequence as set forth in SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

(II) Complexes

[0071] Another aspect of the present disclosure encompasses complexes comprising at least one CRISPR system (*i.e.*, CRISPR protein and guide RNA) and at least one nucleosome interacting protein domain. In some embodiments, the at least one nucleosome interacting protein domain can be linked to the CRISPR protein of the CRISPR system (*i.e.*, the complex comprises a CRISPR fusion protein as described in section (I) above). In other embodiments, the at least one nucleosome interacting protein domain can be linked to the guide RNA of the CRISPR system. The linkage can be direct or indirect, essentially as described above in section (I)(c). For example, a nucleosome interacting protein domain can be linked to an RNA aptamer binding protein, and the guide RNA can comprise aptamer sequences, such that binding of the RNA aptamer binding protein to the RNA aptamer sequence links the nucleosome interacting protein domain to the guide RNA.

[0072] Nucleosome interacting protein domains are described above in section (I)(a), and CRISPR proteins are detailed above in section (I)(b). The CRISPR protein can have nuclease or nickase activity (*e.g.*, can be a type II CRISPR/Cas9, type V CRISPR/Cpf1, or type VI CRISPR/Cas13). For example, a complex can comprise a CRISPR nuclease, or a complex can comprise two CRISPR nickases. Alternatively, the CRISPR protein can be modified to lack all nuclease activity and linked to non-nuclease domains (*e.g.*, domains having cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity). In some embodiments, the non-nuclease domain also can be linked to an RNA aptamer binding protein.

[0073] A guide RNA comprises (i) a CRISPR RNA (crRNA) that contains a guide sequence at the 5' end that hybridizes with a target sequence and (ii) a transacting crRNA (tracrRNA) sequence that interacts with the CRISPR protein. The crRNA guide sequence of each guide RNA is different (*i.e.*, is sequence specific). The tracrRNA sequence is generally the same in guide RNAs designed to complex with a CRISPR protein from a particular bacterial species.

[0074] The crRNA guide sequence is designed to hybridize with a target sequence (*i.e.*, protospacer) that is bordered by a protospacer adjacent motif (PAM) in a double-stranded sequence. PAM sequences for Cas9 proteins include 5'-NGG, 5'-NGGNG, 5'-NNAGAAW, and 5'-ACAY, and PAM sequences for Cpf1 include 5'-TTN (wherein N is defined as any nucleotide, W is defined as either A or T, and Y is defined an either C or T). In general, the complementarity between the crRNA guide sequence and the target sequence is at least 80%, at least 85%, at least 90%, at least 95%, or at least 99%. In specific embodiments, the complementarity is complete (*i.e.*, 100%). In various embodiments, the length of the crRNA guide sequence can range from about 15 nucleotides to about 25 nucleotides. For example, the crRNA guide sequence can be about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In specific embodiments, the crRNA can be about 19, 20, 21, or 22 nucleotides in length.

[0075] The crRNA and tracrRNA comprise repeat sequences that form one or more one stem loop structures, which can interact with the CRISPR protein. The length of each loop and stem can vary. For example, the one or more loops can range from about 3 to about 10 nucleotides in length, and the one or more stems can range from about 6 to about 20 base pairs in length. The one or more stems can comprise one or more bulges of 1 to about 10 nucleotides.

[0076] The crRNA can range in length from about 25 nucleotides to about 100 nucleotides. In various embodiments, the crRNA can range in length from about 25 to about 50 nucleotides, from about 590 to about 75 nucleotides, or from about 75 to about 100 nucleotides. The tracrRNA can range in length from about 50 nucleotides to about 300 nucleotides. In various embodiments, the tracrRNA can range in length from about 50 to about 90 nucleotides, from about 90 to about 110 nucleotides, from about 110 to about 130 nucleotides, from about 130 to about 150 nucleotides, from about 150 to about 170 nucleotides, from about 170 to about 200 nucleotides, from about 200 to about 250 nucleotides, or from about 250 to about 300 nucleotides.

[0077] The tracrRNA sequence in the guide RNA generally is based upon the coding sequence of wild type tracrRNA in the bacterial species of interest. In some embodiments, the wild-type tracrRNA sequence (or the crRNA constant repeat region

and the corresponding 5' region of the tracrRNA that forms a duplex structure with the crRNA constant repeat region) can be modified to facilitate secondary structure formation, increased secondary structure stability, facilitate expression in eukaryotic cells, increase editing efficiency, and so forth. For example, one or more nucleotide changes can be introduced into the constant guide RNA sequence (see Example 8, below).

[0078] The guide RNA can be a single molecule (*i.e.*, a single guide RNA or sgRNA), wherein the crRNA sequence is linked to the tracrRNA sequence. Alternatively, the guide RNA can be two separate molecules. A first molecule comprising the crRNA guide sequence at the 5' end and additional sequence at 3' end that is capable of base pairing with the 5' end of a second molecule, wherein the second molecule comprises 5' sequence that is capable of base pairing with the 3' end of the first molecule, as well as additional tracrRNA sequence. In some embodiments, the guide RNA of type V CRISPR/Cpf1 systems can comprise only crRNA.

[0079] In some embodiments, the one or more stem-loop regions of the guide RNA can be modified to comprise one or more aptamer sequences (Konermann et al., Nature, 2015, 517(7536):583-588; Zalatan et al., Cell, 2015, 160(1-2):339-50). Examples of suitable RNA aptamer protein domains include MS2 coat protein (MCP), PP7 bacteriophage coat protein (PCP), Mu bacteriophage Com protein, lambda bacteriophage N22 protein, stem-loop binding protein (SLBP), Fragile X mental retardation syndrome-related protein 1 (FXR1), proteins derived from bacteriophage such as AP205, BZ13, f1, f2, fd, fr, ID2, JP34/GA, JP501, JP34, JP500, KU1, M11, M12, MX1, NL95, PP7, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, Qβ, R17, SP-β, TW18, TW19, and VK, fragments thereof, or derivatives thereof. The length of the additional aptamer sequence can range from about 20 nucleotides to about 200 nucleotides.

[0080] The guide RNA can comprise standard ribonucleotides, modified ribonucleotides (e.g., pseudouridine), ribonucleotide isomers, and/or ribonucleotide analogs. In some embodiments, the guide RNA can further comprise at least one detectable label. The detectable label can be a fluorophore (e.g., FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a

detection tag (e.g., biotin, digoxigenin, and the like), quantum dots, or gold particles. Those skilled in the art are familiar with gRNA design and construction, e.g., gRNA design tools are available on the internet or from commercial sources.

[0081] The guide RNA can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example the guide RNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the guide RNA can be synthesized *in vitro* by operably linking DNA encoding the guide RNA to a promoter control sequence that is recognized by a phage RNA polymerase. Examples of suitable phage promoter sequences include T7, T3, SP6 promoter sequences, or variations thereof. In embodiments in which the guide RNA comprises two separate molecules (*i.e.*, crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

(III) Nucleic Acids

[0082] A further aspect of the present disclosure provides nucleic acids encoding the fusion proteins described above in section (I) and the CRISPR complexes described in section (II). The CRISPR complexes can be encoded by single nucleic acids or multiple nucleic acids. The nucleic acids can be DNA or RNA, linear or circular, single-stranded or double-stranded. The RNA or DNA can be codon optimized for efficient translation into protein in the eukaryotic cell of interest. Codon optimization programs are available as freeware or from commercial sources.

[0083] In some embodiments, the nucleic acid encoding the fusion protein or the protein components of the CRISPR complex or the can be RNA. The RNA can be enzymatically synthesized *in vitro*. For this, DNA encoding the protein of interest can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for *in vitro* RNA synthesis. For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. The DNA encoding the protein can be part of a vector, as detailed below. In such embodiments, the *in vitro*-transcribed RNA can be purified, capped, and/or polyadenylated. In other embodiments, the RNA encoding the fusion protein or protein

component of the complex can be part of a self-replicating RNA (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254). The self-replicating RNA can be derived from a noninfectious, self-replicating Venezuelan equine encephalitis (VEE) virus RNA replicon, which is a positive-sense, single-stranded RNA that is capable of self-replicating for a limited number of cell divisions, and which can be modified to code proteins of interest (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254).

[0084] In other embodiments, the nucleic acid encoding the fusion protein or the CRISPR protein and guide RNA of complex can be DNA. The DNA coding sequence can be operably linked to at least one promoter control sequence for expression in the cell of interest. In certain embodiments, the DNA coding sequence can be operably linked to a promoter sequence for expression of the protein or RNA in bacterial (*e.g.*, *E. coli*) cells or eukaryotic (*e.g.*, yeast, insect, or mammalian) cells. Suitable bacterial promoters include, without limit, T7 promoters, *lac* operon promoters, *trp* promoters, *tac* promoters (which are hybrids of *trp* and *lac* promoters), variations of any of the foregoing, and combinations of any of the foregoing. Non-limiting examples of suitable eukaryotic Pol II promoters include constitutive, regulated, or cell- or tissue-specific promoters. Suitable eukaryotic constitutive promoter control sequences include, but are not limited to, cytomegalovirus immediate early promoter (CMV), simian virus (SV40) promoter, adenovirus major late promoter, Rous sarcoma virus (RSV) promoter, mouse mammary tumor virus (MMTV) promoter, phosphoglycerate kinase (PGK) promoter, elongation factor (ED1)-alpha promoter, ubiquitin promoters, actin promoters, tubulin promoters, immunoglobulin promoters, fragments thereof, or combinations of any of the foregoing. Examples of suitable eukaryotic regulated promoter control sequences include without limit those regulated by heat shock, metals, steroids, antibiotics, or alcohol. Non-limiting examples of tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45 promoter, CD68 promoter, desmin promoter, elastase-1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM-2 promoter, INF-β promoter, Mb promoter, NphsI promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter. The promoter sequence can be wild type or it can be

modified for more efficient or efficacious expression. In some embodiments, the DNA coding sequence also can be linked to a polyadenylation signal (*e.g.*, SV40 polyA signal, bovine growth hormone (BGH) polyA signal, *etc.*) and/or at least one transcriptional termination sequence. The sequence encoding the guide RNA is operably linked to a Pol III promoter control sequence for expression in eukaryotic cells. Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters. In some situations, the fusion protein or components of the complex can be purified from bacterial or eukaryotic cells.

[0085] In various embodiments, nucleic acid encoding the fusion protein or the CRISPR protein and guide RNA of the complex can be present in a vector. Suitable vectors include plasmid vectors, viral vectors, and self-replicating RNA (Yoshioka *et al.*, Cell Stem Cell, 2013, 13:246-254). In some embodiments, the nucleic acid encoding the fusion protein or the components of the complex can be present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. In other embodiments, the nucleic acid encoding the fusion protein or the components of the complex or can be part of a viral vector (*e.g.*, lentiviral vectors, adeno-associated viral vectors, adenoviral vectors, and so forth). The plasmid or viral vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, *etc.*), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like. Additional information about vectors and use thereof can be found in "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001.

(IV) Kits

[0086] A further aspect of the present disclosure provides kits comprising the at least one of the fusion proteins detailed above in section (I), at least one of the CRISPR complexes described above in section (II), and/or at least one of the nucleic acids described above in section (III). The kits can further comprise transfection

reagents, cell growth media, selection media, *in-vitro* transcription reagents, nucleic acid purification reagents, protein purification reagents, buffers, and the like. The kits provided herein generally include instructions for carrying out the methods detailed below. Instructions included in the kits may be affixed to packaging material or may be included as a package insert. While the instructions are typically written or printed materials, they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (*e.g.*, magnetic discs, tapes, cartridges, chips), optical media (*e.g.*, CD ROM), and the like. As used herein, the term "instructions" can include the address of an internet site that provides the instructions.

(V) Cells

[0087] The present disclosure also provides cells comprising the at least one of the fusion proteins detailed above in section (I), at least one of the CRISPR complexes described above in section (II), and/or at least one of the nucleic acids described above in section (III). In general, the cell is a eukaryotic cell. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, or a single cell eukaryotic organism.

(VI) Methods for Increasing Efficiency of Targeted Genome, Transcriptional, or Epigenetic Modification

[0088] Another aspect of the present disclosure encompasses methods for increasing the efficiency of targeted genome modification, targeted transcriptional modification, or targeted epigenetic modification in eukaryotic cells by increasing the accessibility of a programmable DNA modification protein to its target sequence in chromosomal DNA. In some embodiments, the method comprises introducing into the eukaryotic cell of interest at least one of the fusion proteins described above in section (I), at least one of the CRISPR complexes described above in section (II), or nucleic

acid encoding the at least one fusion protein or CRISPR complex as described above in section (III), and optionally, a donor polynucleotide.

[0089] The programmable DNA modification protein of the fusion protein is engineered to recognize and bind to a target sequence in chromosomal DNA, and the one or more nucleosome interacting protein domains of the fusion protein interact with nucleosomes at or near the target sequence to alter or remodel nucleosomal and/or chromatin structure. As a consequence, the DNA modification protein has increased access to the target chromosomal sequence such that efficiency of modification by the DNA modification protein is increased. In specific embodiments, the fusion protein comprises at least one nucleosome interacting protein domain linked to a CRISPR nuclease, such that interactions between the nucleosome interacting protein domain(s) and nucleosomes/chromatin at or near the target sequence increases the efficiency to targeted genomic modifications (*see*, Examples 1-8).

[0090] Thus, the methods disclosed herein can increase the efficiency targeted genome editing (*e.g.*, gene corrections, gene knock-outs, gene knock-ins, and the like), targeted epigenetic modifications, and targeted transcriptional regulation.

(a) Introduction into the Cell

[0091] As mentioned above, the method comprises introducing into the cell at least one fusion protein, at least one CRISPR complex, or nucleic acid(s) encoding said fusion protein or CRISPR complex (and, optionally, a donor polynucleotide). The at least one fusion protein, CRISPR complex, or nucleic acid(s) can be introduced into the cell of interest by a variety of means.

[0092] In some embodiments, the cell can be transfected with the appropriate molecules (*i.e.*, protein, DNA, and/or RNA). Suitable transfection methods include nucleofection (or electroporation), calcium phosphate-mediated transfection, cationic polymer transfection (*e.g.*, DEAE-dextran or polyethylenimine), viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, gene gun

delivery, impalefection, sonoporation, optical transfection, and proprietary agent-enhanced uptake of nucleic acids. Transfection methods are well known in the art (*see*, *e.g.*, "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001). In other embodiments, the molecules can be introduced into the cell by microinjection. For example, the molecules can be injected into the cytoplasm or nuclei of the cells of interest. The amount of each molecule introduced into the cell can vary, but those skilled in the art are familiar with means for determining the appropriate amount.

[0093] The various molecules can be introduced into the cell simultaneously or sequentially. For example, the fusion protein or CRISPR complex (or encoding nucleic acids) and the donor polynucleotide can be introduced at the same time. Alternatively, one can be introduced first and then the other can be introduced later into the cell.

[0094] In general, the cell is maintained under conditions appropriate for cell growth and/or maintenance. Suitable cell culture conditions are well known in the art and are described, for example, in Santiago *et al.*, Proc. Natl. Acad. Sci. USA, 2008, 105:5809-5814; Moehle *et al.* Proc. Natl. Acad. Sci. USA, 2007, 104:3055-3060; Urnov *et al.*, Nature, 2005, 435:646-651; and Lombardo *et al.*, Nat. Biotechnol., 2007, 25:1298-1306. Those of skill in the art appreciate that methods for culturing cells are known in the art and can and will vary depending on the cell type. Routine optimization may be used, in all cases, to determine the best techniques for a particular cell type.

(b) Targeted Genome or Epigenetic Modification

[0095] The one or more nucleosome interacting protein domains of the fusion protein or CRISPR complex interacts with nucleosomes and/or chromosomal DNA at or near the target chromosomal sequence such that nucleosomal and/or chromatin structure is altered/remodeled, thereby increasing accessibility of the programmable DNA modification protein of the fusion protein or the CRISPR protein of the CRISPR complex to the target chromosomal sequence. Increased access to the

target chromosomal sequence results in increased frequency/efficiency of targeted genome, transcriptional, or epigenetic modification.

[0096] In embodiments in which the fusion protein comprises a programmable DNA modification protein having nuclease activity, the fusion protein can cleave one or both strands of the targeted chromosomal sequence. Double-stranded breaks can be repaired by a non-homologous end-joining (NHEJ) repair process. Because NHEJ is error-prone, indels (*i.e.*, deletions or insertions) of at least one base pair, substitutions of at least one base pair, or combinations thereof can occur during the repair of the break. Accordingly, the targeted chromosomal sequence can be modified, mutated, or inactivated. For example, a deletion, insertion, or substitution in the reading frame of a coding sequence can lead to an altered protein product, or no protein product (which is termed a "knock out"). In some iterations, the method can further comprise introducing into the cell a donor polynucleotide (*see* below) comprising a donor sequence that is flanked by sequence having substantial sequence identity to sequences located on either side of the target chromosomal sequence, such that during repair of the double-stranded break by a homology directed repair process (HDR) the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence at the target chromosomal sequence. Integration of an exogenous sequence is termed a "knock in."

[0097] In various iterations, therefore, the efficiency of targeted genome modification can be increased by at least about 0.1-fold, at least about 0.5-fold, at least about 1-fold, at least about 2-fold, at least about 5-fold, at least about 10-fold, or at least about 20-fold, at least about 50-fold, at least about 100-fold, or more than about 100-fold relative to the parental programmable DNA modification protein that is not linked to at least one nucleosome interacting protein domain.

[0098] In embodiments in which the fusion protein comprises a programmable DNA modification protein having non-nuclease activity, the fusion protein can modify DNA or associated proteins at the target chromosomal sequence or modify expression of the target chromosomal sequence. For example, when the programmable DNA modification protein comprises epigenetic modification activity, the

status of histone acetylation, methylation, phosphorylation, adenylation, *etc.* can be modified or the status of DNA methylation, amination, *etc.* can be modified. As an example, in embodiments in which the programmable DNA modification protein comprises cytidine deaminase activity, one or more cytidine residues at the target chromosomal sequence can be converted to uridine residues. Alternatively, when the programmable DNA modification protein comprises transcriptional activation or repressor activity, transcription at target chromosomal sequence can be increased or decreased.

[0099] The resultant epigenetic modification or transcriptional regulation can be increased by at least about 0.1-fold, at least about 0.5-fold, at least about 1-fold, at least about 2-fold, at least about 5-fold, at least about 10-fold, or at least about 20-fold, at least about 50-fold, at least about 100-fold, or more than about 100-fold relative to the parental programmable DNA modification protein that is not linked to at least one nucleosome interacting protein domain.

[0100] The targeted genome, transcriptional, epigenetic modifications detailed above can be performed singly or multiplexed (*i.e.*, two or more chromosomal sequences can be targeted simultaneously).

(c) Optional Donor Polynucleotide

[0101] In embodiments in which the fusion protein comprises a programmable DNA modification protein having nuclease activity, the method can further comprise introducing at least one donor polynucleotide into the cell. The donor polynucleotide can be single-stranded or double-stranded, linear or circular, and/or RNA or DNA. In some embodiments, the donor polynucleotide can be a vector, *e.g.*, a plasmid vector.

[0102] The donor polynucleotide comprises at least one donor sequence. In some aspects, the donor sequence of the donor polynucleotide can be a modified version of an endogenous or native chromosomal sequence. For example, the donor sequence can be essentially identical to a portion of the chromosomal sequence at or near the sequence targeted by the DNA modification protein, but which comprises at

least one nucleotide change. Thus, upon integration or exchange with the native sequence, the sequence at the targeted chromosomal location comprises at least one nucleotide change. For example, the change can be an insertion of one or more nucleotides, a deletion of one or more nucleotides, a substitution of one or more nucleotides, or combinations thereof. As a consequence of the "gene correction" integration of the modified sequence, the cell can produce a modified gene product from the targeted chromosomal sequence.

[0103] In other aspects, the donor sequence of the donor polynucleotide can be an exogenous sequence. As used herein, an "exogenous" sequence refers to a sequence that is not native to the cell, or a sequence whose native location is in a different location in the genome of the cell. For example, the exogenous sequence can comprise protein coding sequence, which can be operably linked to an exogenous promoter control sequence such that, upon integration into the genome, the cell is able to express the protein coded by the integrated sequence. Alternatively, the exogenous sequence can be integrated into the chromosomal sequence such that its expression is regulated by an endogenous promoter control sequence. In other iterations, the exogenous sequence can be a transcriptional control sequence, another expression control sequence, an RNA coding sequence, and so forth. As noted above, integration of an exogenous sequence into a chromosomal sequence is termed a "knock in."

[0104] As can be appreciated by those skilled in the art, the length of the donor sequence can and will vary. For example, the donor sequence can vary in length from several nucleotides to hundreds of nucleotides to hundreds of thousands of nucleotides.

[0105] Typically, the donor sequence in the donor polynucleotide is flanked by an upstream sequence and a downstream sequence, which have substantial sequence identity to sequences located upstream and downstream, respectively, of the sequence targeted by the programmable DNA modification protein. Because of these sequence similarities, the upstream and downstream sequences of the donor polynucleotide permit homologous recombination between the donor polynucleotide and

the targeted chromosomal sequence such that the donor sequence can be integrated into (or exchanged with) the chromosomal sequence.

[0106] The upstream sequence, as used herein, refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence upstream of the sequence targeted by the programmable DNA modification protein. Similarly, the downstream sequence refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence downstream of the sequence targeted by the programmable DNA modification protein. As used herein, the phrase "substantial sequence identity" refers to sequences having at least about 75% sequence identity. Thus, the upstream and downstream sequences in the donor polynucleotide can have about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with sequence upstream or downstream to the target sequence. In an exemplary embodiment, the upstream and downstream sequences in the donor polynucleotide can have about 95% or 100% sequence identity with chromosomal sequences upstream or downstream to the sequence targeted by the programmable DNA modification protein.

[0107] In some embodiments, the upstream sequence shares substantial sequence identity with a chromosomal sequence located immediately upstream of the sequence targeted by the programmable DNA modification protein. In other embodiments, the upstream sequence shares substantial sequence identity with a chromosomal sequence that is located within about one hundred (100) nucleotides upstream from the target sequence. Thus, for example, the upstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides upstream from the target sequence. In some embodiments, the downstream sequence shares substantial sequence identity with a chromosomal sequence located immediately downstream of the sequence targeted by the programmable DNA modification protein. In other embodiments, the downstream sequence shares substantial sequence identity with a chromosomal sequence that is

located within about one hundred (100) nucleotides downstream from the target sequence. Thus, for example, the downstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides downstream from the target sequence.

[0108] Each upstream or downstream sequence can range in length from about 20 nucleotides to about 5000 nucleotides. In some embodiments, upstream and downstream sequences can comprise about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, or 5000 nucleotides. In specific embodiments, upstream and downstream sequences can range in length from about 50 to about 1500 nucleotides.

(d) Cell Types

[0109] A variety of cells are suitable for use in the methods disclosed herein. In general, the cell is a eukaryotic cell. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, or a single cell eukaryotic organism. In some embodiments, the cell can also be a one cell embryo. For example, a non-human mammalian embryo including rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos. In still other embodiments, the cell can be a stem cell such as embryonic stem cells, ES-like stem cells, fetal stem cells, adult stem cells, and the like. In one embodiment, the stem cell is not a human embryonic stem cell. Furthermore, the stem cells may include those made by the techniques disclosed in WO2003/046141, which is incorporated herein in its entirety, or Chung *et al*. (Cell Stem Cell, 2008, 2:113-117). The cell can be *in vitro* or *in vivo* (*i.e.*, within an organism). In exemplary embodiments, the cell is a mammalian cell or mammalian cell line. In particular embodiments, the cell is a human cell or human cell line.

[0110] Non-limiting examples of suitable mammalian cells or cell lines include human embryonic kidney cells (HEK293, HEK293T); human cervical carcinoma

cells (HELA); human lung cells (W138); human liver cells (Hep G2); human U2-OS osteosarcoma cells, human A549 cells, human A-431 cells, and human K562 cells; Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells; mouse myeloma NS0 cells, mouse embryonic fibroblast 3T3 cells (NIH3T3), mouse B lymphoma A20 cells; mouse melanoma B16 cells; mouse myoblast C2C12 cells; mouse myeloma SP2/0 cells; mouse embryonic mesenchymal C3H-10T1/2 cells; mouse carcinoma CT26 cells, mouse prostate DuCuP cells; mouse breast EMT6 cells; mouse hepatoma Hepa1c1c7 cells; mouse myeloma J5582 cells; mouse epithelial MTD-1A cells; mouse myocardial MyEnd cells; mouse renal RenCa cells; mouse pancreatic RIN-5F cells; mouse melanoma X64 cells; mouse lymphoma YAC-1 cells; rat glioblastoma 9L cells; rat B lymphoma RBL cells; rat neuroblastoma B35 cells; rat hepatoma cells (HTC); buffalo rat liver BRL 3A cells; canine kidney cells (MDCK); canine mammary (CMT) cells; rat osteosarcoma D17 cells; rat monocyte/macrophage DH82 cells; monkey kidney SV-40 transformed fibroblast (COS7) cells; monkey kidney CVI-76 cells; African green monkey kidney (VERO-76) cells. An extensive list of mammalian cell lines may be found in the American Type Culture Collection catalog (ATCC, Manassas, VA).

(VII) Methods for Detecting Specific Genomic Loci

[0111] In embodiments in which the fusion protein comprises a programmable DNA modification having non-nuclease activity or the CRISPR complex comprises a catalytically inactive CRISPR protein having non-nuclease activity, said fusion protein or CRISPR complex can be used in methods for detecting or visualizing specific genomic loci in eukaryotic cells. In such embodiments, the fusion protein or CRISPR protein of the complex further comprises at least one detectable label, such as a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles. Alternatively, the guide RNA of the CRISPR complex can further comprise a detectable label for *in situ* detection (*e.g.,* FISH or CISH). The at least one nucleosome interacting protein domain of the fusion protein or CRISPR complex increases access of the programmable DNA modification protein or

CRISPR protein having non-nuclease activity to the target chromosomal sequence, thereby enhancing detection of specific genomic loci or targeted chromosomal sequences.

[0112] The method comprises introducing into the eukaryotic cell the detectably labeled fusion protein, detectably labeled CRISPR complex, or encoding nucleic acid, and detecting the labeled programmable DNA modification protein or labeled CRISPR protein bound to the target chromosomal sequence. The detecting can be via dynamic live cell imaging, fluorescent microscopy, confocal microscopy, immunofluorescence, immunodetection, RNA-protein binding, protein-protein binding, and the like. The detecting step can be performed in live cells or fixed cells.

[0113] In embodiments in which the method comprises detecting chromatin structural dynamics in live cells, the detectably labeled fusion protein or detectably labeled CRISPR complex can be introduced into the cell as proteins or nucleic acids. In embodiments in which the method comprises detecting the targeted chromosomal sequence in fixed cells, the detectably labeled fusion protein or detectably labeled CRISPR complex can be introduced into the cell as proteins (or protein-RNA complexes). Means for fixing and permeabilizing cells are well known in the art. In some embodiments, the fixed cells can be subjected to chemical and/or thermal denaturation processes to convert double-stranded chromosomal DNA into single-stranded DNA. In other embodiments, the fixed cells are not subjected to chemical and/or thermal denaturation processes.

(VIII) Applications

[0114] The compositions and methods disclosed herein can be used in a variety of therapeutic, diagnostic, industrial, and research applications. In some embodiments, the present disclosure can be used to modify any chromosomal sequence of interest in a cell, animal, or plant in order to model and/or study the function of genes, study genetic or epigenetic conditions of interest, or study biochemical pathways involved in various diseases or disorders. For example, transgenic organisms can be created that model diseases or disorders, wherein the

expression of one or more nucleic acid sequences associated with a disease or disorder is altered. The disease model can be used to study the effects of mutations on the organism, study the development and/or progression of the disease, study the effect of a pharmaceutically active compound on the disease, and/or assess the efficacy of a potential gene therapy strategy.

[0115] In other embodiments, the compositions and methods can be used to perform efficient and cost effective functional genomic screens, which can be used to study the function of genes involved in a particular biological process and how any alteration in gene expression can affect the biological process, or to perform saturating or deep scanning mutagenesis of genomic loci in conjunction with a cellular phenotype. Saturating or deep scanning mutagenesis can be used to determine critical minimal features and discrete vulnerabilities of functional elements required for gene expression, drug resistance, and reversal of disease, for example.

[0116] In further embodiments, the compositions and methods disclosed herein can be used for diagnostic tests to establish the presence of a disease or disorder and/or for use in determining treatment options. Examples of suitable diagnostic tests include detection of specific mutations in cancer cells (*e.g.*, specific mutation in EGFR, HER2, and the like), detection of specific mutations associated with particular diseases (*e.g.*, trinucleotide repeats, mutations in β-globin associated with sickle cell disease, specific SNPs, *etc.*), detection of hepatitis, detection of viruses (*e.g.*, Zika), and so forth.

[0117] In additional embodiments, the compositions and methods disclosed herein can be used to correct genetic mutations associated with a particular disease or disorder such as, *e.g.*, correct globin gene mutations associated with sickle cell disease or thalassemia, correct mutations in the adenosine deaminase gene associated with severe combined immune deficiency (SCID), reduce the expression of HTT, the disease-causing gene of Huntington's disease, or correct mutations in the rhodopsin gene for the treatment of retinitis pigmentosa. Such modifications may be made in cells *ex vivo*.

[0118] In still other embodiments, the compositions and methods disclosed herein can be used to generate crop plants with improved traits or increased resistance to environmental stresses. The present disclosure can also be used to generate farm animal with improved traits or production animals. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine or xenotransplantation.

(IX) ENUMERATED EMBODIMENTS

[0119] The following enumerated embodiments are presented to illustrate certain aspects of the present invention, and are not intended to limit its scope.

[0120] 1. A fusion protein comprising at least one nucleosome interacting protein domain linked to a programmable DNA modification protein.

[0121] 2. The fusion protein of embodiment 1, wherein the at least one nucleosome interacting protein domain is a DNA binding domain from a high mobility group (HMG) box (HMGB) protein chosen from HMGB1, HMGB2, or HMGB3; a HMG nucleosome-binding (HMGN) protein chosen from HMGN1, HMGN2, HMGN3a, HMGN3b, HMGN4, or HMGN5; a central globular domain from a histone H1 variant; a DNA binding domain from a chromatin remodeling complex protein chosen from switch/sucrose non-fermentable (SWI/SNF) complex, imitation switch (ISWI) complex, chromodomain-helicase-DNA binding (CHD) complex, nucleosome remodeling and deacetylase (NuRD) complex, INO80 complex, SWR1 complex, RSC complex, or combination thereof.

[0122] 3. The fusion protein of embodiment 2, wherein the at least one nucleosome interacting protein domain is HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular domain, ISWI protein DNA binding domain, CHD1 protein DNA binding domain, or combination thereof.

[0123] 4. The fusion protein of any one of embodiments 1 to 3, wherein the programmable DNA modification protein has nuclease activity.

[0124] 5. The fusion protein of embodiment 4, wherein the programmable DNA modification protein is a clustered regularly interspersed short palindromic repeats (CRISPR) nuclease or nickase, a zinc finger nuclease (ZFN), a transcription activator-like effector nuclease (TALEN), a meganuclease, or a chimeric protein comprising a programmable DNA binding domain linked to a nuclease domain.

[0125] 6. The fusion protein of any one of embodiments 1 to 3, wherein the programmable DNA modification protein has non-nuclease activity.

[0126] 7. The fusion protein of embodiment 6, wherein the programmable DNA modification protein is a chimeric protein comprising a programmable DNA binding domain linked to a non-nuclease domain.

[0127] 8. The fusion protein of embodiment 7, wherein the programmable DNA binding domain is a CRISPR protein modified to lack all nuclease activity, a zinc finger protein, or a transcription activator-like effector.

[0128] 9. The fusion protein of embodiment 7, wherein the non-nuclease domain has acetyltransferase activity, deacetylase activity, methyltransferase activity, demethylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, demyristoylation activity, citrullination activity, helicase activity, amination activity, deamination activity, alkylation activity, dealkylation activity, oxidation activity, transcriptional activation activity, or transcriptional repressor activity.

[0129] 10. The fusion protein of embodiment 9, wherein the non-nuclease domain has cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0130] 11. The fusion protein of any one of embodiments 1 to 10, wherein the at least one nucleosome interacting protein domain is linked to the programmable DNA modification protein directly via a chemical bond, indirectly via a linker, or combination thereof.

[0131] 12. The fusion protein of any one of embodiments 1 to 11, wherein the at least one nucleosome interacting protein domain is linked to the programmable

DNA modification protein at its N-terminus, C-terminus, an internal location, or combination thereof.

[0132] 13. The fusion protein of any one of embodiments 1 to 12, further comprising at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or combination thereof.

[0133] 14. A fusion protein comprising a clustered regularly interspersed short palindromic repeats (CRISPR) protein linked to at least one nucleosome interacting protein domain.

[0134] 15. The fusion protein of embodiment 14, wherein the CRISPR protein is a type II CRISPR/Cas9 nuclease or nickase, or the CRISPR protein is a type V CRISPR/Cpf1 nuclease or nickase.

[0135] 16. The fusion protein of embodiment 14, wherein the CRISPR protein is a type II CRISPR/Cas9 protein modified to lack all nuclease activity and linked to a non-nuclease domain, or a type V CRISPR/Cpf1 protein modified to lack all nuclease activity and linked to a non-nuclease domain.

[0136] 17. The fusion protein of embodiment 16, wherein the non-nuclease domain has cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0137] 18. The fusion protein of any one of embodiments 14 to 17, wherein the at least one nucleosome interacting protein domain is a high mobility group (HMG) box (HMGB) DNA binding domain, a HMG nucleosome-binding (HMGN) protein, a central globular domain from a histone H1 variant, a DNA binding domain from a chromatin remodeling complex protein, or a combination thereof.

[0138] 19. The fusion protein of embodiment 18, wherein at least one nucleosome interacting protein domain is HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular domain, imitation switch (ISWI) protein DNA binding domain, chromodomain-helicase-DNA protein 1 (CHD1) DNA binding domain, or a combination thereof.

[0139] 20. The fusion protein of any one of embodiments 14 to 19, wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein directly via a chemical bond, indirectly via a linker, or a combination thereof.

[0140] 21. The fusion protein of any one of embodiments 14 to 20, wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein at its N-terminus, C-terminus, an internal location, or a combination thereof.

[0141] 22. The fusion protein of any one of embodiments 14 to 21, further comprising at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof.

[0142] 23. The fusion protein of any one of embodiments 14 to 22, wherein the CRISPR protein is *Streptococcus pyogenes* Cas9 (SpCas9), *Streptococcus thermophilus* Cas9 (StCas9), *Streptococcus pasteurianus* (SpaCas9), *Campylobacter jejuni* Cas9 (CjCas9), *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9), *Neisseria cinerea* Cas9 (NcCas9), *Neisseria meningitis* Cas9 (NmCas9), *Francisella novicida* Cpf1 (FnCpf1), *Acidaminococcus sp.* Cpf1 (AsCpf1), or *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1).

[0143] 24. The fusion protein of any one of embodiments14 to 23, wherein the fusion protein has an amino acid sequence having at least about 90% sequence identity with SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

[0144] 25. The fusion protein of any one of embodiments 14 to 24, wherein the fusion protein has an amino acid sequence as set forth in SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

[0145] 26. A complex comprising at least one fusion protein of any one of embodiments 14 to 25 and at least one guide RNA.

[0146] 27. A nucleic acid encoding the fusion protein of any one of embodiments 1 to 25.

[0147] 28. The nucleic acid of embodiment 27, which is codon optimized for translation in a eukaryotic cell.

[0148] 29. The nucleic acid of embodiments 27 or 28, which is part of a viral vector, a plasmid vector, or a self-replicating RNA.

[0149] 30. A method for increasing efficiency of targeted genome or epigenetic modification in a eukaryotic cell, the method comprising introducing into the eukaryotic cell at least one fusion protein as set forth in any one of embodiments 1 to 25, or nucleic acid encoding the at least one fusion protein as set forth in any one of embodiments 27 to 29, wherein the programmable DNA modification protein of the at least one fusion protein is targeted to a target chromosomal sequence and the at least one nucleosome interacting protein domain of the at least one fusion protein alters nucleosomal or chromatin structure such that the at least one fusion protein has increased access to the target chromosomal sequence, thereby increasing efficiency of targeted genome or epigenetic modification.

[0150] 31. The method of embodiment 30, wherein the DNA modification protein of the at least one fusion protein comprises a CRISPR protein and the method further comprises introducing into the eukaryotic cell at least one guide RNA or nucleic acid encoding the at least one guide RNA.

[0151] 32. The method of embodiments 30 or 31, wherein the method further comprises introducing into the eukaryotic cell at least one donor polynucleotide, the donor polynucleotide comprising at least one donor sequence.

[0152] 33. The method of any one of embodiments 30 to 32, wherein the eukaryotic cell is *in vitro*.

[0153] 34. The method of any one of embodiments 30 to 32, wherein the eukaryotic cell is *in vivo*.

[0154] 35. The method of any one of embodiments 30 to 34, wherein the eukaryotic cell is a mammalian cell.

[0155] 36. The method of any one of embodiments 30 to 35, wherein the eukaryotic cell is a human cell.

[0156] 37. A method for increasing efficiency of targeted genome or epigenetic modification in a eukaryotic cell, the method comprising introducing into the eukaryotic cell: (a) at least one fusion protein or nucleic acid encoding at least one fusion protein, each fusion protein comprising a CRISPR protein linked to at least one nucleosome interacting protein domain, wherein the CRISPR protein (i) has nuclease or nickase activity or (ii) is modified to lack all nuclease activity and is linked to a non-nuclease domain; and (b) at least one guide RNA or nucleic acid encoding at least one guide RNA; wherein the CRISPR protein of the at least one fusion protein is targeted to a target chromosomal sequence and the at least one nucleosome interacting protein domain of the at least one fusion protein alters nucleosomal or chromatin structure such that the at least one fusion protein has increased access to the target chromosomal sequence, thereby increasing efficiency of targeted genome or epigenetic modification.

[0157] 38. The method of embodiment 37, wherein the CRISPR protein is a type II CRISPR/Cas9 protein or a type V CRISPR/Cpf1 protein.

[0158] 39. The method of embodiments 37 or 38, wherein the non-nuclease domain has cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

[0159] 40. The method of any one of embodiments 37 to 39, wherein the at least one nucleosome interacting protein domain is a high mobility group (HMG) box (HMGB) DNA binding domain, a HMG nucleosome-binding (HMGN) protein, a central globular domain from a histone H1 variant, a DNA binding domain from a chromatin remodeling complex protein, or a combination thereof.

[0160] 41. The method of any one of embodiments 37 to 40, wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein directly via a chemical bond, indirectly via a linker, or a combination thereof.

[0161] 42. The method of any one of embodiments 37 to 41, wherein the at least one nucleosome interacting protein domain is linked to the N-terminus, C-terminus, and/or an internal location of the CRISPR protein.

[0162] 43. The method of any one of embodiments 37 to 42, wherein the at least one fusion protein further comprises at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof.

[0163] 44. The method of any one of embodiments 37 to 43, wherein nucleic acid encoding the at least one fusion protein is codon optimized for translation in the eukaryotic cell.

[0164] 45. The method of any one of embodiments 37 to 44, wherein nucleic acid encoding the at least one fusion protein is part of a viral vector, a plasmid vector, or a self-replicating RNA.

[0165] 46. The method of any one of embodiments 37 to 45, wherein the method further comprises introducing into the eukaryotic cell at least one donor polynucleotide, the donor polynucleotide comprising at least one donor sequence.

[0166] 47. The method of any one of embodiments 37 to 46, wherein the eukaryotic cell is *in vitro*.

[0167] 48. The method of any one of embodiments 37 to 46, wherein the eukaryotic cell is *in vivo*.

[0168] 49. The method of any one of embodiments 37 to 48, wherein the eukaryotic cell is a mammalian cell.

[0169] 50. The method of any one of embodiments 37 to 48, wherein the eukaryotic cell is a human cell.

DEFINITIONS

[0170] 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.

[0171] When introducing elements of the present disclosure or the preferred embodiments(s) thereof, the articles “a”, “an”, “the” and “said" are intended to mean that there are one or more of the elements. The terms “comprising”, “including” and “having” are intended to be inclusive and mean that there may be additional elements other than the listed elements.

[0172] The term “about” when used in relation to a numerical value, x, for example means x ± 5%.

[0173] As used herein, the terms “complementary” or “complementarity” refer to the association of double-stranded nucleic acids by base pairing through specific hydrogen bonds. The base paring may be standard Watson-Crick base pairing (*e.g.*, 5’-A G T C-3’ pairs with the complementary sequence 3’-T C A G-5’). The base pairing also may be Hoogsteen or reversed Hoogsteen hydrogen bonding. Complementarity is typically measured with respect to a duplex region and thus, excludes overhangs, for example. Complementarity between two strands of the duplex region may be partial and expressed as a percentage (*e.g.*, 70%), if only some (*e.g.*, 70%) of the bases are complementary. The bases that are not complementary are “mismatched.” Complementarity may also be complete (*i.e.*, 100%), if all the bases in the duplex region are complementary.

[0174] As used herein, the term “CRISPR system” refers to a complex comprising a CRISPR protein (*i.e.*, nuclease, nickase, or catalytically dead protein) and a guide RNA.

[0175] The term “endogenous sequence,” as used herein, refers to a chromosomal sequence that is native to the cell.

[0176] As used herein, the term “exogenous” refers to a sequence that is not native to the cell, or a chromosomal sequence whose native location in the genome of the cell is in a different chromosomal location.

[0177] A “gene,” as used herein, refers to a DNA region (including exons and introns) encoding a gene product, as well as all DNA regions which regulate the

production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites, and locus control regions.

[0178] The term "heterologous" refers to an entity that is not endogenous or native to the cell of interest. For example, a heterologous protein refers to a protein that is derived from or was originally derived from an exogenous source, such as an exogenously introduced nucleic acid sequence. In some instances, the heterologous protein is not normally produced by the cell of interest.

[0179] The term "nickase" refers to an enzyme that cleaves one strand of a double-stranded nucleic acid sequence (*i.e.*, nicks a double-stranded sequence). For example, a nuclease with double strand cleavage activity can be modified by mutation and/or deletion to function as a nickase and cleave only one strand of a double-stranded sequence.

[0180] The term "nuclease," as used herein, refers to an enzyme that cleaves both strands of a double-stranded nucleic acid sequence.

[0181] The terms "nucleic acid" and "polynucleotide" refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogs of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analog of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analog of A will base-pair with T.

[0182] The term "nucleotide" refers to deoxyribonucleotides or ribonucleotides. The nucleotides may be standard nucleotides (*i.e.*, adenosine, guanosine, cytidine, thymidine, and uridine), nucleotide isomers, or nucleotide analogs. A nucleotide analog refers to a nucleotide having a modified purine or pyrimidine base

or a modified ribose moiety. A nucleotide analog may be a naturally occurring nucleotide (*e.g.*, inosine, pseudouridine, etc.) or a non-naturally occurring nucleotide. Non-limiting examples of modifications on the sugar or base moieties of a nucleotide include the addition (or removal) of acetyl groups, amino groups, carboxyl groups, carboxymethyl groups, hydroxyl groups, methyl groups, phosphoryl groups, and thiol groups, as well as the substitution of the carbon and nitrogen atoms of the bases with other atoms (*e.g.*, 7-deaza purines). Nucleotide analogs also include dideoxy nucleotides, 2'-O-methyl nucleotides, locked nucleic acids (LNA), peptide nucleic acids (PNA), and morpholinos.

[0183] The terms "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues.

[0184] As used herein, the term "programmable DNA modification protein" refers to a protein that is engineered to bind a specific target sequence in chromosomal DNA and which modifies the DNA or protein(s) associated with DNA at or near the target sequence.

[0185] The term "sequence identity" as used herein, indicates a quantitative measure of the degree of identity between two sequences of substantially equal length. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is

BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Details of these programs can be found on the GenBank website. In general, the substitutions are conservative amino acid substitutions: limited to exchanges within members of group 1: glycine, alanine, valine, leucine, and Isoleucine; group 2: serine, cysteine, threonine, and methionine; group 3: proline; group 4: phenylalanine, tyrosine, and tryptophan; group 5: aspartate, glutamate, asparagine, and glutamine.

[0186] The terms “target sequence,” “target chromosomal sequence,” and “target site” are used interchangeably to refer to the specific sequence in chromosomal DNA to which the programmable DNA modification protein is targeted, and the site at which the programmable DNA modification protein modifies the DNA or protein(s) associated with the DNA.

[0187] Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion.

[0188] In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their percent identity.

[0189] As various changes could be made in the above-described cells and methods without departing from the scope of the invention, it is intended that all matter contained in the above description and in the examples given below, shall be interpreted as illustrative and not in a limiting sense.

EXAMPLES

[0190] The following examples illustrate certain aspects of the disclosure. Table 1 lists the peptide sequences of nucleosome interacting domains and Table 2 presents target chromosomal sequences used in Examples 1-8 presented below.

Table 1. Peptide Sequences of Nucleosome Interacting Domains

Nucleosome Interacting Domain	Sequence (NH_2-COOH)	SEQ ID NO:
Human HMGB1 box A domain (1-84 aa)	MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNF SEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMK TYIPPKGE	40
Human HMGN1 protein	MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKA AAKDKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAE NGETKTEESPASDEAGEKEAKSD	41
Human HMGN2 protein	MPKRKAEGDAKGDKAKVKDEPQRRSARLSAKPAPPKPE PKPKKAPAKKGEKVPKGKKGKADAGKEGNNPAENGDAK TDQAQKAEGAGDAK	42
Human HMGN3a protein	MPKRKSPENTEGKDGSKVTKQEPTRRSARLSAKPAPPK PEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKEGTAP SENGETKAEEAQKTESVDNEGE	43
Human HMGN3b protein	MPKRKSPENTEGKDGSKVTKQEPTRRSARLSAKPAPPK PEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKEGTEN	44
Human histone H1 central globular domain (22-101 aa)	STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVG ENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP	45

Yeast ISWI chromatin-remodeling complex ATPase ISW1 DNA binding domain	LLNPTKRERKENYSIDNYYKDVLNTGRSSTPSHPRMPKP HVFHSHQLQPPQLKVLYEKERMWTAKKTGYVPTMDDVK AAYGDISDEEEKKQKLELLKLSVNNSQPLTEEEEKMKAD WESEGFTNWNKLEFRKFITVSGKYGRNSIQAIARELAPGK TLEEVRAYAKAFWSNIERIEDYEKYLKIIENEEEKIKRVKM QQEALRRKLSEYKNPFFDLKLKHPPSSNNKRTYSEEEDR FILLMLFKYGLDRDDVYELVRDEIRDCPLFELDFYFRSRTP VELARRGNTLLQCLEKEFNAGIVLDDATKDRMKKEDENG KRIREEFADQTANEKENVDGVESKKAKIEDTSNVGTEQLV AEKIPENETTH	46
Yeast chromo domain-containing protein 1 (CHD1) DNA binding domain	DMDSIGESEVRALYKAILKFGNLKEILDELIADGTLPVKSFE KYGETYDEMMEAAKDCVHEEEKNRKEILEKLEKHATAYR AKLKSGEIKAENQPKDNPLTRLSLKKREKKAVLFNFKGVK SLNAESLLSRVEDLKYLKNLINSNYKDDPLKFSLGNNTPK PVQNWSSNWTKEEDEKLLIGVFKYGYGSWTQIRDDPFL GITDKIFLNEVHNPVAKKSASSSDTTPTPSKKGKGITGSSK KVPGAIHLGRRVDYLLSFLRGGLNTKSPS	47

Table 2. Chromosomal Target Sites			
Locus	**Site**	**Sequence (5'-3')**	**SEQ ID NO:**
***Streptococcus pyogenes* Cas9 (SpCas9)**			
POR	#1	AGCCGTGAGTGGAGGGAGCGTGG	48
POR	#2	AGAGGGAGGGGTTGGACTACAGG	49
POR	#3	CATTCGCCAGTACGAGCTTGTGG	50
CAR	#1	CTTTAATGCGCTGACTTGTGAGG	51
EMX1	#1	GTGGCGCATTGCCACGAAGCAGG	52
EMX1	#2	TTCTTCTTCTGCTCGGACTCAGG	53
***Streptococcus pasteurianus* Cas9 (SpaCas9)**			
POR	#1	TGCTGGAAAGGGGAGACCAAGGGTGA	54

POR	#2	AGAGCTACGAGAACCAGAAGCCGTGA	55
***Francisella novicida* Cpf1 (FnCpf1)**			
POR	#1	TTCCCGGCCTCACCCTTGGTCTCCCC	56
POR	#2	TTGGTCTCCCCTTTCCAGCATTCGCC	57
POR	#3	TTCCAGCATTCGCCAGTACGAGCTTG	58
***Campylobacter jejuni* Cas9 (CjCas9)**			
POR	#1	GATCAACATGGGAGACTCCCACGTGGACAC	59
POR	#2	AGATACTTCTTCGGCCACCGCCTCGGACAC	60

Example 1. Improvement of *Streptococcus pyogenes* Cas9 (SpCas9) activity using human HMGB1 box A domain

[0191] A human HMGB1 box A domain (SEQ ID NO:40) was fused with SpCas9 (+NLS) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and the HMGB1 box A domain. Human K562 cells (1×10^6) were transfected with plasmid DNA encoding the fusion protein or wild type SpCas9 protein in molar equivalent amounts (5.2 and 5.0 μg for the fusion protein and the wild type Cas9 protein, respectively) in combination with 3 μg of a sgRNA plasmid for targeting a genomic site (#1) in the human cytochrome p450 oxidoreductase (*POR*) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 nuclease target cleavage activities (% indels) were measured using Cel-I assays. As shown in Table 3, fusion of the human HMGB1 box A domain with the nuclease increased SpCas9 cleavage efficiency at the target site.

Table 3. Cleavage Efficiency		
Nuclease	**Target Site**	**Indel (%)**
Wild type SpCas9	*POR*/site #1	8.5
SpCas9-HMGB1 box A fusion	*POR*/site #1	21.3

Example 2. Improvement of *Streptococcus pyogenes* Cas9 (SpCas9) activity using human HMGN1, HMGN2, HMGN3a, and HMGN3b

[0192] Human HMGN1, HMGN2, HMGN3a, and HMGN3b (SEQ ID NOS:41-44, respectively) were each fused with SpCas9 (+NLS) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and each of the HMGN peptides. Human K562 cells (1×10^6) were transfected with plasmid DNA encoding each of the fusion proteins or the wild type SpCas9 protein in molar equivalent amounts (5.2 and 5.0 µg for each of the fusion proteins and the wild type Cas9 protein, respectively) in combination with 3 µg of a sgRNA plasmid for targeting a genomic site (#1) in the human cytochrome p450 oxidoreductase (*POR*) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. The results, as summarized in Table 4, show that fusion of each of the human HMGN peptides with the nuclease increased SpCas9 cleavage efficiency at the target site.

Table 4. Cleavage Efficiency

Nuclease	**Target Site**	**Indel (%)**
Wild type SpCas9	*POR*/site #1	8.5
SpCas9-HMGN1 fusion	*POR*/site #1	18.3
SpCas9-HMGN2 fusion	*POR*/site #1	13.3
SpCas9-HMGN3a fusion	*POR*/site #1	13.5
SpCas9-HMGN3b fusion	*POR*/site #1	14.4

Example 3. Improvement of *Streptococcus pyogenes* Cas9 (SpCas9) activity using human histone H1 central globular domain

[0193] A human histone H1 central globular domain (SEQ ID NO:45) was fused with SpCas9 (+NLS) at the nuclease carboxyl terminus with the linker LEGGGS

(SEQ ID NO:1) between Cas9 and the globular domain. Human K562 cells (1 x 10^6) were transfected with plasmid DNA encoding the fusion protein or the wild type SpCas9 protein in molar equivalent amounts (5.2 and 5.0 µg for the fusion protein and the wild type Cas9 protein, respectively) in combination with 3 µg of a sgRNA plasmid for targeting a genomic site (#1) in the human cytochrome p450 oxidoreductase (POR) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. The results are presented in Table 5. Fusion of the human histone H1 central globular domain with the nuclease increased SpCas9 cleavage efficiency at the target site.

Table 5. Cleavage Efficiency

Nuclease	**Target Site**	**Indel (%)**
Wild type SpCas9	*POR*/site #1	8.5
SpCas9-H1 central globular domain fusion	*POR*/site #1	19.4

Example 4. Improvement of *Streptococcus pyogenes* Cas9 (SpCas9) activity using a chromatin remodeling protein DNA binding domain

[0194] SpCas9 (+NLS) was fused with the DNA binding domain of the yeast ISWI chromatin-remodeling complex ATPase ISW1 (SEQ ID NO:46) at the nuclease amino terminus with the linker TGSG (SEQ ID NO:2) between Cas9 and the DNA binding domain. Independently, the wild type SpCas9 was fused with the DNA binding domain of the yeast chromo domain-containing protein 1 (CHD1) (SEQ ID NO:47) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and the DNA binding domain. Human K562 cells (1 x 10^6) were transfected with plasmid DNA encoding each of the fusion proteins or the wild type SpCas9 protein in molar equivalent amounts (6.0 and 5.0 µg for each of the fusion proteins and the wild type Cas9 protein, respectively) in combination with 3 µg of a sgRNA plasmid for targeting a genomic site (#1) in the human cytochrome p450

oxidoreductase (*POR*) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. The results, as summarized in Table 6, show that the fusion of each of the DNA binding domains with the nuclease increased SpCas9 cleavage efficiency at the target site.

Table 6. Cleavage Efficiency

Nuclease	**Target Site**	**Indel (%)**
Wild type SpCas9	*POR*/site #1	8.5
ISW1 DNA binding domain-SpCas9 fusion	*POR*/site #1	21.1
SpCas9-CHD1 DNA binding domain fusion	*POR*/site #1	20.8

Example 5. Improvement of *Streptococcus pyogenes* Cas9 (SpCas9) activity using combinations of nucleosome interacting domains

[0195] SpCas9 (+NLS) was fused with the human HMGN1 (SEQ ID N:41) at the nuclease amino terminus with the linker TGSG (SEQ ID NO:2) between Cas9 and HMGN1 and with the human HMGB1 box A domain (SEQ ID NO:40) or the human histone H1 central globular domain (SEQ ID NO: 45) or the yeast chromo domain-containing protein 1 (CHD1) DNA binding domain (SEQ ID NO:47) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and each of the protein domains. Human K562 cells (1×10^6) were transfected with plasmid DNA encoding each of the fusion proteins or the wild type SpCas9 protein in molar equivalent amounts (5.4 µg for the HMGB1 box A and H1 central globular domain fusion proteins, 6.0 µg for the CHD1 DNA binding domain fusion protein, and 5.0 µg for the wild type Cas9 protein) in combination with 3 µg of a sgRNA plasmid for targeting a genomic site (#1, #2, #3) in the human cytochrome p450 oxidoreductase (POR) locus, or a genomic site (#1) the human nuclear receptor subfamily 1 group I member 3 (CAR) locus, or a genomic site (#1, #2) the human empty spiracles homeobox 1 (EMX1) locus.

Transfection was carried out using nucleofection on an Amaxi nucleofector. Five days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and each targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. The results, as summarized in the Table 7, show that the combinatory fusion of these protein domains with the nuclease increased SpCas9 cleavage efficiency at the target sites.

Table 7. Cleavage Efficiency

Nuclease	**Target Site**	**Indel (%)**
Wild type SpCas9	*POR*/site #1	3.4
	POR/site #2	1.3
	POR/site #3	22.2
	CAR/site #1	2.1
	EMX1/site #1	2.2
	EMX1/site #2	1.1
HMGN1-SpCas9-HMGB1 box A fusion	*POR*/site #1	28.2
	POR/site #2	8.3
	POR/site #3	42.7
	CAR/site #1	14.3
	EMX1/site #1	29.0
	EMX1/site #2	12.1
HMGN1-SpCas9-H1 central globular domain fusion	*POR*/site #1	24.3
	POR/site #2	6.5
	POR/site #3	44.2
	CAR/site #1	23.9
	EMX1/site #1	26.9
	EMX1/site #2	21.0
HMGN1-SpCas9- CHD1 DNA binding domain fusion	*POR*/site #1	21.5
	POR/site #2	3.6

	POR/site #3	39.8
	CAR/site #1	9.0
	EMX1/site #1	23.5
	EMX1/site #2	20.2

Example 6. Improvement of *Streptococcus pasteurianus* Cas9 (SpaCas9) activity using combinations of nucleosome interacting domains

[0196] *Streptococcus pasteurianus* Cas9 (SpaCas9) (+NLS) was fused with the human HMGN1 (SEQ ID NO:41) at the nuclease amino terminus with the linker TGSG (SEQ ID NO:2) between Cas9 and HMGN1 and with the human HMGB1 box A domain (SEQ ID NO:41) or the human histone H1 central globular domain (SEQ ID NO:45) or the yeast chromo domain-containing protein 1 (CHD1) DNA binding domain (SEQ ID NO:47) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and each of the protein domains. Human K562 cells (1×10^6) were transfected with plasmid DNA encoding each of the fusion proteins or the wild type SpaCas9 protein in molar equivalent amounts (5.4 and 5.0 µg for each of the fusion proteins and the wild type Cas9 protein, respectively) in combination with 3 µg of a sgRNA plasmid for targeting a genomic site (#1, #2) in the human cytochrome p450 oxidoreductase (*POR*) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. As summarized in Table 8, the combinatory fusion of these protein domains with the nuclease increased SpaCas9 cleavage efficiency at the target sites.

Table 8. Cleavage Efficiency		
Nuclease	**Target Site**	**Indel (%)**
Wild type SpaCas9	*POR*/site #1	16.6
	POR/site #2	12.9

HMGN1-SpaCas9-HMGB1 box A fusion	*POR*/site #1	20.6
	POR/site #2	35.8
HMGN1-SpaCas9- H1 central globular domain fusion	*POR*/site #1	28.6
	POR/site #2	31.7
HMGN1-SpaCas9-CHD1 DNA binding domain fusion	*POR*/site #1	19.4
	POR/site #2	18.5

Example 7. Improvement of *Francisella novicida* Cpf1 (FnCpf1) activity using combinations of nucleosome interacting domains

[0197] *Francisella novicida* Cpf1 (FnCpf1) (+NLS) was fused with the human HMGN1 (SEQ ID NO:41) at the nuclease amino terminus with the linker TGSG (SEQ ID NO:2) between Cpf1 and HMGN1 and with the human HMGB1 box A domain (SEQ ID NO:40) or the human histone H1 central globular domain (SEQ ID NO:45) or the yeast chromo domain-containing protein 1 (CHD1) DNA binding domain (SEQ ID NO:47) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cpf1 and each of the protein domains. Human K562 cells (1×10^6) were transfected with plasmid DNA encoding each of the fusion proteins or the wild type FnCpf1 protein in molar equivalent amounts (5.4 and 5.0 μg for each of the fusion proteins and the wild type Cas9 protein, respectively) in combination with 3 μg of a sgRNA plasmid for targeting a genomic site (#1, #2, #3) in the human cytochrome p450 oxidoreductase (*POR*) locus. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with a DNA extraction solution (QuickExtract™) and the targeted genomic region was PCR amplified. Cas9 target cleavage activities (% indels) were measured using Cel-I assays. The results, as summarized in Table 9, show that the combinatory fusion of these protein domains with the nuclease increased FnCpf1 cleavage efficiency on the target sites.

Table 9. Cleavage Efficiency		
Nuclease	**Target Site**	**Indel (%)**
Wild typeFnCpf1	*POR*/site #1	2.3
	POR/site #2	5.3
	POR/site #3	3.0
HMGN1-FnCpf1-HMGB1 box A fusion	*POR*/site #1	8.2
	POR/site #2	12.8
	POR/site #3	13.2
HMGN1-FnCpf1-H1 central globular domain fusion	*POR*/site #1	8.7
	POR/site #2	12.9
	POR/site #3	13.2
HMGN1-FnCpf1-CHD1 DNA binding domain fusion	*POR*/site #1	7.7
	POR/site #2	7.5
	POR/site #3	9.4

Example 8. Improvement of *Campylobacter jejuni* Cas9 (CjCas9) gene editing efficiency

[0198] *Campylobacter jejuni* Cas9 (CjCas9) (+NLS) was fused with the human HMGN1 (SEQ ID NO:41) at the nuclease amino terminus with the linker TGSG (SEQ ID NO:2) between Cas9 and HMGN1 and with the human HMGB1 box A domain (SEQ ID NO:40) or the human histone H1 central globular domain (SEQ ID NO:45) at the nuclease carboxyl terminus with the linker LEGGGS (SEQ ID NO:1) between Cas9 and each of the protein domains. The wild type CjCas9 gRNA was modified by introducing a U to C mutation into the crRNA constant repeat region and a corresponding A to G mutation into the 5' region of the tracrRNA sequence. The modified sgRNA sequence is: 5'-NNNNNNNNNNNNNNNNNNNNNGUUCUAGUCCCUGAAAAGGGACUAGAAUAAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUU-3', where the mutated nucleotides in the crRNA and tracrRNA moieties are underlined. Guide

sequences targeting two different sites (#1, #2) in the human cytochrome p450 oxidoreductase gene (*POR*) were cloned into the wild type and the modified CjCas9 sgRNA scaffold, respectively. The expression of the sgRNAs was under the control of a U6 promoter. Human K562 cells (1x 10^6) were transfected with 4 µg of CjCas9 plasmid DNA and 3 µg of a sgRNA plasmid DNA. Transfection was carried out using nucleofection on an Amaxi nucleofector. Three days after transfection, cells were lysed with QuickExtract and the targeted genomic regions were PCR amplified. CjCas9 target DNA cleavage activities (% indels) were measured using Cel-I assays. The results are presented in FIG. 1 and show that the fusion proteins had increased cleavage efficiency on the target sites, and that modified CjCas9 sgRNA scaffold effectively increased CjCas9 cleavage efficiency on target sites.

[0199] Table 10 presents the amino acid sequences of the specific fusion proteins. The nucleosome interacting protein domains are shown in bold, the linkers are shown in italics, and the NLS is underlined.

Table 10. CRISPR Fusion Proteins

SpCas9- HMGB1 box A fusion (SEQ ID NO:61)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA

GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***MGKGDPKKPRGKMSSYAFFVQTCREEH**
KKKHPDASVNFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

SpCas9-HMGN1 fusion (SEQ ID NO:62)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHIVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***MPKRKVSSAEGAAKEEPKRRSARLSAK**
PPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAENGETKTEESPASDE
AGEKEAKSD

SpCas9-HMGN2 fusion (SEQ ID NO:63)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL

SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***MPKRKAEGDAKGDKAKVKDEPQRRSAR**
LSAKPAPPKPEPKPKKAPAKKGEKVPKGKKGKADAGKEGNNPAENGDAKTDQAQKAEGAGDAK

SpCas9-HMGN3a fusion (SEQ ID NO:64)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***MPKRKSPENTEGKDGSKVTKQEPTRRS**
ARLSAKPAPPKPEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKEGTAPSENGETKAEEAQKT
ESVDNEGE

SpCas9-HMGN3b fusion (SEQ ID NO:65)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI

PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***MPKRKSPENTEGKDGSKVTKQEPTRRS**
ARLSAKPAPPKPEPKPRKTSAKKEPGAKISRGAKGKKEEKQEAGKEGTEN

SpCas9-Histone H1 globular fusion (SEQ ID NO:66)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQ**
SIQKYIKSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

ISWI-SpCas9 fusion (SEQ ID NO:67)

LLNPTKRERKENYSIDNYYKDVLNTGRSSTPSHPRMPKPHVFHSHQLQPPQLKVLYEKERMWTA
KKTGYVPTMDDVKAAYGDISDEEEKKQKLELLKLSVNNSQPLTEEEEKMKADWESEGFTNWNKL

EFRKFITVSGKYGRNSIQAIARELAPGKTLEEVRAYAKAFWSNIERIEDYEKYLKIIENEEEKI
KRVKMQQEALRRKLSEYKNPFFDLKLKHPPSSNNKRTYSEEEDRFILLMLFKYGLDRDDVYELV
RDEIRDCPLFELDFYFRSRTPVELARRGNTLLQCLEKEFNAGIVLDDATKDRMKKEDENGKRIR
EEFADQTANEKENVDGVESKKAKIEDTSNVGTEQLVAEKIPENETTH*TGSG*MDKKYSIGLDIGT
NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD
TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFF
YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD<u>PKKKRKV</u>

SpCas9-CHD1 fusion (SEQ ID NO:68)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM

PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGDPKKKRKV*LEGGGGS***DMDSIGESEVRALYKAILKFGNLKEIL**
DELIADGTLPVKSFEKYGETYDEMMEAAKDCVHEEEKNRKEILEKLEKHATAYRAKLKSGEIKA
ENQPKDNPLTRLSLKKREKKAVLFNFKGVKSLNAESLLSRVEDLKYLKNLINSNYKDDPLKFSL
GNNTPKPVQNWSSNWTKEEDEKLLIGVFKYGYGSWTQIRDDPFLGITDKIFLNEVHNPVAKKSA
SSSDTTPTPSKKGKGITGSSKKVPGAIHLGRRVDYLLSFLRGGLNTKSPS

HMGN1-SpCas9-HMGB1 box A fusion (SEQ ID NO:69)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MDKKYSIGLDIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSN
EMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR
KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDP
KKKRKV*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKTM**
SAKEKGKFEDMAKADKARYEREMKTYIPPKGE

HMGN1-SpCas9-Histone H1 globular fusion (SEQ ID NO:70)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MDKKYSIGLDIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSN
EMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP

EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR
KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDP
KKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKLSI**
KRLVTTGVLKQTKGVGASGSFRLAKSDEP

HMGN1-SpCas9-CDH1 fusion (SEQ ID NO:71)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MDKKYSIGLDIGTNSVGWAVITDE
YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSN
EMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR
KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ
AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDP
KKKRKV*LEGGGGS***DMDSIGESEVRALYKAILKFGNLKEILDELIADGTLPVKSFEKYGETYDEM**
MEAAKDCVHEEEKNRKEILEKLEKHATAYRAKLKSGEIKAENQPKDNPLTRLSLKKREKKAVLF

NFKGVKSLNAESLLSRVEDLKYLKNLINSNYKDDPLKFSLGNNTPKPVQNWSSNWTKEEDEKLL
IGVFKYGYGSWTQIRDDPFLGITDKIFLNEVHNPVAKKSASSSDTTPTPSKKGKGITGSSKKVP
GAIHLGRRVDYLLSFLRGGLNTKSPS

HMGN1-SpaCas9-HMGB1 box A fusion (SEQ ID NO:72)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MTNGKILGLDIGIASVGVGIIEAK
TGKVVHANSRLFSAANAENNAERRGFRGSRRLNRRKKHRVKRVRDLFEKYGIVTDFRNLNLNPY
ELRVKGLTEQLKNEELFAALRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQI
QLERLEKYGQLRGNFTVYDENGEAHRLINVFSTSDYEKEARKILETQADYNKKITAEFIDDYVE
ILTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCNFYPDEYRASKASYTAQEYNFL
NDLNNLKVSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGKP
DLHTFEPYRKLKFNLESINIDDLSREVIDKLADILTLNTEREGIEDAIKRNLPNQFTEEQISEI
IKVRKSQSTAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKKSSKNTKTIDEKE
VTDEIYNPVVAKSVRQTIKIINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDD
ALKRAAYLYNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLYSGKPISIQELVHNSNNFEIDHI
LPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSFREMKDYVLKQKGLGKKKRDYL
LTTENIDKIEVKKKFIERNLVDTRYASRVVLNSLQSALRELGKDTKVSVVRGQFTSQLRRKWKI
DKSRETYHHHAVDALIIAASSQLKLWEKQDNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQ
PPYQGFVNTISSKGFEDEILFSYQVDSKYNRKVSDATIYSTRKAKIGKDKKEETYVLGKIKDIY
SQNGFDTFIKKYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRRE
NGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVYFNPETLKYEL
MGLKYSDLSFEKGTGNYHISQEKYDAIKEKEGIGKKSEFKFTLYRNDLILIKDIASGEQEIYRF
LSRTMPNVNHYVELKPYDKEKFDNVQELVEALGEADKVGRCIKGLNKPNISIYKVRTDVLGNKY
FVKKKGDKPKLDFKNNKK*PKKKRKV**LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPD**
ASVNFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

HMGN1-SpaCas9-Histone H1 globular fusion (SEQ ID NO:73)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MTNGKILGLDIGIASVGVGIIEAK
TGKVVHANSRLFSAANAENNAERRGFRGSRRLNRRKKHRVKRVRDLFEKYGIVTDFRNLNLNPY
ELRVKGLTEQLKNEELFAALRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQI
QLERLEKYGQLRGNFTVYDENGEAHRLINVFSTSDYEKEARKILETQADYNKKITAEFIDDYVE
ILTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCNFYPDEYRASKASYTAQEYNFL
NDLNNLKVSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGKP
DLHTFEPYRKLKFNLESINIDDLSREVIDKLADILTLNTEREGIEDAIKRNLPNQFTEEQISEI
IKVRKSQSTAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKKSSKNTKTIDEKE
VTDEIYNPVVAKSVRQTIKIINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDD
ALKRAAYLYNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLYSGKPISIQELVHNSNNFEIDHI
LPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSFREMKDYVLKQKGLGKKKRDYL
LTTENIDKIEVKKKFIERNLVDTRYASRVVLNSLQSALRELGKDTKVSVVRGQFTSQLRRKWKI
DKSRETYHHHAVDALIIAASSQLKLWEKQDNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQ
PPYQGFVNTISSKGFEDEILFSYQVDSKYNRKVSDATIYSTRKAKIGKDKKEETYVLGKIKDIY
SQNGFDTFIKKYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRRE
NGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVYFNPETLKYEL

MGLKYSDLSFEKGTGNYHISQEKYDAIKEKEGIGKKSEFKFTLYRNDLILIKDIASGEQEIYRF
LSRTMPNVNHYVELKPYDKEKFDNVQELVEALGEADKVGRCIKGLNKPNISIYKVRTDVLGNKY
FVKKKGDKPKLDFKNNKKPKKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKY**
IKSHYKVGENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

HMGN1-SpaCas9-CHD1 fusion (SEQ ID NO:74)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MTNGKILGLDIGIASVGVGIIEAK
TGKVVHANSRLFSAANAENNAERRGFRGSRRLNRRKKHRVKRVRDLFEKYGIVTDFRNLNLNPY
ELRVKGLTEQLKNEELFAALRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKNKTPGQI
QLERLEKYGQLRGNFTVYDENGEAHRLINVFSTSDYEKEARKILETQADYNKKITAEFIDDYVE
ILTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCNFYPDEYRASKASYTAQEYNFL
NDLNNLKVSTETGKLSTEQKESLVEFAKNTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGKP
DLHTFEPYRKLKFNLESINIDDLSREVIDKLADILTLNTEREGIEDAIKRNLPNQFTEEQISEI
IKVRKSQSTAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKKSSKNTKTIDEKE
VTDEIYNPVVAKSVRQTIKIINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDD
ALKRAAYLYNSSDKLPDEVFHGNKQLETKIRLWYQQGERCLYSGKPISIQELVHNSNNFEIDHI
LPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSFREMKDYVLKQKGLGKKKRDYL
LTTENIDKIEVKKKFIERNLVDTRYASRVVLNSLQSALRELGKDTKVSVVRGQFTSQLRRKWKI
DKSRETYHHHAVDALIIAASSQLKLWEKQDNPMFVDYGKNQVVDKQTGEILSVSDDEYKELVFQ
PPYQGFVNTISSKGFEDEILFSYQVDSKYNRKVSDATIYSTRKAKIGKDKKEETYVLGKIKDIY
SQNGFDTFIKKYNKDKTQFLMYQKDSLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRRE
NGLICKYSKKGKGTPIKSLKYYDKKLGNCIDITPEESRNKVILQSINPWRADVYFNPETLKYEL
MGLKYSDLSFEKGTGNYHISQEKYDAIKEKEGIGKKSEFKFTLYRNDLILIKDIASGEQEIYRF
LSRTMPNVNHYVELKPYDKEKFDNVQELVEALGEADKVGRCIKGLNKPNISIYKVRTDVLGNKY
FVKKKGDKPKLDFKNNKKPKKKRKV*LEGGGGS***DMDSIGESEVRALYKAILKFGNLKEILDELIA**
DGTLPVKSFEKYGETYDEMMEAAKDCVHEEEKNRKEILEKLEKHATAYRAKLKSGEIKAENQPK
DNPLTRLSLKKREKKAVLFNFKGVKSLNAESLLSRVEDLKYLKNLINSNYKDDPLKFSLGNNTP
KPVQNWSSNWTKEEDEKLLIGVFKYGYGSWTQIRDDPFLGITDKIFLNEVHNPVAKKSASSSDT
TPTPSKKGKGITGSSKKVPGAIHLGRRVDYLLSFLRGGLNTKSPS

HMGN1-FnCpf1-HNGB1 fusion (SEQ ID NO:75)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MSIYQEFVNKYSLSKTLRFELIPQ
GKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKK
SDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIEL
FKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKA
KYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFD
DYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPG
ANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI

SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKD
FSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDN
PKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGE
RHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKT
GGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKF
DKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNNPKKK
RKV*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKTMSAK
EKGKFEDMAKADKARYEREMKTYIPPKGE**

HMGN1-FnCpf1-Histone H1 globular fusion (SEQ ID NO:76)

**MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD***TGSG***MSIYQEFVNKYSLSKTLRFELIPQ**
GKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKK
SDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIEL
FKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKA
KYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFD
DYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPG
ANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKD
FSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDN
PKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGE
RHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKT
GGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKF
DKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNNPKKK
RKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKLSIKRL
VTTGVLKQTKGVGASGSFRLAKSDEP**

HMGN1-FnCpf1-CHD1 fusion (SEQ ID NO:77)

**MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD***TGSG***MSIYQEFVNKYSLSKTLRFELIPQ**
GKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKK
SDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIEL
FKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKA
KYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK

FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFD
DYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQ
CRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNL
LHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPG
ANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI
SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKD
FSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDN
PKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGE
RHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKT
GGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKF
DKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEK
LLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNNPKKK
RKV*LEGGGGS***DMDSIGESEVRALYKAILKFGNLKEILDELIADGTLPVKSFEKYGETYDEMMEA
AKDCVHEEEKNRKEILEKLEKHATAYRAKLKSGEIKAENQPKDNPLTRLSLKKREKKAVLFNFK
GVKSLNAESLLSRVEDLKYLKNLINSNYKDDPLKFSLGNNTPKPVQNWSSNWTKEEDEKLLIGV
FKYGYGSWTQIRDDPFLGITDKIFLNEVHNPVAKKSASSSDTTPTPSKKGKGITGSSKKVPGAI
HLGRRVDYLLSFLRGGLNTKSPS**

HMGN1-CjCas9-HMGB1 box A fusion (SEQ ID NO:78)

**MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE
VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD***TGSG***MARILAFDIGISSIGWAFSENDEL
KDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDE
SLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQN
EEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGF
SFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKN
TEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGE
HNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLM
LEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGK
VHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEF
CAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAK
WQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDE
NTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDF
KKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETF
RKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFAL
KVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLI
VSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK
PKKKRKV*LEGGGGS***GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKT
MSAKEKGKFEDMAKADKARYEREMKTYIPPKGE**

HMGN1-CjCas9-Histone H1 globular fusion (SEQ ID NO:79)

MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAE

VANQETKEDLPAENGETKTEESPASDEAGEKEAKSD*TGSG*MARILAFDIGISSIGWAFSENDEL
KDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDE
SLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQN
EEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGF
SFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKN
TEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGE
HNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLM
LEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGK
VHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEF
CAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAK
WQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDE
NTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDF
KKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETF
RKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFAL
KVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLI
VSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK
PKKKRKV*LEGGGGS***STDHPKYSDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVGENADSQIKLS
IKRLVTTGVLKQTKGVGASGSFRLAKSDEP**

CLAIMS

1. A fusion protein comprising a clustered regularly interspersed short palindromic repeats (CRISPR) protein linked to at least one nucleosome interacting protein domain,
wherein the CRISPR protein is a type II CRISPR/Cas9 nuclease or nickase, or the CRISPR protein is a type V CRISPR/Cpf1 nuclease or nickase, and
wherein the fusion protein further comprises at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof, and
wherein the at least one nucleosome interacting protein domain is a high mobility group (HMG) box (HMGB) DNA binding domain; a HMG nucleosome-binding (HMGN) protein; a central globular domain from histone H1; a DNA binding domain from a chromatin remodeling complex protein chosen from switch/sucrose non-fermentable (SWI/SNF) complex, imitation switch (ISWI) complex, chromodomain-helicase-DNA binding (CHD) complex, nucleosome remodeling and deacetylase (NuRD) complex, INO80 complex, SWR1 complex, or RSC complex; or a combination thereof.

2. The fusion protein of claim 1, wherein the CRISPR protein is a type II CRISPR/Cas9 protein modified to lack all nuclease activity and linked to a non-nuclease domain, or a type V CRISPR/Cpf1 protein modified to lack all nuclease activity and linked to a non-nuclease domain.

3. The fusion protein of claim 2, wherein the non-nuclease domain has cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

4. The fusion protein of any preceding claim, wherein at least one nucleosome interacting protein domain is HMGB1 box A domain, HMGN1 protein, HMGN2 protein, HMGN3a protein, HMGN3b protein, histone H1 central globular

domain, imitation switch (ISWI) protein DNA binding domain, chromodomain-helicase-DNA protein 1 (CHD1) DNA binding domain, or a combination thereof.

5. The fusion protein of any preceding claim, wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein directly via a chemical bond, indirectly via a linker, or a combination thereof, and/or wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein at its N-terminus, C-terminus, an internal location, or a combination thereof.

6. The fusion protein of any preceding claim, comprising at least one nuclear localization signal.

7. The fusion protein of claim 1, wherein the CRISPR protein is *Streptococcus pyogenes* Cas9 (SpCas9), *Streptococcus thermophilus* Cas9 (StCas9), *Streptococcus pasteurianus* (SpaCas9), *Campylobacter jejuni* Cas9 (CjCas9), *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9), *Neisseria cinerea* Cas9 (NcCas9), *Neisseria meningitis* Cas9 (NmCas9), *Francisella novicida* Cpf1 (FnCpf1), *Acidaminococcus sp*. Cpf1 (AsCpf1), or *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1).

8. The fusion protein of claim 1, wherein the fusion protein has an amino acid sequence having at least 90% sequence identity with SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79;
optionally wherein the fusion protein has an amino acid sequence as set forth in SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74,

SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, or SEQ ID NO:79.

9. A nucleic acid encoding the fusion protein of any preceding claim, optionally which is codon optimized for translation in a eukaryotic cell, and/or optionally which is part of a viral vector, a plasmid vector, or a self-replicating RNA.

10. A complex comprising at least one fusion protein of any of claims 1-8 and at least one guide RNA.

11. A kit comprising the fusion protein according to any one of claims 1-8, the nucleic acid according to claim 9 or the complex according to claim 10.

12. A eukaryotic cell comprising the fusion protein according to any one of claims 1-8, the nucleic acid according to claim 9 or the complex according to claim 10.

13. A method for increasing efficiency of targeted genome or epigenetic modification in a eukaryotic cell, the method comprising introducing into the eukaryotic cell:
(a) at least one fusion protein or nucleic acid encoding at least one fusion protein, each fusion protein comprising a CRISPR protein linked to at least one nucleosome interacting protein domain, wherein the CRISPR protein (i) has nuclease or nickase activity or (ii) is modified to lack all nuclease activity and is linked to a non-nuclease domain, wherein the CRISPR protein is a type II CRISPR/Cas9 protein or a type V CRISPR/Cpf1 protein, and wherein the at least one fusion protein further comprises at least one nuclear localization signal, at least one cell-penetrating domain, at least one marker domain, or a combination thereof;
(b) at least one guide RNA or nucleic acid encoding at least one guide RNA;
wherein the CRISPR protein of the at least one fusion protein is targeted to a target chromosomal sequence and the at least one nucleosome interacting

protein domain of the at least one fusion protein alters nucleosomal or chromatin structure such that the at least one fusion protein has increased access to the target chromosomal sequence, thereby increasing efficiency of targeted genome or epigenetic modification,
wherein the at least one nucleosome interacting protein domain is a high mobility group (HMG) box (HMGB) DNA binding domain; a HMG nucleosome-binding (HMGN) protein; a central globular domain from histone H1; a DNA binding domain from a chromatin remodeling complex protein chosen from switch/sucrose non-fermentable (SWI/SNF) complex, imitation switch (ISWI) complex, chromodomain-helicase-DNA binding (CHD) complex, nucleosome remodeling and deacetylase (NuRD) complex, INO80 complex, SWR1 complex, or RSC complex; or a combination thereof, and
wherein the method does not comprise a method for treatment of the human or animal body by surgery or therapy.

14. The method of claim 13, wherein the non-nuclease domain has cytosine deaminase activity, histone acetyltransferase activity, transcriptional activation activity, or transcriptional repressor activity.

15. The method of any of claim 13 or claim 14, wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein directly via a chemical bond, indirectly via a linker, or a combination thereof; and/or wherein the at least one nucleosome interacting protein domain is linked to the CRISPR protein at its N-terminus, C-terminus, an internal location, or a combination thereof.

16. The method of any of claims 13-15, wherein the at least one fusion protein comprises at least one nuclear localization signal.

17. The method of any of claims 13-16, wherein the nucleic acid encoding the at least one fusion protein is codon optimized for translation in the eukaryotic

cell, and/or wherein the nucleic acid encoding the at least one fusion protein is part of a viral vector, a plasmid vector, or a self-replicating RNA.

18. The method of any of claims 13-17, wherein the method further comprises introducing into the eukaryotic cell at least one donor polynucleotide, the donor polynucleotide comprising at least one donor sequence.

19. The method of any of claims 13-18, wherein the eukaryotic cell is *in vitro* or *in vivo*, and/or wherein the eukaryotic cell is a mammalian cell or a human cell.