



(12) **DEMANDE DE BREVET CANADIEN  
CANADIAN PATENT APPLICATION**

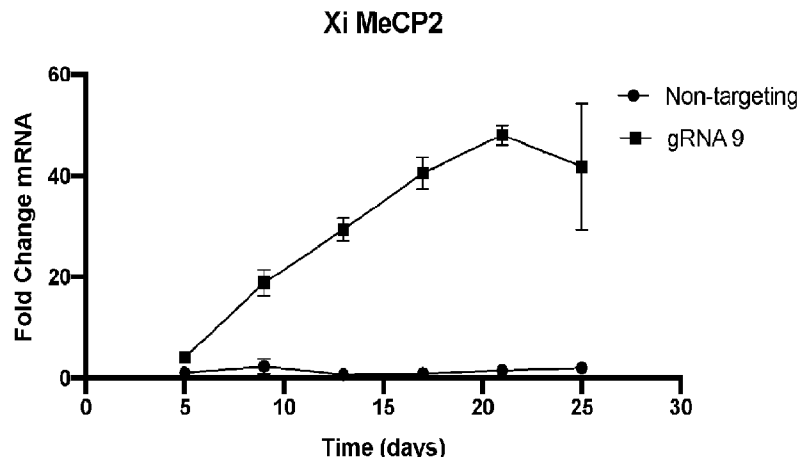
(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/07/29  
 (87) **Date publication PCT/PCT Publication Date:** 2023/02/02  
 (85) **Entrée phase nationale/National Entry:** 2024/01/25  
 (86) **N° demande PCT/PCT Application No.:** US 2022/074355  
 (87) **N° publication PCT/PCT Publication No.:** 2023/010135  
 (30) **Priorités/Priorities:** 2021/07/30 (US63/228,014);  
 2022/05/24 (US63/345,392)

(51) **Cl.Int./Int.Cl. A61K 31/7105** (2006.01),  
**C12N 15/113** (2010.01), **C12N 15/62** (2006.01)  
 (71) **Demandeur/Applicant:**  
 TUNE THERAPEUTICS, INC., US  
 (72) **Inventeurs/Inventors:**  
 BLACK, JOSHUA B., US;  
 SANCHEZ-PEREZ, LUIS, US;  
 GEMBERLING, MATTHEW P., US;  
 KWON, JENNIFER, US;  
 TTOFALI, FANI, US;  
 GERSBACH, CHARLES A., US;  
 PEERS, DILARA, US  
 (74) **Agent:** SMART & BIGGAR LP

(54) **Titre : COMPOSITIONS ET PROCEDES POUR MODULER L'EXPRESSION DE LA PROTEINE 2 DE LIAISON METHYLE-CPG (MECP2)**  
 (54) **Title: COMPOSITIONS AND METHODS FOR MODULATING EXPRESSION OF METHYL-CPG BINDING PROTEIN 2 (MECP2)**

**FIG. 3A**



(57) **Abrégé/Abstract:**

Provided in some aspects are compositions, such as DNA-targeting systems, fusion proteins, guide RNAs (gRNAs), and pluralities and combinations thereof, that bind to or target a methyl-CpG-binding protein 2 (MeCP2) locus. In particular, the present disclosure relates to the modulation of expression of the MeCP2 gene. In some aspects, also provided are polynucleotides, vectors, cells and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or pluralities or combinations thereof, and methods and uses related to the provided compositions, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders that involve the activity, function or expression of MeCP2, such as Rett syndrome.

**Date Submitted:** 2024/01/25

**CA App. No.:** 3227105

**Abstract:**

Provided in some aspects are compositions, such as DNA-targeting systems, fusion proteins, guide RNAs (gRNAs), and pluralities and combinations thereof, that bind to or target a methyl-CpG-binding protein 2 (MeCP2) locus. In particular, the present disclosure relates to the modulation of expression of the MeCP2 gene. In some aspects, also provided are polynucleotides, vectors, cells and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or pluralities or combinations thereof, and methods and uses related to the provided compositions, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders that involve the activity, function or expression of MeCP2, such as Rett syndrome.

## COMPOSITIONS AND METHODS FOR MODULATING EXPRESSION OF METHYL-CPG BINDING PROTEIN 2 (MECP2)

### Cross-Reference to Related Applications

**[0001]** This application claims priority from U.S. provisional application No. 63/228,014, filed July 30, 2021, entitled “COMPOSITIONS AND METHODS FOR MODULATING EXPRESSION OF METHYL-CPG BINDING PROTEIN 2 (MECP2),” and U.S. provisional application No. 63/345,392, filed May 24, 2022, entitled “COMPOSITIONS AND METHODS FOR MODULATING EXPRESSION OF METHYL-CPG BINDING PROTEIN 2 (MECP2),” the contents of which are incorporated by reference in their entireties.

### Incorporation by Reference of Sequence Listing

**[0002]** The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled 224742000240SeqList.xml, created July 29, 2022, which is 379 kilobytes in size. The information in the electronic format of the Sequence Listing is incorporated by reference in its entirety.

### Field

**[0003]** The present disclosure relates in some aspects to compositions, such as DNA-targeting systems, fusion proteins, guide RNAs (gRNAs), and pluralities and combinations thereof, that bind to or target a methyl-CpG-binding protein 2 (MeCP2) locus. In particular, the present disclosure relates to the modulation of expression of the MeCP2 gene. In some aspects, the present disclosure also relates to polynucleotides, vectors, cells and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or pluralities or combinations thereof, and methods and uses related to the provided compositions, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders that involve the activity, function or expression of MeCP2, such as Rett syndrome.

### Background

**[0004]** Several genetic development disorders, including Rett syndrome, are associated with reduced activity, inactivation, mutation and/or dysregulation of expression of the methyl-CpG-binding protein 2 (MeCP2) gene, present on the X chromosome. Rett syndrome affects cells of the nervous system, and can result in a slowing of development resulting in loss of control of the hands, loss of speech, breathing problems, slowed brain and head growth, ambulatory

problems, seizures, and mental retardation. Existing treatment of such genetic disorders are directed towards symptoms and providing support. Treatments that address the fundamental etiology and disease mechanism and needed. Provided are embodiments that meet such needs.

### Summary

**[0005]** Provided herein DNA-targeting systems that bind to or target a methyl-CpG-binding protein 2 (MeCP2) locus. In some aspects, the DNA-targeting systems include fusion proteins. In some aspects, the DNA-targeting systems include guide RNAs (gRNAs). In some aspects, the DNA-targeting systems include fusion proteins and gRNAs. Provided herein are compositions, such as DNA-targeting systems, including fusion proteins, gRNAs, and pluralities and combinations thereof, that bind to or target a MeCP2 locus. Also provided are fusion proteins that bind to or target MeCP2. Also provided are gRNAs that bind to or target MeCP2. In some aspects, the provided DNA-targeting systems, including fusion proteins, gRNAs, bind to, target, and/or modulate the expression of MeCP2. Also provided are compositions, such as polynucleotides, vectors, cells, and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or components thereof. Also provided are methods and uses related to any of the provided compositions and combinations, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders that involve the activity, function or expression of MeCP2, such as Rett syndrome.

**[0006]** Provided herein are DNA-targeting systems comprising a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus. In some of any embodiments, the DNA-targeting system also includes at least one effector domain that increases transcription of the MeCP2 locus.

**[0007]** In some aspects, provided herein is a DNA-targeting system comprising (a) a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and (b) at least one effector domain that increases transcription of the MeCP2 locus.

**[0008]** In some of any of the provided embodiments, binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

**[0009]** In some of any of the provided embodiments, the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof. In some of any



of the provided embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

**[0010]** In some of any of the provided embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA. In some of any of the provided embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein. In some of any of the provided embodiments, the variant Cas protein is a deactivated Cas (dCas) protein.

**[0011]** Also provided herein are DNA-targeting systems comprising a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus and comprises a Cas-guide RNA (gRNA) combination that includes: (a) a variant Cas protein that lacks nuclease activity or that is a deactivated Cas (dCas) protein; and (b) at least one gRNA, comprising a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

**[0012]** In some of any of the provided embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some of any of the provided embodiments, the gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

**[0013]** In some of any of the provided embodiments, the Cas protein or a variant thereof is a Cas9 protein or a variant thereof. In some of any of the provided embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein. In some of any of the provided embodiments, the variant Cas protein is a deactivated Cas (dCas) protein.

**[0014]** In some of any of the provided embodiments, the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof.

**[0015]** In some of any of the provided embodiments, provided herein is a DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising: (a) a *Streptococcus pyogenes* dCas9 (dSpCas9) protein; (b) at least one effector domain that increases transcription of a methyl-CpG-binding protein 2 (MeCP2) locus; and (c) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a MeCP2 locus or is complementary to the target site.

**[0016]** In some of any of the provided embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96. In some of any of the provided embodiments, the variant Cas9 protein comprises the sequence set forth in

SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0017]** In some of any of the provided embodiments, the Cas9 protein or a variant thereof is a *Staphylococcus aureus* Cas9 (SaCas9) protein or a variant thereof. In some of any of the provided embodiments, the variant Cas9 is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99. In some of any of the provided embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0018]** In some of any of the provided embodiments, the variant Cas protein is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein. In some of any of the provided embodiments, the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein. In some of any of the provided embodiments, the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the N-terminal fragment of the variant Cas protein comprises: the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the first polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of

the foregoing. In some of any of the provided embodiments, the C-terminal fragment of the variant Cas protein comprises: the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the second polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0019]** In some of any of the provided embodiments, the target site comprises the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nucleotides (nt), or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158. In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

**[0020]** In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the at least one gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt. In some of any of the provided embodiments, the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30. In some of any of the provided embodiments, the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:69. In some of any of the provided embodiments, the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:69.

**[0021]** In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the at least one gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt. In some of any of the provided embodiments, the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30. In some of any of the provided embodiments, the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:87. In some of any of the provided embodiments, the at least

one gRNA is the gRNA sequence set forth in SEQ ID NO:87.

**[0022]** In some of any of the provided embodiments, the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length. In some of any of the provided embodiments, the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

**[0023]** In some of any of the provided embodiments, the gRNA comprises modified nucleotides for increased stability.

**[0024]** In some of any of the provided embodiments, the DNA-targeting system also includes at least one effector domain. In some of any of the provided embodiments, the DNA-targeting domain or a component thereof is fused to the at least one effector domain.

**[0025]** In some of any of the provided embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA, and the component thereof fused to the at least one effector domain is the Cas protein or a variant thereof.

**[0026]** In some of any of the provided embodiments, the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces transcription de-repression. In some of any of the provided embodiments, the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces transcription de-repression, DNA demethylation or DNA base oxidation.

**[0027]** Also provided herein are DNA-targeting systems comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination that includes: (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and (b) at least one gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:39.

**[0028]** Also provided herein are DNA-targeting systems comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination that includes: (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at

least one effector domain that induces transcription de-repression; and (b) at least one gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:57.

**[0029]** In some aspects, provided herein is a DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising: (a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

**[0030]** In some aspects, provided herein is a DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising: (a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

**[0031]** In some of any of the provided embodiments, the DNA-targeting system further comprises a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of the dSpCas9 fused to a C-terminal Intein.

**[0032]** In some aspects, provided herein is a DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising (a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an C-terminal intein and at least one effector domain that induces transcription de-repression; and (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

**[0033]** In some aspects, provided herein is a DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising: (a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to a C-terminal intein and at least one effector domain that induces transcription de-repression; and (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

**[0034]** In some of any of the provided embodiments, the DNA-targeting system further comprises a first polypeptide of a split variant Cas9 protein an N-terminal fragment of the dSpCas9 fused to an N-terminal Intein. In some of any of the provided embodiments, when the first polypeptide and the second polypeptide of the split variant Cas9 are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-

terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein. In some of any of the provided embodiments, the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0035]** In some of any of the provided embodiments, the N-terminal fragment of the variant Cas9 comprises: the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0036]** In some of any of the provided embodiments, the first polypeptide of the split variant Cas9 comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0037]** In some of any of the provided embodiments, the C-terminal fragment of the variant Cas9 comprises: the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0038]** In some of any of the provided embodiments, the second polypeptide of the split variant Cas9 comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0039]** In some of any of the provided embodiments, the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof. In some of any of the provided embodiments, the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1

(TET1) or a portion or a variant thereof. In some of any of the provided embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0040]** In some of any of the provided embodiments, the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof. In some of any of the provided embodiments, the DNA-targeting system also includes one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

**[0041]** In some of any of the provided embodiments, the DNA-targeting system comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0042]** In some of any of the provided embodiments, the DNA-targeting domain is a first DNA-targeting domain, and the DNA-targeting system further comprises one or more second DNA-targeting domain.

**[0043]** In some aspects, provided herein is a combination comprising: a first DNA-targeting domain comprising any DNA targeting domain provided herein, and one or more second DNA-targeting domains. In some of any embodiments, the one or more second DNA-targeting domains comprises any DNA targeting domain provided herein.

**[0044]** In some of any of the provided embodiments, the first DNA-targeting domain binds a first target site in a MeCP2 locus; and the second DNA-targeting domain binds a second target site in a MeCP2 locus.

**[0045]** Also provided herein are DNA-targeting systems that binds to one or more target sites in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, the DNA-targeting system comprising: a first DNA-targeting domain that binds a first target site in a MeCP2 locus; and a second DNA-targeting domain that binds a second target site in a MeCP2 locus.

**[0046]** Also provided herein is a combination comprising: a first DNA-targeting domain that binds a first target site in a MeCP2 locus; and a second DNA-targeting domain that binds a second target site in a MeCP2 locus.

**[0047]** In some of any of the provided embodiments, the first target site and the second target site independently are located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

**[0048]** In some of any of the provided embodiments, the first DNA-targeting domain comprises a first Cas-gRNA combination that includes (a) a first Cas protein or a variant thereof and (b) a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site; and the second DNA-targeting domain comprises a second Cas-gRNA combination that includes (a) a second Cas protein or a variant thereof and (b) a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site.

**[0049]** In some of any of the provided embodiments, the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein. In some of any of the provided embodiments, the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a deactivated Cas9 (dCas9) protein.

**[0050]** In some of any of the provided embodiments, the first variant Cas protein and/or the second variant Cas protein is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96; or comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0051]** In some of any of the provided embodiments, the first variant Cas protein and/or the second variant Cas protein is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; or comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0052]** In some of any of the provided embodiments, the first variant Cas protein and/or the second variant Cas protein is a split variant Cas9 protein, wherein the split variant Cas9 protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas9 and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas9 and a C-terminal Intein.

**[0053]** In some of any of the provided embodiments, the first Cas protein and the second Cas protein are the same. In some of any of the provided embodiments, the first Cas protein and the second Cas protein are different.

**[0054]** In some of any of the provided embodiments, the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is fused to at least one effector domain.

**[0055]** In some of any of the provided embodiments, the effector domain induces, catalyzes



or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, heterochromatin formation, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces transcription de-repression. In some of any of the provided embodiments, the effector domain induces transcription de-repression.

**[0056]** In some of any of the provided embodiments, the first DNA-targeting domain and the second DNA-targeting domain are encoded in a first polynucleotide. In some of any of the provided embodiments, the first Cas protein and the second Cas protein are encoded in a first polynucleotide. In some of any of the provided embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence. In some of any of the provided embodiments, the first gRNA and the second gRNA are encoded in a first polynucleotide. In some of any of the provided embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence, and the Cas protein, the first gRNA, and the second gRNA are encoded in a first polynucleotide.

**[0057]** In some of any of the provided embodiments, the first DNA-targeting domain is encoded in a first polynucleotide and the second DNA-targeting domain is encoded in a second polynucleotide. In some of any of the provided embodiments, the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide. In some of any of the provided embodiments, the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide. In some of any of the provided embodiments, the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

**[0058]** Also provided are gRNAs that bind a target site located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

**[0059]** Also provided are gRNAs that bind a target site comprising the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a

complementary sequence of any of the foregoing.

**[0060]** Also provided are guide RNAs (gRNAs) that bind a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

**[0061]** In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

**[0062]** In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt. In some of any of the provided embodiments, the gRNA further comprises the sequence set forth in SEQ ID NO:30. In some of any of the provided embodiments, the gRNA comprises the sequence set forth in SEQ ID NO:69. In some of any of the provided embodiments, the at least one gRNA is set forth in SEQ ID NO:69.

**[0063]** In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt. In some of any of the provided embodiments, the gRNA further comprises the sequence set forth in SEQ ID NO:30. In some of any of the provided embodiments, the gRNA comprises the sequence set forth in SEQ ID NO:87. In some of any of the provided embodiments, the gRNA is set forth in SEQ ID NO:87.

**[0064]** In some of any of the provided embodiments, the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length. In some of any of the provided embodiments, the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

**[0065]** In some of any of the provided embodiments, the gRNA comprises modified nucleotides for increased stability. In some of any of the provided embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof.

**[0066]** In some of any of the provided embodiments, the gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

**[0067]** Also provided herein are combinations, such as combinations of gRNAs, that includes a first gRNA comprising any of the gRNAs described herein, and one or more second

gRNAs that binds to a second target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus. In some of any of the provided embodiments, the second gRNA comprises any of the gRNAs described herein.

**[0068]** Also provided herein are combinations, such as combinations of gRNAs, that include: a first gRNA that binds a first target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158; and a second gRNA that binds a second target site in a regulatory DNA element of a MeCP2 locus, wherein the second target site is located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

**[0069]** In some aspects, provided herein is a fusion protein comprising: (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein: the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and the effector domain increases transcription of the MeCP2 locus.

**[0070]** Also provided are fusion proteins that include (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein: the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. Also provided are fusion proteins that include (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein: the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

**[0071]** In some of any of the provided embodiments, binding of the DNA-targeting domain or a component thereof to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

**[0072]** In some of any of the provided embodiments, the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA

(gRNA) combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof. In some of any of the provided embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

**[0073]** In some of any of the provided embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes a Cas protein or a variant thereof and at least one gRNA, and the component of the DNA-targeting domain is a Cas protein or a variant thereof. In some of any of the provided embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein. In some of any of the provided embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof.

**[0074]** In some of any of the provided embodiments, the gRNA binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

**[0075]** In some aspects, provided herein is a fusion protein comprising (1) a Cas protein or a variant thereof and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

**[0076]** In some aspects, provided herein is a fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

**[0077]** In some aspects, provided herein is a fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus. In some of any of the provided embodiments, the first polypeptide of the split variant Cas protein, and a second polypeptide of the split variant Cas protein comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

**[0078]** In some aspects, provided herein is a fusion protein comprising (1) a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain induces

transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

**[0079]** In some aspects, provided herein is a fusion protein comprising (1) a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

**[0080]** In some of any of the provided embodiments, the second polypeptide of the split variant Cas protein, and a first polypeptide of the split variant Cas protein comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

**[0081]** In some of any of the provided embodiments, the Cas protein or a variant thereof is capable of complexing with at least one gRNA. In some of any embodiments, the gRNA binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

**[0082]** In some of any of the provided embodiments, the DNA-targeting domain or a component thereof targeted to the target site does not introduce a genetic disruption or a DNA break at or near the target site

**[0083]** In some of any of the provided embodiments, the Cas protein or a variant thereof is a Cas9 protein or a variant thereof. In some of any of the provided embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein. In some of any of the provided embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein.

**[0084]** In some of any of the provided embodiments, the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof. In some of any of the provided embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96. In some of any of the provided embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0085]** In some of any of the provided embodiments, the Cas9 protein or a variant thereof is a *Streptococcus pyogenes* Cas9 (SaCas9) protein or a variant thereof. In some of any of the

provided embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99. In some of any of the provided embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0086]** In some of any of the provided embodiments, the variant Cas protein is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein. In some of any of the provided embodiments, when the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein. In some of any of the provided embodiments, the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0087]** In some of any of the provided embodiments, the N-terminal fragment of the variant Cas protein comprises: the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0088]** In some of any of the provided embodiments, the first polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0089]** In some of any of the provided embodiments, the C-terminal fragment of the variant Cas protein comprises: the C-terminal fragment of variant SpCas9 from position 574 to the C-

terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

**[0090]** In some of any of the provided embodiments, the second polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing. In some of any of the provided embodiments, the target site comprises the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

**[0091]** In some of any of the provided embodiments, the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158. In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some of any of the provided embodiments, the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

**[0092]** In some of any of the provided embodiments, the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation. In some of any of the provided embodiments, the effector domain induces transcription de-repression.

**[0093]** In some of any of the provided embodiments, the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof. In some of any of the provided embodiments, the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof. In some of any of the provided embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0094]** In some of any of the provided embodiments, the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the

DNA-targeting domain or a component thereof. In some of any of the provided embodiments, the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the Cas protein or a variant thereof. In some of any of the provided embodiments, the fusion protein also includes one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

**[0095]** In some of any of the provided embodiments, the fusion protein also includes one or more linkers connecting the Cas protein or variant thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

**[0096]** In some of any of the provided embodiments, the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0097]** Also provided are combinations comprising any of the fusion proteins described herein, and at least one gRNA. In some of any of the provided embodiments, the at least one gRNA comprises any of the gRNA described herein.

**[0098]** Also provided are polynucleotides encoding any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing.

**[0099]** Also provided are polynucleotides encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of any of the DNA-targeting systems described herein or any of the combinations described herein.

**[0100]** Also provided are polynucleotides encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of any of the DNA-targeting systems described herein or any of the combinations described herein.

**[0101]** Also provided are polynucleotides that include any of the polynucleotides described herein, and one or more additional polynucleotides encoding an additional portion or an additional component of any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing.

**[0102]** Also provided are pluralities of polynucleotides, that includes a first polynucleotide comprising any of the polynucleotides described herein; and a second polynucleotide comprising any of the polynucleotides described herein.

**[0103]** Also provided are vectors that include any of the polynucleotides described herein,



any of the pluralities of polynucleotides described herein, or a first polynucleotide or a second polynucleotide of any of the pluralities of polynucleotides described herein, or a portion or a component of any of the foregoing.

**[0104]** In some of any of the provided embodiments, the vector is a viral vector. In some of any of the provided embodiments, the viral vector is an AAV vector. In some of any of the provided embodiments, the AAV vector is an AAV vector engineered for central nervous system (CNS) tropism. In some of any of the provided embodiments, the AAV vector exhibits tropism for a cell of the central nervous system (CNS), a heart cell, such as a cardiomyocyte, a skeletal muscle cell, a fibroblast, an induced pluripotent stem cell, or a cell derived from any of the foregoing. In some of any of the provided embodiments, the AAV vector is selected from among AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or AAV-DJ vector. In some of any of the provided embodiments, the AAV vector is an AAV5 vector or an AAV9 vector. In some of any of the provided embodiments, the viral vector is an AAV9 vector.

**[0105]** In some of any of the provided embodiments, the vector is a non-viral vector selected from: a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide

**[0106]** Also provided are pluralities of vectors that include comprising any of the vectors described herein, and one or more additional vectors comprising one or more additional polynucleotides encoding an additional portion or an additional component of any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing.

**[0107]** Also provided are pluralities of vectors, that include: a first vector comprising any of the polynucleotides described herein; and a second vector comprising any of the polynucleotides described herein.

**[0108]** Also provided are cells comprising any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing.

**[0109]** In some of any of the provided embodiments, the cell is a nervous system cell, or an induced pluripotent stem cell.

**[0110]** In some of any of the provided embodiments, the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman

syndrome, or PPM-X syndrome. In some of any of the provided embodiments, the cell is from a subject that has or is suspected of having Rett syndrome.

**[0111]** Also provided are methods for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a cell, that involve: introducing any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, into the cell.

**[0112]** In some of any of the provided embodiments, the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some of any of the provided embodiments, the cell is from a subject that has or is suspected of having Rett syndrome.

**[0113]** Also provided are methods for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a subject, that involve: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to the subject.

**[0114]** In some of any of the provided embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some of any of the provided embodiments, the subject has or is suspected of having Rett syndrome.

**[0115]** Also provided are methods of treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, that involve: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0116]** Also provided are methods of treating Rett syndrome, that involve: administering any

of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

**[0117]** In some of any of the provided embodiments, a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome. In some of any of the provided embodiments, the mutant MeCP2 allele comprises a mutation corresponding to R255X. In some of any of the provided embodiments, a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, for example the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome. In some of any of the provided embodiments, a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome. In some of any of the provided embodiments, a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject. In some of any of the provided embodiments, the cell is a nervous system cell, or an induced pluripotent stem cell.

**[0118]** In some of any of the provided embodiments, the introducing, contacting or administering is carried out in vivo or ex vivo.

**[0119]** In some of any of the provided embodiments, following the introducing, contacting or administering, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject. In some of any of the provided embodiments, the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 7.5-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold. In some of any of the provided embodiments, the expression is increased by less than about 200-fold, 150-fold, or 100-fold. In some of any of the provided embodiments, the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

**[0120]** In some of any of the provided embodiments, the subject is a human.

**[0121]** Also provided are pharmaceutical compositions that include any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component

of any of the foregoing.

**[0122]** Also provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0123]** Also provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in treating Rett syndrome.

**[0124]** Also provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0125]** Also provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in the manufacture of a medicament for treating Rett syndrome.

**[0126]** In some of any of the provided embodiments, the pharmaceutical composition is to be administered to a subject. In some of any of the provided embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some of any of the provided embodiments, the subject has or is suspected of having Rett syndrome.

**[0127]** Also provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0128]** Also provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for treating Rett syndrome.

**[0129]** Also provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0130]** Also provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, in the manufacture of a medicament for treating Rett syndrome.

**[0131]** In some of any of the provided embodiments, the pharmaceutical composition is to be administered to a subject. In some of any of the provided embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some of any of the provided embodiments, the subject has

or is suspected of having Rett syndrome.

**[0132]** In some of any of the provided embodiments, a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome. In some of any of the provided embodiments, a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome. In some of any of the provided embodiments, a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, for example the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome. In some of any of the provided embodiments, a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

**[0133]** In some of any of the provided embodiments, the cell is a nervous system cell, or an induced pluripotent stem cell.

**[0134]** In some of any of the provided embodiments, the administration is carried out in vivo or ex vivo.

**[0135]** In some of any of the provided embodiments, following the administration, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject. In some of any of the provided embodiments, the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold. In some of any of the provided embodiments, the expression is increased by less than about 200-fold, 150-fold, or 100-fold. In some of any of the provided embodiments, the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

**[0136]** In some of any of the provided embodiments, the subject is a human.

#### Brief Description of the Drawings

**[0137] FIGS. 1A-1C** show allele-specific activation of MeCP2 in Rett syndrome patient-derived induced pluripotent stem cells (iPSCs). **FIG. 1A** illustrates that mutant R255X-iPSCs harbor one nonsense mutation allele of MeCP2 (R255X) on the X-chromosome. In this cell line, the wild-type (WT) allele is present on the inactive X chromosome (Xi), and the R255X mutant allele is present on the active X chromosome (Xa). **FIGS. 1B** and **1C** show expression of the WT Xi (**FIG. 1B**) and mutant Xa (**FIG. 1C**) alleles of MeCP2, following transduction of R255X-iPSCs with dSpCas9-TET1 and indicated gRNAs, as assessed by RT-qPCR. Control conditions as follows: Ctrl Tct1 (dSpCas9-TET1 expression vector without gRNA), Ctrl VP64

(dSpCas9-2xVP64 expression vector without gRNA), Pool 1 (combined gRNAs 1-5 with dSpCas9-TET1), Pool 2 (combined gRNAs 6-10 with dSpCas9-TET1), Pool1 + VP64 (gRNAs 1-5 and gRNA 9 tested with dSpCas9-Tet1 and dSpCas9-2xVP64), Pool2 + VP64 (gRNAs 6-10 tested with dSpCas9-Tet1 and dSpCas9-2xVP64).

[0138] **FIG. 2** shows location of 29 tested gRNAs with respect to the MeCP2 gene. gRNAs found to increase expression of the Xi WT MeCP2 allele are indicated as Active gRNA.

[0139] **FIGS. 3A-3B** show allele-specific activation of the Xi WT MeCP2 allele (**FIG. 3A**) and Xa R255X (**FIG. 3B**) in R225X-iPSCs after indicated days post-transduction with dSpCas9-TET1 and indicated gRNA, as assessed by RT-qPCR.

[0140] **FIGS. 4A** and **4B** show expression of MeCP2 in R255X-iPSCs following transduction with dSpCas9-TET1 and gRNA 9, using two vector system (**FIG. 4A**) or one vector system (**FIG. 4B**).

[0141] **FIG. 4C** shows expression of MeCP2 in R255X-iPSCs following transduction of dSpCas9-TET1 with gRNA 9 (left), dSpCas9-TET1 with gRNA 27 (middle) or dSpCas9-TET1 with gRNA 9 and gRNA 27 (right).

[0142] **FIG. 5** shows expression of neuronal protein TUBB3 and MeCP2 protein as assessed by immunofluorescence in neurons derived from R255X-iPSCs that were transduced with dSpCas9-TET1 and gRNA 9.

[0143] **FIG. 6** shows results of bisulfite sequencing to determine methylation levels in the MeCP2 promoter in R255X-iPSCs following transduction with dSpCas9-TET1 and a non-targeting gRNA or the MeCP2 promoter-targeting gRNA 9. Cells transduced with gRNA 9 were sorted into MeCP2- and MeCP2+ populations prior to bisulfite sequencing. Lines represent cells from indicated conditions. Dots represent results from individual CpGs. x-axis represents CpG position relative to transcriptional start site (TSS), to scale. y-axis represents % cytosine methylation. The location of promoter region targeted by gRNA 9 is also indicated.

[0144] **FIG. 7A** shows a schematic illustrating a dSpCas9-TET1 fusion protein and modified dSpCas9-TET1 fusion protein with a modified 80-amino acid linker sequence. **FIG. 7B** shows % of MeCP2 positive cells as assessed by flow cytometry after transduction of the indicated fusion protein with gRNA 9, at 11 or 17 days post-transduction.

[0145] **FIG. 8** shows a schematic illustrating an engineered self-assembling split dCas9-TET1 fusion protein. An N-terminal fragment had a TET1 catalytic domain and an N-terminal fragment of dSpCas9, followed by an N terminal Npu Intein. The C-terminal fragment had a C terminal Npu Intein, followed by a C-terminal fragment of dSpCas9. The N-terminal Npu Intein and C-terminal Npu Intein were engineered to self-excise and ligate the N- and C-terminal

fragments, forming the full-length self-assembled dSpCas9-TET1 fusion protein when expressed in a cell.

**[0146] FIG. 9** shows results of flow cytometry to measure % of MeCP2 positive cells following transduction with gRNA 9 and indicated dSpCas9-TET1 components, including the dSpCas9 C-terminal fragment of the split fusion protein alone (left; negative control), a non-split dSpCas9-TET1 fusion protein (center; positive control), or both the C-terminal and N-terminal fragment of the split dSpCas9-TET1 fusion protein.

**[0147] FIG. 10** shows expression of a transgenic inactive X (Xi) allele of MeCP2 with a luciferase reporter allele in mouse fibroblasts, at Day 15 and Day 29 post-transduction with mouse MeCP2-targeting gRNAs (gRNA m1-m7) or control non-targeting gRNA, and a dCas9-TET1 effector, as assessed by RT-qPCR. The fold change in mRNA expression relative to a non-targeting gRNA control and normalized to a Gapdh loading control gene, are depicted.

#### Detailed Description

**[0148]** Provided herein DNA-targeting systems that bind to or target a methyl-CpG-binding protein 2 (MeCP2) locus. In some aspects, the DNA-targeting systems include fusion proteins. In some aspects, the DNA-targeting systems include guide RNAs (gRNAs). In some aspects, the DNA-targeting systems include fusion proteins and gRNAs. Provided herein are compositions, such as DNA-targeting systems, including fusion proteins, gRNAs, and pluralities and combinations thereof, that bind to or target a MeCP2 locus. Also provided are fusion proteins that bind to or target MeCP2. Also provided are gRNAs that bind to or target MeCP2. In some aspects, the provided DNA-targeting systems, including fusion proteins, gRNAs, bind to, target, and/or modulate the expression of MeCP2. Also provided are polynucleotides, vectors, cells, and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or components thereof.

**[0149]** Also provided are methods and uses related to any of the provided compositions and combinations, for example, in modulating the expression of MeCP2, and/or in the treatment of diseases or disorders associated with reduced activity, mutation and/or dysregulation of expression of MeCP2, such as Rett syndrome. In some aspects, also provided are methods and uses related to any of the provided compositions and combinations, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders associated with the activity, function or expression, for example dysregulation or reduced activity, function or expression of MeCP2, such as Rett syndrome.

**[0150]** In some aspects, the provided embodiments are based on an observation described

herein that the level of a MeCP2 locus expression in cells from patients with Rett syndrome, including in induced pluripotent stem cells (iPSCs) generated from Rett syndrome patient cells, can be increased or restored using an exemplary DNA-targeting system comprising a deactivated Cas9 (dCas9)-transcriptional activator fusion protein and a gRNA targeting a human MeCP2 locus. The embodiments described herein demonstrate consistent and effective increase or restoration of MeCP2 expression, in cells from patients with Rett syndrome supporting the utility of the approaches in treating Rett syndrome or other diseases or disorders that are associated with reduced activity, mutation and/or dysregulation of expression of MeCP2.

**[0151]** Certain genetic development disorders, including Rett syndrome, are associated with reduced activity, mutation and/or dysregulation of expression of the methyl-CpG-binding protein 2 (MeCP2) gene, present on the X chromosome. Rett syndrome affects cells of the nervous system, and can result in a slowing of development resulting in loss of control of the hands, loss of speech, breathing problems, slowed brain and head growth, ambulatory problems, seizures, and mental retardation. Existing treatment of such genetic disorders only are directed towards symptoms and providing support, and there is a need for therapies and treatments that address the fundamental etiology and disease mechanism. Provided are embodiments, including DNA-targeting systems, fusion proteins, guide RNAs (gRNAs), polynucleotides, vectors, cells, kits, and pluralities and combinations thereof, and methods and uses thereof, that meet such needs.

**[0152]** In some aspects, the provided embodiments offer an advantage of targeting regulatory DNA elements of an MeCP2 locus for modulating transcription. In some aspects, the provided embodiments offer an advantage of facilitating controlled de-repression or activation of MeCP2, for example to a level that is therapeutically relevant for subjects having a disease or disorder that involve the activity, function or expression of MeCP2, such as Rett syndrome.

**[0153]** In certain aspects, the provided embodiments offer the ability to fine tune and tightly regulate the level of expression and/or activity of MeCP2 in a cell or a subject. As described further below, the control of the expression and/or activity of MeCP2 at a particular level is critical for the survival and normal function of the subject, as the reduction of expression can result in diseases or disorders such as Rett syndrome. Accordingly, the level of expression and/or activity of MeCP2 must be de-repressed, in some cases controlled to be at or near a particular level. The provided embodiments permit such de-repression or activation of expression of MeCP2 without the need for introducing additional copies of MeCP2 into the cell, which could result in adverse effects.

**[0154]** All publications, including patent documents, scientific articles and databases,



referred to in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication were individually incorporated by reference. If a definition set forth herein is contrary to or otherwise inconsistent with a definition set forth in the patents, applications, published applications and other publications that are herein incorporated by reference, the definition set forth herein prevails over the definition that is incorporated herein by reference.

**[0155]** The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

## **I. COMPOSITIONS AND METHODS FOR MODULATING EXPRESSION OF METHYL-CPG BINDING PROTEIN 2 (MeCP2)**

**[0156]** Provided herein are compositions such a DNA-targeting systems that bind to or target a MeCP2 locus. In some aspects, the provided DNA-targeting systems include fusion proteins and/or guide RNAs (gRNAs). In some aspects, provided are polynucleotides, vectors that encode any of the DNA-targeting systems, fusion proteins and/or components of kits. In some embodiments, provided are cells, kits, systems and pluralities and combinations thereof, that comprise any of the DNA-targeting systems, fusion proteins or gRNAs described herein.

**[0157]** Provided herein are DNA-targeting systems comprising a DNA-targeting domain that binds to a target site in a target site at a MeCP2 locus. In some of any of the embodiments provided herein, binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site. In some aspects, the provided DNA-targeting systems comprise a fusion protein comprising a DNA-targeting domain and an effector domain, and binds to a target site in a MeCP2 locus. In some aspects, the DNA-targeting system also comprises a guide RNA (gRNA). In some aspects, when administered to a subject or delivered or introduced into a cell that exhibits dysregulation or reduced activity, function or expression of MeCP2, the provided DNA-targeting systems can lead to an increase of or a restoration of, the activity, function or expression of MeCP2. Also provided are methods and uses related to any of the provided compositions, for example, in modulating the expression of MeCP2, and/or in the treatment or therapy of diseases or disorders that involve the activity, function or expression of MeCP2, such as Rett syndrome.

**[0158]** In some embodiments, the DNA-targeting systems are targeted to one or more target sites located within a regulatory DNA element of a MeCP2 locus, such as a promoter or an enhancer. In some embodiments, the DNA-targeting systems are targeted to at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 target sites within a regulatory DNA element of a MeCP2 locus. In some

embodiments, the DNA-targeting systems are targeted to one or more target sites located within a promoter of a MeCP2 locus, and one or more target sites located within an enhancer of a MeCP2 locus.

**[0159]** In some embodiments, the DNA-targeting system comprises a DNA-targeting domain comprising a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof. In some aspects, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing. In some embodiments, the DNA-targeting system comprises a DNA-targeting domain comprising a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof, and (b) at least one gRNA. In some embodiments, the at least one gRNA comprises at least 2, 3, 4, 5, 6, 7, 8, 9 or 10 gRNAs. In some embodiments, the gRNAs are targeted to one or more target sites located within a MeCP2 locus, such as a regulatory DNA element of MeCP2.

**[0160]** In some aspects, the provided embodiments involve modulating transcription of an endogenous MeCP2 locus in a cell. In some aspects, the provided embodiments involve depressing or increasing transcription of an endogenous MeCP2 locus, such as the wild-type MeCP2 allele on an inactive X chromosome of in a cell or a subject. In some embodiments, the cell, such as the cell to be treated with the provided embodiments, has a mutation, such as a R255X mutation, in the MeCP2 locus of the active X chromosome. In some embodiments, the cell, such as the cell to be treated with the provided embodiments, is from or in a subject with Rett syndrome. In some embodiments, the cell, such as the cell to be treated with the provided embodiments, exhibits reduced expression of MeCP2 compared to a cell from a subject without Rett syndrome.

**[0161]** In some aspects, in a cell introduced with or contacted with any of the DNA-targeting systems, gRNA, combinations, fusion proteins, polynucleotides, plurality of polynucleotides, vectors, plurality of vectors or components or portions thereof provided herein, the expression of MeCP2 is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold, compared to a cell that has not been introduced or contacted. In some embodiments, the expression is increased by less than about 200-fold, 150-fold, or 100-fold. In some of any of the provided embodiments, the expression MeCP2 is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

**[0162]** In some embodiments, the subject is a human. In some embodiments, the cell is a

heart cell, a skeletal muscle cell, a nervous system cell, or an induced pluripotent stem cell. In some embodiments, the introducing, contacting or administering is carried out in vivo or ex vivo.

#### **A. MeCP2 and Rett Syndrome**

**[0163]** Several genetic development disorders, including Rett syndrome, are associated with reduced activity, inactivation, mutation and/or dysregulation of expression of the methyl-CpG-binding protein 2 (MeCP2) gene, present on the X chromosome. Rett syndrome affects cells of the nervous system, and can result in a slowing of development resulting in loss of control of the hands, loss of speech, breathing problems, slowed brain and head growth, ambulatory problems, seizures, and mental retardation. Existing treatment of such genetic disorders only are directed towards symptoms and providing support, and there is a need for therapies and treatments that address the fundamental etiology and disease mechanism. Provided are embodiments that meet such needs.

**[0164]** Rett syndrome is a developmental disorder of the brain occurring mostly in females characterized by normal early development, followed by a slowing of development resulting in loss of control of the hands, loss of speech, breathing problems, slowed brain and head growth, ambulatory problems, seizures, and mental retardation. Rett syndrome affects approximately 1 in 10,000 live female births. Most cases of Rett syndrome are associated with a mutation in the methyl CpG binding protein 2, or MeCP2 gene, on the X chromosome that causes reduced activity or inactivation of MeCP2.

**[0165]** MeCP2 (exemplary amino acid sequences of human MeCP2 Isoform A: Uniprot P51608-1 (486 aa), SEQ ID NO:177; exemplary amino acid sequences of human MeCP2 Isoform B: Uniprot P51608-1 (498 aa) SEQ ID NO:221) is a transcriptional repressor that binds to methylated DNA and is present in large quantities in mature nerve cells. MeCP2 represses transcription from methylated gene promoters through interaction with histone deacetylase and the corepressor SIN3A. Many of the genes that are known to be regulated by the MeCP2 protein play a role in normal brain function, particularly the maintenance of synapses. Mouse studies have demonstrated MeCP2 mutations cause defects in synaptic function, especially in synaptic plasticity.

**[0166]** In some aspects, activity, expression or function of MeCP2 is associated with Angelman syndrome (AS), also known as happy puppet syndrome. AS is a neurodevelopmental disorder characterized by severe mental retardation, absent speech, ataxia, sociable affect and dysmorphic facial features. AS and Rett syndrome have overlapping clinical features.

**[0167]** In some aspects, activity, expression or function of MeCP2 is associated with mental retardation syndromic X-linked type 13 (MRXS13). Mental retardation is a mental disorder characterized by significantly sub-average general intellectual functioning associated with impairments in adaptive behavior and manifested during the developmental period. MRXS13 patients manifest mental retardation associated with other variable features such as spasticity, episodes of manic depressive psychosis, increased tone and macroorchidism.

**[0168]** In some aspects, activity, expression or function of MeCP2 is associated with Rett syndrome (RTT). RTT is an X-linked dominant disease, it is a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females. Patients appear to develop normally until 6 to 18 months of age, then gradually lose speech and purposeful hand movements and develop microcephaly, seizures, autism, ataxia, intermittent hyperventilation, and stereotypic hand movements. After initial regression, the condition stabilizes and patients usually survive into adulthood.

**[0169]** In some aspects, activity, expression or function of MeCP2 is associated with susceptibility autism X-linked type 3 (AUTSX3). AUTSX3 is a pervasive developmental disorder (PDD), prototypically characterized by impairments in reciprocal social interaction and communication, restricted and stereotyped patterns of interests and activities, and the presence of developmental abnormalities by 3 years of age.

**[0170]** In some aspects, activity, expression or function of MeCP2 is associated with encephalopathy neonatal severe due to MeCP2 mutations (ENS-MeCP2). Although it was first thought that MeCP2 mutations causing Rett syndrome were lethal in males, later reports identified a severe neonatal encephalopathy in surviving male sibs of patients with Rett syndrome. Additional reports have confirmed a severe phenotype in males with Rett syndrome-associated MeCP2 mutations.

**[0171]** In some aspects, activity, expression or function of MeCP2 is associated with mental retardation syndromic X-linked Lubs type (MRXSL). Mental retardation is characterized by significantly below average general intellectual functioning associated with impairments in adaptive behavior and manifested during the developmental period. MRXSL patients manifest mental retardation associated with variable features. They include swallowing dysfunction and gastroesophageal reflux with secondary recurrent respiratory infections, hypotonia, mild myopathy and characteristic facies such as downslanting palpebral fissures, hypertelorism and a short nose with a low nasal bridge. In some aspects, increased dosage of MeCP2 due to gene duplication appears to be responsible for the mental retardation phenotype.

## **B. Modulating Expression of MeCP2**

**[0172]** In some aspects, provided are compositions, methods and related uses, that can be employed to modulate the expression of MeCP2, such as in a cell or a subject. In some aspects, the provided compositions, methods and uses can be employed to de-repress or increase the expression of wild-type MeCP2 allele on an inactive X chromosome of the cell or the subject. In some aspects, the subject has or is suspected of having a disease or disorder associated with reduced activity, inactivation, mutation and/or dysregulation of expression of the methyl-CpG-binding protein 2 (MeCP2) gene, such as Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some aspects, by modulating, such as by de-repressing or increasing the expression of the wild-type MeCP2 allele on an inactive X chromosome, the provided compositions, methods and uses can be employed to treat or ameliorate the disease or disorder associated with reduced activity, inactivation, mutation and/or dysregulation of MeCP2.

**[0173]** In some aspects, the MeCP2 locus on the inactive X (Xi) in somatic cells is typically silenced by virtue of heterochromatin-mediated transcriptional silencing. The Xi exhibits characteristic features of heterochromatin including inhibitory histone modifications, such as histone H3-lysine 27 trimethylation (H3K27me3) and histone H2A ubiquitination (H2Aub), and hypermethylated DNA regions. Reversal of heterochromatin formation and silencing, and de-repression of the transcription from the MeCP2 locus on the Xi, can lead to recovery of expression of the MeCP2 gene and be used for treatment and/or prevention of such diseases or disorders.

**[0174]** In some aspects, by modulating, such as by activating, de-repressing or increasing the expression of MeCP2, the provided compositions, methods and uses can be employed to restore or recover the expression or activity of MeCP2 in a subject or a cell with a disease or disorder associated with reduced activity, mutation and/or dysregulation of MeCP2, such that the expression or activity of MeCP2 is increased at least about 1.2-fold, 1.25-fold, 1.3-fold, 1.4-fold, 1.5-fold, 1.6-fold, 1.7-fold, 1.75-fold, 1.8-fold, 1.9-fold, 2-fold, 2.5-fold, 3-fold, 4-fold, or 5-fold, compared to the expression or activity of MeCP2 in the subject or cell with the disease or disorder in the absence of the provided compositions or uses. In some aspects, the expression or activity is increased by less than about 10-fold, 9-fold, 8-fold, 7-fold or 6-fold. In some aspects, by modulating, such as by activating, de-repressing or increasing the expression of MeCP2, the provided compositions, methods and uses can be employed to restore or recover the expression or activity of MeCP2 in a subject or a cell with a disease or disorder associated with reduced activity, mutation and/or dysregulation of MeCP2, such that the expression or activity of MeCP2

is increased to at least at or about 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 120%, 125%, 150%, 175%, 200%, 225%, 250%, 300%, 400%, or 500%, of the expression or activity of MeCP2 in an individual or a cell without the disease or disorder or in a wild-type cell. Increasing the expression of MeCP2 mRNA and/or protein, can lead to recovery or restoration of expression of the MeCP2 gene and be used for treatment and/or prevention of such diseases or disorders.

## II. DNA-TARGETING SYSTEMS

**[0175]** Provided herein are DNA-targeting systems comprising a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus. Exemplary components and features of the DNA-targeting systems are provided herein. In some aspects, the DNA-targeting system comprises one or more of any of the components described herein, such as one or more DNA-targeting domains, one or more fusion proteins, such as one or more fusion proteins comprising one or more DNA-targeting domains and one or more effector domains, one or more gRNAs, or any component, portion or fragment thereof, or any combination thereof.

**[0176]** In some aspects, the DNA-targeting system comprises a DNA-targeting domain and one or more guide RNAs (gRNAs). In some aspects, the DNA-targeting system comprises a fusion protein and one or more gRNAs. In some aspects, the DNA-targeting system comprises a DNA-targeting domain and a gRNA. In some aspects, the DNA-targeting system comprises a fusion protein. In some aspects, the DNA-targeting system comprises a fusion protein and a gRNA. In some aspects, the DNA-targeting system comprises a DNA-targeting domain.

**[0177]** In some embodiments, binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

**[0178]** In some embodiments, provided are DNA-targeting systems capable of specifically targeting a target site in a MeCP2 gene or DNA regulatory element thereof, and increasing transcription of the MeCP2 gene. In some embodiments, the DNA-targeting systems include a DNA-targeting domain that binds to a target site in the MeCP2 gene or regulatory DNA element thereof. In provided embodiments, the DNA-targeting systems additionally include at least one effector domain that is able to epigenetically modify one or more DNA bases of the MeCP2 gene or regulatory element thereof, in which the epigenetic modification results in an increase in transcription of the MeCP2 gene (e.g. de-represses, re-activates, activates transcription or increases transcription of MeCP2 compared to the absence of the DNA-targeting system). Hence, the terms DNA-targeting system and epigenetic-modifying DNA targeting system may

be used herein interchangeably. In some embodiments, the DNA-targeting system includes a fusion protein comprising (a) a DNA-targeting domain capable of being targeted to the target site; and (b) at least one effector domain capable of increasing transcription of the MeCP2 gene. For instance, the at least one effector domain is a transcription activation domain.

**[0179]** In some embodiments, the DNA-targeting domain comprises or is derived from a CRISPR associated (Cas) protein, zinc finger protein (ZFP), transcription activator-like effectors (TALE), meganuclease, homing endonuclease, I-SceI enzyme, or variants thereof. In some embodiments, the DNA-targeting domain comprises a catalytically inactive (e.g. nuclease-inactive or nuclease-inactivated) variant of any of the foregoing. In some embodiments, the DNA-targeting domain comprises a deactivated Cas9 (dCas9) protein or variant thereof that is a catalytically inactivated so that it is inactive for nuclease activity and is not able to cleave the DNA.

**[0180]** In some embodiments, the DNA-targeting domain comprises or is derived from a Cas protein or variant thereof, such as a nuclease-inactive Cas or dCas (e.g. dCas9, and the DNA-targeting system comprises one or more guide RNAs (gRNAs). In some embodiments, the gRNA comprises a spacer sequence that is capable of targeting and/or hybridizing to the target site. In some embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some aspects, the gRNA directs or recruits the Cas protein or variant thereof to the target site.

**[0181]** In some embodiments, the DNA-targeting system comprises a DNA-targeting domain. In some embodiments the DNA-targeting domain comprises a DNA-binding protein or DNA-binding nucleic acid. In some embodiments, the DNA-targeting domain specifically binds to or hybridizes to a particular site or position in the genome, e.g., a target, target site, or target position. In some aspects, the DNA-targeting domain is coupled to, fused to or complexed with an effector domain, such as any effector domain described herein, for example, in Section II.D.

**[0182]** In some embodiments, the DNA-targeting system comprises various components, such as an RNA-guided nuclease, variant thereof, or fusion protein comprising the RNA-guided nuclease or variant thereof, or a fusion protein comprising a DNA-targeting domain and an effector domain. In some embodiments, the DNA-targeting system comprises a DNA-targeting molecule that comprises a DNA-binding protein such as one or more zinc finger protein (ZFP) or transcription activator-like effectors (TALEs), fused to an effector domain.

**[0183]** In some embodiments, the DNA-targeting system specifically targets at least one target site in a regulatory DNA element of a MeCP2 locus. In some embodiments, the DNA-targeting system comprises a ZFP, TALE or a CRISPR/Cas9 combination that specifically binds

to, recognizes, or hybridizes to the target site(s). In some embodiments, the CRISPR/Cas9 system includes an engineered crRNA/tracrRNA (i.e. "single guide RNA"). In some embodiments, the DNA-targeting system comprises nucleases or variants thereof based on the Argonaute system (e.g., from *T. thermophilus*, known as 'TtAgo' (Swarts et al., (2014) Nature 507(7491): 258-261).

**[0184]** In some embodiments, the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof. In some embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing. In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA. In some embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein.

**[0185]** Also provided herein are DNA-targeting systems comprising a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus and comprises a Cas-guide RNA (gRNA) combination that includes: (a) a variant Cas protein that lacks nuclease activity or that is a deactivated Cas (dCas) protein; and (b) at least one gRNA, each comprising a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

**[0186]** In some embodiments, the DNA-targeting system comprises a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus and comprises a Cas-guide RNA (gRNA) combination that includes: (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and (b) a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

**[0187]** In some embodiments, the DNA-targeting system comprises a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus and comprises a Cas-guide RNA (gRNA) combination that includes: (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and (b) a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

**[0188]** In some embodiments, the DNA-targeting system comprises a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein



2 (MeCP2) locus and comprises a Cas-guide RNA (gRNA) combination that includes: (a) a *Staphylococcus aureus* deactivated Cas9 protein (dSaCas9) protein set forth in SEQ ID NO:98 fused to at least one effector domain that induces transcription de-repression; and (b) a gRNA comprising a gRNA spacer sequence set forth in any one of SEQ ID NOS:231-240.

#### **A. Target Site at the MeCP2 Locus**

**[0189]** In some aspects, provided are compositions, methods and uses, such as DNA-targeting system, DNA-targeting domains, components of the DNA-targeting domains, such as at least one gRNA, fusion proteins, and pluralities and combinations thereof, polynucleotides, vectors, cells and pluralities and combinations thereof, that encode or comprise the DNA-targeting systems, fusion proteins, gRNAs or pluralities or combinations thereof, that can target a particular genomic location related to the MeCP2 locus, such as a regulatory DNA element of the MeCP2 locus.

**[0190]** In some embodiments, the target site is in a cell, such as any suitable cell. In some embodiments, the cell is in or from any suitable organism, such as a human, mouse, dog, horse, rabbit, cattle, pig, hamster, gerbil, mouse, ferret, rat, cat, non-human primate, monkey, etc. In some embodiments, the cell is in or from a human. In some embodiments, the cell is any suitable cell, such as an immune cell (e.g. a T cell, B cell, or antigen-presenting cell), a liver cell (e.g. a hepatocyte), a cell of a nervous system (e.g. a neuron or glial cell), a heart cell (e.g. a cardiomyocyte) or a stem cell (e.g. an embryonic stem cell or induced pluripotent stem cell).

**[0191]** In some embodiments, the target site is located in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus. In some embodiments, the target site is located within the promoter, upstream regulatory element (e.g., enhancer), exon, intron, 5' untranslated region (UTR), 3' UTR, or downstream regulatory element.

**[0192]** In some embodiments, the target site is located within a MeCP2 locus. In some embodiments the target site is located within a regulatory DNA element (e.g. a cis-, trans-, distal, proximal, upstream, or downstream regulatory DNA element) of a MeCP2 locus. In some embodiments, the target site is located within a promoter, enhancer, exon, intron, untranslated region (UTR), 5' UTR or 3' UTR. In some embodiments the target site is located within a sequence and/or sequences of unknown or known function that are suspected of being able to control expression of MeCP2.

**[0193]** In some embodiments one or more target sites, such as one or more target sites located within a regulatory DNA element (e.g. a cis-, trans-, distal, proximal, upstream, or downstream regulatory DNA element) of a MeCP2 locus. In some embodiments, the target site

is located within a promoter, enhancer, exon, intron, untranslated region (UTR), 5' UTR or 3' UTR are targeted.

**[0194]** In some aspects, an exemplary human methyl-CpG binding protein 2 (MeCP2) transcript is set forth in RefSeq NM\_004992 (transcript variant 1); Gencode Transcript: ENST00000303391.11; Gencode Gene: ENSG00000169057.24. Genomic coordinates for an exemplary transcript (including UTRs) for MeCP2 include hg38 chrX:154,021,573-154,097,717 (Size: 76,145; Total Exon Count: 4 Strand: -). Genomic coordinates for the coding region for this transcript variant include hg38 chrX:154,030,367-154,092,209 (Size: 61,843 Coding Exon Count: 3).

**[0195]** In some aspects, an exemplary human methyl-CpG binding protein 2 (MeCP2) transcript is set forth in RefSeq NM\_001369393 (transcript variant 6); Gencode Transcript: ENST00000453960.7; Gencode Gene: ENSG00000169057.24. Genomic coordinates for an exemplary transcript (including UTRs) for MeCP2 include hg38 chrX:154,021,573-154,097,717 (Size: 76,145 Total Exon Count: 3 Strand: -). Genomic coordinates for the coding region for this transcript variant include hg38 chrX:154,030,367-154,097,665 (Size: 67,299 Coding Exon Count: 3).

**[0196]** In some embodiments, the regulatory DNA element is located in a genomic region comprising the MeCP2 locus. In some embodiments, the target site is at, near, or within a MeCP2 locus.

**[0197]** In some embodiments, the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

**[0198]** In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 80% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 85% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 90% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 91% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 92% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 93% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 94% sequence identity to all or a portion of the target

site sequence described herein. In some aspects, the target site is a sequence having at least 95% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 96% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 97% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 98% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 99% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 99.5% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having at least 99.9% sequence identity to all or a portion of the target site sequence described herein. In some aspects, the target site is a sequence having 100% sequence identity to all or a portion of the target site sequence described herein.

**[0199]** In some embodiments, the target site is selected from the sequence set forth in any one of SEQ ID NOS:1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some embodiments, the target site is

**[0200]** In some embodiments, the target site comprises a sequence selected from any one of SEQ ID NOS:1-29, a contiguous portion thereof of at least 14 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:1-29 that is 14, 15, 16, 17, 18, 19, or 20 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a contiguous portion of a target site sequence described herein above. In some embodiments, the target site is the sequence set forth in any one of SEQ ID NOS:1-29.

**[0201]** In some embodiments, the target site comprises a sequence selected from any one of SEQ ID NOS:231-240, a contiguous portion thereof of at least 14 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:231-240 that is 14, 15, 16, 17, 18, 19, or 20 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a contiguous portion of a target site sequence described herein above. In some embodiments, the target site is the sequence set forth in any one of SEQ ID NOS:231-240.

**[0202]** In some embodiments, the target site comprises a sequence selected from any one of SEQ ID NOS:122 and 241-249, a contiguous portion thereof of at least 14 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:122 and 241-249 that is 14, 15, 16, 17, 18, 19, or 20 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a contiguous portion of a target site sequence described herein above. In some embodiments, the target site is the sequence set forth in any one of SEQ ID NOS:122 and 241-249.

**[0203]** In some embodiments, the target site comprises SEQ ID NO:1, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:2, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:3, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:4, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:5, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:6, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:7, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:8, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:10, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:11, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:12, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:13, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:14, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:15, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises



comprises SEQ ID NO:238, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:239, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:240, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:241, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:242, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:243, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:244, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:245, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:246, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:247, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:248, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:249, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof. In some embodiments, the target site comprises SEQ ID NO:122, a contiguous portion thereof of at least 14 nt, or a complementary sequence of thereof.

**[0204]** In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a contiguous portion of the sequence set forth in SEQ ID NO:9 that is 14, 15, 16, 17, 18, 19, or 20 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a contiguous portion of a target site sequence described herein above. In some embodiments, the target site is the sequence set forth in SEQ ID NO:9.

**[0205]** In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nucleotides, or a complementary sequence of any of the foregoing. In some embodiments, the target site is a contiguous portion of the sequence set forth in SEQ ID NO:27 that is 14, 15, 16, 17, 18, 19, or 20 nucleotides, or a

complementary sequence of any of the foregoing. In some embodiments, the target site is a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a contiguous portion of a target site sequence described herein above. In some embodiments, the target site is the sequence set forth in SEQ ID NO:27.

**[0206]** In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing. In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:9. In some embodiments, the target site comprises the sequence set forth in SEQ ID NO:27. In some embodiments, the target site comprises a complementary sequence of the sequence set forth in SEQ ID NO:9. In some embodiments, the target site comprises a complementary sequence of the sequence set forth in SEQ ID NO:27.

## **B. Guide RNAs (gRNAs)**

**[0207]** Provided herein are gRNAs, such as gRNAs that target or can bind to a regulatory DNA element of a MeCP2 locus. In some embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some embodiments, the gRNA comprises a gRNA spacer sequence (also known as a spacer sequence or a guide sequence) that is capable of hybridizing to the target site or is complementary to the target site, such as any target site described herein, for example, any target site in a genome. In some embodiments, the gRNA comprises a scaffold sequence that complexes with or binds to the Cas protein. In some embodiments, a gRNA specific to a target locus of interest (*e.g.* a regulatory DNA element of a MeCP2 locus) is used to recruit an RNA-guided protein (*c.g.* a Cas protein) or variant thereof or a fusion protein comprising such RNA-guided protein (*e.g.*, a Cas polypeptide), to the target site.

**[0208]** In some embodiments, the Cas protein (*e.g.* dCas9) is provided in combination or as a complex with one or more guide RNA (gRNA). In some aspects, the gRNA is a nucleic acid that promotes the specific targeting or homing of the gRNA/Cas RNP complex to the target site, such as any described above. In some embodiments, a target site of a gRNA may be referred to as a protospacer.

**[0209]** Provided herein are gRNAs, such as gRNAs that target or bind to a target site in a

MeCP2 gene or DNA regulatory element thereof, such as any described above in Section I.A. In some embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some embodiments, the gRNA comprises a gRNA spacer sequence (i.e. a spacer sequence or a guide sequence) that is capable of hybridizing to the target site, or that is complementary to the target site, such as any target site described in Section I.A or further below. In some embodiments, the gRNA comprises a scaffold sequence that complexes with or binds to the Cas protein.

**[0210]** In some aspects, a “gRNA molecule” is a nucleic acid that promotes the specific targeting or homing of a gRNA molecule/Cas9 molecule complex to a target nucleic acid, such as a locus on the genomic DNA of a cell. gRNA molecules can be unimolecular (having a single RNA molecule), sometimes referred to herein as “chimeric” gRNAs, or modular (comprising more than one, and typically two, separate RNA molecules). In general, a spacer sequence of the guide RNA, is any polynucleotide sequences comprising at least a sequence portion that has sufficient complementarity with a target polynucleotide sequence, such as the at the *MeCP2* locus in humans, to hybridize with the target sequence at the target site and direct sequence-specific binding of the CRISPR complex to the target sequence. In some embodiments, in the context of formation of a CRISPR complex, “target sequence” is to a sequence to which a spacer sequence is designed to have complementarity, where hybridization between the target sequence and a spacer sequence of the guide RNA promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. Generally, a spacer sequence is selected to reduce the degree of secondary structure within the spacer sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm.

**[0211]** In some embodiments, a guide RNA (gRNA) specific to a target locus of interest (e.g. at the *MeCP2* locus in humans) is used with RNA-guided nucleases or variants thereof, e.g., nuclease-inactive Cas variants, to target the provided DNA-targeting system to the target site or target position. Methods for designing gRNAs and exemplary spacer sequences are known. Exemplary gRNA structures that can be associated with particular RNA-guided nucleases or variants thereof, e.g., nuclease-inactive Cas variants, with particular domains and scaffold regions, are also known. In some aspects, gRNA molecules comprise a scaffold sequence, e.g., sequences that can be complexed with the Cas protein. In some aspects, the scaffold sequence is specific for the Cas protein.

**[0212]** In some embodiments, the gRNA is a chimeric gRNA. In general, gRNAs can be



unimolecular (i.e. composed of a single RNA molecule), or modular (comprising more than one, and typically two, separate RNA molecules). Modular gRNAs can be engineered to be unimolecular, wherein sequences from the separate modular RNA molecules are comprised in a single gRNA molecule, sometimes referred to as a chimeric gRNA, synthetic gRNA, or single gRNA. A guide RNA can comprise at least a spacer sequence that hybridizes to a target nucleic acid sequence of interest, and a CRISPR repeat sequence. In Type II systems, the gRNA also comprises a second RNA called the tracrRNA sequence. In the Type II guide RNA (gRNA), the CRISPR repeat sequence and tracrRNA sequence hybridize to each other to form a duplex. In the Type V guide RNA (gRNA), the crRNA forms a duplex. In both systems, the duplex can bind a site-directed polypeptide, such that the guide RNA and site-directed polypeptide form a complex. The gRNA can provide target specificity to the complex by virtue of its association with the site-directed polypeptide. The gRNA thus can direct the activity of the site-directed polypeptide.

**[0213]** In some embodiments, the chimeric gRNA is a fusion of two non-coding RNA sequences: a crRNA sequence and a tracrRNA sequence, for example as described in WO 2013/176772, or Jinek, M. et al. *Science* 337(6096):816-21 (2012). In some embodiments, the chimeric gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II CRISPR/Cas system, wherein the naturally occurring crRNA:tracrRNA duplex acts as a guide for the Cas protein, e.g., Cas9 protein. Exemplary types of CRISPR/Cas systems and associated gRNA structures include those described in, for example, Moon et al. *Exp. Mol. Med.* 51, 1-11 (2019), Zhang, F. Q. *Rev. Biophys.* 52, E6 (2019), Makarova et al. *Methods Mol. Biol.* 1311:47-75 (2015), WO 2013/176772, or Jinek, M. et al. *Science* 337(6096):816-21 (2012).

**[0214]** Methods for designing gRNAs and exemplary targeting domains can include those described in, e.g., International PCT Pub. Nos. WO 2014/197748, WO 2016/130600, WO 2017/180915, WO 2021/226555, WO 2013/176772, WO 2014/152432, WO 2014/093661, WO 2014/093655, WO 2015/089427, WO 2016/049258, WO 2016/123578, WO 2021/076744, WO 2014/191128, WO 2015/161276, WO 2017/193107, and WO 2017/093969.

**[0215]** In some aspects, the spacer sequence of a gRNA is a polynucleotide sequence comprising at least a portion that has sufficient complementarity with the target gene or DNA regulatory element thereof (e.g. any described in Section I.A) to hybridize with a target site in the target gene and direct sequence-specific binding of a CRISPR complex to the sequence of the target site. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. In some embodiments, the gRNA comprises a spacer sequence that is complementary, e.g., at least 80%,

85%, 90%, 95%, 98%, 99%, or 100% (e.g., fully complementary), to the target site. The strand of the target nucleic acid comprising the target site sequence may be referred to as the “complementary strand” of the target nucleic acid. In some aspects, the spacer sequence is a user-defined sequence. Guidance on the selection of spacer sequences can be found, e.g., in Fu *et al.*, Nat Biotechnol 2014 32:279–284 and Sternberg *et al.*, Nature 2014 507:62-67.

**[0216]** In some embodiments, the gRNA spacer sequence is between about 14 nucleotides (nt) and about 26 nt, between about 14 nt and about 24 nt, or between about 16 nt and 22 nt in length. In some embodiments, the gRNA spacer sequence is 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt or 22 nt, 23 nt, 24 nt, 25 nt, or 26 nt in length. In some embodiments, the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length. In some embodiments, the gRNA spacer sequence is 18 nt in length. In some embodiments, the gRNA spacer sequence is 19 nt in length. In some embodiments, the gRNA spacer sequence is 20 nt in length. In some embodiments, the gRNA spacer sequence is 21 nt in length. In some embodiments, the gRNA spacer sequence is 22 nt in length.

**[0217]** A target site of a gRNA may be referred to as a protospacer. In some aspects, the spacer is designed to target a protospacer with a specific protospacer-adjacent motif (PAM), i.e. a sequence immediately adjacent to the protospacer that contributes to and/or is required for Cas binding specificity. Different CRISPR/Cas systems have different PAM requirements for targeting. For example, in some embodiments, *S. pyogenes* Cas9 uses the PAM 5'-NGG-3' (SEQ ID NO:158), where N is any nucleotide. *S. aureus* Cas9 uses the PAM 5'-NNGRRT-3' (SEQ ID NO:159), where N is any nucleotide, and R is G or A. *N. meningitidis* Cas9 uses the PAM 5'-NNNNGATT-3' (SEQ ID NO:160), where N is any nucleotide. *C. jejuni* Cas9 uses the PAM 5'-NNNNRYAC-3' (SEQ ID NO:161) or 5'-NNNNACAC-3' (SEQ ID NO:216), where N is any nucleotide, R is G or A, and Y is C or T. *S. thermophilus* uses the PAM 5'-NNAGAAW-3' (SEQ ID NO:162), where N is any nucleotide and W is A or T. *F. Novicida* Cas9 uses the PAM 5'-NGG-3' (SEQ ID NO:158), where N is any nucleotide. *T. denticola* Cas9 uses the PAM 5'-NAAAAC-3' (SEQ ID NO:163), where N is any nucleotide. Cas12a (also known as Cpf1) from various species, uses the PAM 5'-TTTV-3' (SEQ ID NO:164), where V is A, C, or G. Phage-derived CasPhi (such as CasPhi-2, also known as Cas12j), uses the PAM 5'-TBN-3' (SEQ ID NO:214), where N is any nucleotide, and B is G, T, or C. Archaeal Un1Cas12f1 (also known as Cas14a1), uses the PAM 5'-TTTN-3' (SEQ ID NO:215), where N is any nucleotide. A Cas12f protein (also known as Cas14) uses the PAM 5'-TTTR-3' (SEQ ID NO:222), where R is G or A. A Cas12k protein uses the PAM 5'-GGTT-3' (SEQ ID NO:217). Cas proteins may use or be engineered to use different PAMs from those listed above. For example, variant

SpCas9 proteins may use a PAM selected from: 5'-NGG-3' (SEQ ID NO:158), 5'-NGAN-3' (SEQ ID NO:165), 5'-NGNG-3' (SEQ ID NO:166), 5'-NGAG-3' (SEQ ID NO:167), or 5'-NGCG-3' (SEQ ID NO:168), where N is any nucleotide. Methods for designing or identifying gRNA spacer sequences and/or protospacer sequences in a particular region, are known. gRNA spacer sequences and/or protospacer sequences can be determined based on the type of Cas protein used and the associated PAM sequence.

**[0218]** In some embodiments, the PAM of a gRNA for complexing with *S. pyogenes* Cas9 or variant thereof is set forth in SEQ ID NO:158. In some embodiments, the PAM of a gRNA for complexing with *S. aureus* Cas9 or variant thereof is set forth in SEQ ID NO:159. In some embodiments, the PAM of a gRNA for complexing with a Type V CRISPR/Cas system, such as with Cas12a (also known as Cpf1) or variant thereof is set forth in SEQ ID NO:164.

**[0219]** A spacer sequence may be selected to reduce the degree of secondary structure within the spacer sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm.

**[0220]** In some embodiments, the gRNA (including the spacer sequence) will comprise the base uracil (U), whereas DNA encoding the gRNA molecule will comprise the base thymine (T). While not wishing to be bound by theory, in some embodiments, it is believed that the complementarity of the spacer sequence (i.e. guide sequence) with the target sequence contributes to specificity of the interaction of the gRNA molecule/Cas molecule complex with a target nucleic acid. It is understood that in a spacer sequence (i.e. guide sequence) and target sequence pair, the uracil bases in the spacer sequence (i.e. guide sequence) will pair with the adenine bases in the target sequence. A gRNA spacer sequence herein may be defined by the DNA sequence encoding the gRNA spacer, and/or the RNA sequence of the spacer.

**[0221]** In some embodiments, the gRNA comprises modified nucleotides, e.g., for increased stability. In some embodiments, one, more than one, or all of the nucleotides of a gRNA can have a modification, e.g., to render the gRNA less susceptible to degradation and/or improve bio-compatibility. By way of non-limiting example, the backbone of the gRNA can be modified with a phosphorothioate, or other modification(s). In some cases, a nucleotide of the gRNA can comprise a 2' modification, e.g., a 2'-acetylation, e.g., a 2' methylation, or other modification(s).

**[0222]** In some embodiments the gRNA is a concatenation of two non-coding RNA sequences: a crRNA sequence and a tracrRNA sequence. The gRNA may target a desired DNA sequence by exchanging the sequence encoding a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target. gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II CRISPR/Cas system (e.g.,

Cas9). This duplex, which may include, for example, a 42-nucleotide crRNA and a 75-nucleotide tracrRNA, acts as a guide for the Cas9 protein to cleave the target nucleic acid. The “target region”, “target sequence” or “protospacer” as used interchangeably herein refers to the region of the target gene to which the CRISPR/Cas9-based system targets. The CRISPR/Cas9-based system may include two or more gRNAs, wherein the two or more gRNAs target different DNA sequences. The target DNA sequences may be overlapping or non-overlapping. The target DNA sequences may be located within or near the same gene or different genes. The target sequence or protospacer is followed by a PAM sequence at the 3' end of the protospacer. Different Type II systems have differing PAM requirements. For example, the *Streptococcus pyogenes* Type II system uses an “NGG” sequence, where “N” can be any nucleotide.

[0223] In some aspects, the gRNA comprises scaffold sequences. In some aspects, the scaffold sequence (in some cases including a crRNA sequence and/or a tracrRNA sequence) will be different depending on the Cas protein. In some aspects, different CRISPR/Cas systems have different gRNA scaffold sequences for associating with Cas protein. In some embodiments, an exemplary scaffold sequence for *S. aureus* Cas9 comprises a sequence set forth in SEQ ID NO:219, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:219. In some embodiments, an exemplary scaffold sequence for *S. aureus* Cas9 comprises a sequence set forth in SEQ ID NO:219. In some embodiments, an exemplary scaffold sequence for *S. pyogenes* Cas9 comprises a sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, an exemplary scaffold sequence for *S. pyogenes* Cas9 comprises a sequence set forth in SEQ ID NO:30. In some embodiments, an exemplary scaffold sequence for *Acidaminococcus* sp. Cas12a comprises a sequence set forth in SEQ ID NO:201, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:201. In some embodiments, an exemplary scaffold sequence for CasPhi-2 comprises a sequence set forth in SEQ ID NO:202, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:202. In some embodiments, an exemplary scaffold sequence for Un1Cas12f1 comprises a sequence set forth in SEQ ID NO:203, 204, or 205, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:203, 204, or 205. In some embodiments, an exemplary scaffold sequence for Un1Cas12f1 comprises a sequence set forth in SEQ ID NO:203, or a sequence

having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:203. In some embodiments, an exemplary scaffold sequence for Un1Cas12f1 comprises a sequence set forth in SEQ ID NO:204, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:204. In some embodiments, an exemplary scaffold sequence for Un1Cas12f1 comprises a sequence set forth in SEQ ID NO:205, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:205. In some embodiments, an exemplary scaffold sequence for C. jejuni Cas9 comprises a sequence set forth in SEQ ID NO:206, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:206. In some embodiments, an exemplary scaffold sequence for Cas12k comprises a sequence set forth in SEQ ID NO:207, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:207. In some embodiments, an exemplary scaffold sequence for CasMini comprises a sequence set forth in SEQ ID NO:208, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:208.

**[0224]** In some embodiments, the gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence comprises the sequence set forth in SEQ ID NO:30 (GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the scaffold sequence is set forth in SEQ ID NO:30.

**[0225]** In some aspects, the gRNA can target the DNA-targeting system can direct the activities of an associated polypeptide (e.g., fusion protein, DNA-targeting system, effector domain, etc.) to a specific target site within a target nucleic acid (e.g., regulatory DNA element of a MeCP2 locus).

**[0226]** In some embodiments, a gRNA provided herein targets a target site in a gene in a cell or DNA regulatory element thereof, wherein the gene is MeCP2.

**[0227]** In some embodiments, the gRNA targets a target site that comprises a sequence selected from any one of SEQ ID NOS:1-29, a contiguous portion thereof of at least 14 nucleotides, a complementary sequence of any of the foregoing, or a sequence having at or at

least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:1-29 that is 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides in length. In some embodiments, the target site is set forth in any one of SEQ ID NOS:1-29.

**[0228]** In some embodiments, the gRNA targets a target site that comprises a sequence selected from any one of SEQ ID NOS:231-240, a contiguous portion thereof of at least 14 nucleotides, a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:231-240 that is 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides in length. In some embodiments, the target site is set forth in any one of SEQ ID NOS:231-240.

**[0229]** In some embodiments, the gRNA targets a target site that comprises a sequence selected from any one of SEQ ID NOS:122 and 241-249, a contiguous portion thereof of at least 14 nucleotides, a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the target site is a contiguous portion of any one of SEQ ID NOS:122 and 241-249 that is 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides in length. In some embodiments, the target site is set forth in any one of SEQ ID NOS:122 and 241-249.

**[0230]** In some embodiments, the gRNA comprises a spacer sequence selected from any one of SEQ ID NOS:31-59, or a contiguous portion thereof of at least 14 nt, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the spacer sequence of the gRNA is a contiguous portion of any one of SEQ ID NOS:31-59 that is 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides in length. In some embodiments, the spacer sequence of the gRNA is set forth in any one of SEQ ID NOS:31-59.

**[0231]** In some embodiments, a gRNA provided herein comprises a spacer sequence selected from any one of SEQ ID NOS:31-59. In some embodiments, the gRNA further comprises a scaffold sequence set forth in SEQ ID NO:30. In some embodiments, the gRNA comprises the sequence selected from any one of SEQ ID NOS:61-89, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any one of SEQ ID NO:61-89. In some embodiments, the gRNA

is set forth in any one of SEQ ID NOS:61-89.

**[0232]** In some embodiments, the gRNA targets a target site in a MeCP2 locus or a DNA regulatory element thereof that comprises the sequence selected from any one of SEQ ID NO:231-240, a contiguous portion thereof of at least 14 nucleotides (e.g., 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence comprises the sequence set forth in SEQ ID NO:21, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 231, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 232, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 233, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 234, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 235, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 236, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 237, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 238, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 239, and a scaffold sequence of SEQ ID NO:219. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 240, and a scaffold sequence of SEQ ID NO:219. In some embodiments, a provided DNA-targeting system comprises any of the aforementioned gRNAs complexed with a Cas protein, such as a *S. aureus* Cas9 protein. In some embodiments, the Cas9 is a dCas9. In some embodiments, the dCas9 is a dSaCas9, such as a dSaCas9 set forth in SEQ ID NO:98, or a variant and/or fusion thereof.

**[0233]** In some embodiments, the gRNA targets a target site in a MeCP2 locus or a DNA regulatory element thereof that comprises the sequence selected from any one of SEQ ID NO:122 and 241-249, a contiguous portion thereof of at least 14 nucleotides (e.g., 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence comprises the sequence set forth in SEQ ID NO:201, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 241, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 242, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 243, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 244, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 245, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 246, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 247, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 248, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO: 249, and a scaffold sequence of SEQ ID NO:201. In some embodiments, the gRNA comprises, in 5' to 3' order, a spacer targeting SEQ ID NO:122, and a scaffold sequence of SEQ ID NO:201. In some embodiments, a provided DNA-targeting system comprises any of the aforementioned gRNAs complexed with a Cas protein, such as a Cas12a (also known as Cpf1) protein. In some embodiments, the Cas12a is a dCas12a. In some embodiments, the dCas12a is a dSaCas12a, such as a dSaCas12a set forth in SEQ ID NO:182, or a variant and/or fusion thereof.

**[0234]** In some embodiments, the gRNA targets a target site in MeCP2 or a DNA regulatory element thereof that comprises SEQ ID NO:9, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA comprises a spacer sequence comprising SEQ ID NO:39, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA further comprises a scaffold sequence. In some embodiments, the



scaffold sequence comprises the sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, the gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:69, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the gRNA targeting MeCP2 or a DNA regulatory element thereof, is set forth in SEQ ID NO:69. In some embodiments, a provided DNA-targeting system includes any of the above gRNAs complexed with a Cas protein, such as a Cas9 protein. In some embodiments, the Cas9 is a dCas9. In some embodiments, the dCas9 is a dSpCas9, such as a dSpCas9 set forth in SEQ ID NO:95.

**[0235]** In some embodiments, the gRNA targets a target site in MeCP2 or a DNA regulatory element thereof that comprises SEQ ID NO:27, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA comprises a spacer sequence comprising SEQ ID NO:57, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence comprises the sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, the gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:87, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the gRNA targeting MeCP2 or a DNA regulatory element thereof, is set forth in SEQ ID NO:87. In some embodiments, a provided DNA-targeting system includes any of the above gRNAs complexed with a Cas protein, such as a Cas9 protein. In some embodiments, the Cas9 is a dCas9. In some embodiments, the dCas9 is a dSpCas9, such as a dSpCas9 set forth in SEQ ID NO:95.

**[0236]** In some embodiments, any of the provided gRNA sequences is complexed with or is provided in combination with a Cas9. In some embodiments, the Cas9 is a dCas9. In some embodiments, the dCas9 is a dSpCas9, such as a dSpCas9 set forth in SEQ ID NO:95.

**[0237]** Also provided are guide RNAs (gRNAs) that binds a target site in a regulatory DNA

element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

**[0238]** In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) a gRNA; and the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt. In some embodiments, the gRNA further comprises the sequence set forth in SEQ ID NO:30. In some embodiments, the gRNA comprises the sequence set forth in SEQ ID NO:69. In some of any of the provided embodiments, the gRNA is set forth in SEQ ID NO:69.

**[0239]** In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) a gRNA; and the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt. In some embodiments, the gRNA further comprises the sequence set forth in SEQ ID NO:30. In some embodiments, the gRNA comprises the sequence set forth in SEQ ID NO:87. In some of any of the provided embodiments, the gRNA is set forth in SEQ ID NO:87.

**[0240]** In some embodiments, the gRNA comprises a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion of the gRNA sequence or a gRNA spacer sequence described herein.

### **C. Combinations of gRNAs**

**[0241]** Provided herein are combinations, such as combinations of gRNAs, that includes a first gRNA comprising any of the gRNAs described herein, and one or more second gRNAs that binds to a second target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus. In some embodiments, the second gRNA comprises any of the gRNAs described herein.

**[0242]** Also provided herein are combinations, such as combinations of gRNAs, that include: a first gRNA that binds a first target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158; and a second gRNA that binds a second target site in a regulatory DNA element of a MeCP2 locus, wherein the second target site is located within the genomic coordinates hg38

chrX:154,097,151-154,098,158.

**[0243]** In some aspects, the combination of gRNAs comprises a first gRNA and a second gRNA.

**[0244]** In some embodiments, the first gRNA targets a target site that comprises SEQ ID NO:9, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the first gRNA comprises a spacer sequence comprising SEQ ID NO:39, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the first gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence of the first gRNA comprises the sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, the first gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:69, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the first gRNA is set forth in SEQ ID NO:69. In any of the preceding embodiments, the second gRNA may be any gRNA disclosed herein.

**[0245]** In some embodiments, the second gRNA targets a target site that comprises SEQ ID NO:27, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the second gRNA comprises a spacer sequence comprising SEQ ID NO:57, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the scaffold sequence of the second gRNA comprises the sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, the second gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:87, or a sequence having at

or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the second gRNA is set forth in SEQ ID NO:87. In any of the preceding embodiments, the first gRNA may be any gRNA disclosed herein.

**[0246]** In some embodiments, the first gRNA targets a target site that comprises SEQ ID NO:9, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing, and the second gRNA targets a target site that comprises SEQ ID NO:27, a contiguous portion thereof of at least 14 nucleotides (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), a complementary sequence of any of the foregoing, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the first gRNA comprises a spacer sequence comprising SEQ ID NO:39, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing, and the second gRNA comprises a spacer sequence comprising SEQ ID NO:57, a contiguous portion thereof of at least 14 nt (e.g. 14, 15, 16, 17, 18, 19, 20, 21, or 22 nucleotides), or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to any of the foregoing. In some embodiments, the first and/or second gRNA further comprises a scaffold sequence. In some embodiments, the scaffold sequence of the first and/or second gRNA comprises the sequence set forth in SEQ ID NO:30, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to SEQ ID NO:30. In some embodiments, the first gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:69, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof, and the second gRNA, including a spacer sequence and a scaffold sequence, comprises SEQ ID NO:87, or a sequence having at or at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, or 100% sequence identity to all or a portion thereof. In some embodiments, the first gRNA is set forth in SEQ ID NO:69, and the second gRNA is set forth in SEQ ID NO:87.

**[0247]** In some embodiments, the first gRNA comprises a gRNA spacer sequence set forth

in any one of SEQ ID NO:1-29 or a contiguous portion thereof of at least 14 nt. In some embodiments, the first gRNA comprises a gRNA spacer sequence set forth in SEQ ID NO:9 or 27 or a contiguous portion thereof of at least 14 nt. In some embodiments, the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:1-29 or a contiguous portion thereof of at least 14 nt. In some embodiments, the second gRNA comprises a gRNA spacer sequence set forth in SEQ ID NO: 9 or 27 or a contiguous portion thereof of at least 14 nt.

**[0248]** In some embodiments, the combination comprises: the first gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:1-29 or a contiguous portion thereof of at least 14 nt; and the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:1-29 or a contiguous portion thereof of at least 14 nt.

**[0249]** In some embodiments, the combination comprises: the first gRNA comprises a gRNA spacer sequence set forth in SEQ ID NO:9 or a contiguous portion thereof of at least 14 nt; and the second gRNA comprises a gRNA spacer sequence set forth in SEQ ID NO:27 or a contiguous portion thereof of at least 14 nt.

**[0250]** In some embodiments, the first gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:231-240 or a contiguous portion thereof of at least 14 nt. In some embodiments, the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:231-240 or a contiguous portion thereof of at least 14 nt. In some embodiments, the combination comprises: the first gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:231-240 or a contiguous portion thereof of at least 14 nt; and the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:231-240 or a contiguous portion thereof of at least 14 nt.

**[0251]** In some embodiments, the first gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:122 and 241-249 or a contiguous portion thereof of at least 14 nt. In some embodiments, the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:122 and 241-249 or a contiguous portion thereof of at least 14 nt. In some embodiments, the combination comprises: the first gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:122 and 241-249 or a contiguous portion thereof of at least 14 nt; and the second gRNA comprises a gRNA spacer sequence set forth in any one of SEQ ID NO:122 and 241-249 or a contiguous portion thereof of at least 14 nt.

#### **D. DNA Targeting Domains**

**[0252]** In some embodiments, the provided DNA-targeting systems or fusion proteins

comprise a DNA-targeting domain. In some aspects, the DNA-targeting domain provides sequence specificity and targets the DNA targeting system or fusion protein at a particular location of the genome, such as a target site specified by a component of the DNA-targeting domain. In some embodiments, exemplary DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant of any of the foregoing. In some embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing. In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA. In some embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein. In some aspects, for a DNA-targeting domain that comprises a Cas-gRNA combination, the gRNA component (such as any described herein, for example, in Section II.B) provides the sequence specificity to target the DNA-targeting system, DNA-targeting domain or fusion protein to a target site specified by the gRNA.

### *1. Cas and Variants*

**[0253]** In some embodiments, the DNA-targeting systems comprise a DNA-targeting domain, that binds to a target site in a regulatory DNA element of a MeCP2 locus and comprises a Cas-guide RNA (gRNA) combination. In some embodiments, the Cas-gRNA combination includes a variant Cas protein that lacks nuclease activity or that is a deactivated Cas (dCas) protein. In some embodiments, the Cas-gRNA combination includes at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

**[0254]** In some aspects, the DNA-targeting domain comprises a CRISPR-associated (Cas) protein or variant thereof, or comprises a protein that is derived from a Cas protein or variant thereof. In particular embodiments here, the Cas protein is nuclease-inactive (i.e. is a dCas protein).

**[0255]** In some aspects, provided herein are DNA-targeting systems based on CRISPR/Cas systems, i.e. CRISPR/Cas-based DNA-targeting systems, that are able to bind to a target site in a MeCP2 gene or regulatory DNA element thereof. In some embodiments, the CRISPR/Cas DNA-targeting domain is nuclease inactive, such as includes a dCas (e.g. dCas9) so that the system binds to the target site in a target gene without mediating nucleic acid cleavage at the

target site. The CRISPR/Cas-based DNA-targeting systems may be used to modulate expression of MeCP2 in a cell. In some embodiments, the CRISPR/Cas-based DNA-targeting system can include any known Cas enzyme, such as a nuclease-inactive or dCas. In some embodiments, the CRISPR/Cas-based DNA-targeting system includes a fusion protein of a nuclease-inactive Cas protein or a variant thereof and an effector domain that increases transcription of a gene (e.g. a transcription activation domain), and at least one gRNA.

**[0256]** The CRISPR system (also known as CRISPR/Cas system, or CRISPR-Cas system) refers to a conserved microbial nuclease system, found in the genomes of bacteria and archaea, that provides a form of acquired immunity against invading phages and plasmids. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), refers to loci containing multiple repeating DNA elements that are separated by non-repeating DNA sequences called spacers. Spacers are short sequences of foreign DNA that are incorporated into the genome between CRISPR repeats, serving as a 'memory' of past exposures. Spacers encode the DNA-targeting portion of RNA molecules that confer specificity for nucleic acid cleavage by the CRISPR system. CRISPR loci contain or are adjacent to one or more CRISPR-associated (Cas) genes, which can act as RNA-guided nucleases for mediating the cleavage, as well as non-protein coding DNA elements that encode RNA molecules capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

**[0257]** CRISPR/Cas systems, such as those with Cas9, have been engineered to allow efficient programming of Cas/RNA RNPs to target desired sequences in cells of interest, both for gene-editing and modulation of gene expression. The tracrRNA and crRNA have been engineered to form a single chimeric guide RNA molecule, commonly referred to as a guide RNA (gRNA), for example as described in WO 2013/176772, WO 2014/093661, WO 2014/093655, Jinck et al. *Science* 337(6096):816-21 (2012), or Cong et al. *Science* 339(6121):819-23 (2013), and as described herein, for example, in Section II.B. The spacer sequence of the gRNA can be chosen by a user to target the Cas/gRNA RNP complex to a desired locus, e.g. a desired target site in the target gene, e.g., MeCP2.

**[0258]** CRISPR/Cas systems may be multi-protein systems or single effector protein systems. Multi-protein, or Class 1, CRISPR systems include Type I, Type III, and Type IV systems. In some aspects, Class 2 systems include a single effector molecule and include Type II, Type V, and Type VI. In some embodiments, the DNA targeting system comprises components of CRISPR/Cas systems, such as a Type I, Type II, Type III, Type IV, Type V, or Type VI CRISPR system. In some embodiments, the Cas protein is from a Class 1 CRISPR system (i.e. multiple Cas protein system), such as a Type I, Type III, or Type IV CRISPR

system. In some embodiments, the Cas protein is from a Class 2 CRISPR system (i.e. single Cas protein system), such as a Type II, Type V, or Type VI CRISPR system.

**[0259]** In some embodiments, the Cas protein is derived from a Cas9 protein or variant thereof, for example as described in WO 2013/176772, WO 2014/152432, WO 2014/093661, WO 2014/093655, Jinek, M. et al. *Science* 337(6096):816-21 (2012), Mali, P. et al. *Science* 339(6121):823-6 (2013), Cong, L. et al. *Science* 339(6121):819-23 (2013), Perez-Pinera, P. et al. *Nat. Methods* 10, 973–976 (2013), or Mali, P. et al. *Nat. Biotechnol.* 31, 833–838 (2013). Various CRISPR/Cas systems and associated Cas proteins for use in gene editing and regulation have been described, for example in Moon et al. *Exp. Mol. Med.* 51, 1–11 (2019), Zhang, F. Q. *Rev. Biophys.* 52, E6 (2019), and Makarova et al. *Methods Mol. Biol.* 1311:47-75 (2015).

**[0260]** Type I CRISPR/Cas systems employ a large multisubunit ribonucleoprotein (RNP) complex called Cascade that recognizes double-stranded DNA (dsDNA) targets. After target recognition and verification, Cascade recruits the signature protein Cas3, a fused helicase-nuclease, to degrade DNA.

**[0261]** In some embodiments, the Cas protein is from a Type II CRISPR system. Exemplary Cas proteins of a Type II CRISPR system include Cas9. In some embodiments, the Cas protein is from a Cas9 protein or variant thereof, for example as described in WO 2013/176772, WO 2014/152432, WO 2014/093661, WO 2014/093655, Jinek. et al. *Science* 337(6096):816-21 (2012), Mali et al. *Science* 339(6121):823-6 (2013), Cong et al. *Science* 339(6121):819-23 (2013), Perez-Pinera et al. *Nat. Methods* 10, 973–976 (2013), or Mali et al. *Nat. Biotechnol.* 31, 833–838 (2013). In Type II CRISPR/Cas systems with the Cas protein Cas9, two RNA molecules and the Cas9 protein form a ribonucleoprotein (RNP) complex to direct Cas9 nuclease activity. The CRISPR RNA (crRNA) contains a spacer sequence that is complementary to a target nucleic acid sequence (target site), and that encodes the sequence specificity of the complex. The trans-activating crRNA (tracrRNA) base-pairs to a portion of the crRNA and forms a structure that complexes with the Cas9 protein, forming a Cas/RNA RNP complex. Cas9 mediates cleavage of target DNA if a correct protospacer-adjacent motif (PAM) is also present at the 3' end of the protospacer. For protospacer targeting, the sequence must be immediately followed by the protospacer-adjacent motif (PAM), a short sequence recognized by the Cas9 nuclease that is required for DNA cleavage.

**[0262]** Different Type II systems have differing PAM requirements. The *S. pyogenes* CRISPR system may have the PAM sequence for this Cas9 (SpCas9) as 5'-NRG-3', where R is either A or G, and characterized the specificity of this system in human cells. A unique capability of the CRISPR/Cas9 system is the straightforward ability to simultaneously



target multiple distinct genomic loci by co-expressing a single Cas9 protein with two or more sgRNAs. For example, the *Streptococcus pyogenes* Type II system typically prefers to use an “NGG” sequence, where “N” can be any nucleotide, but also accepts other PAM sequences, such as “NAG” in engineered systems (Hsu et al., *Nature Biotechnology* (2013) doi:10.1038/nbt.2647). Similarly, the Cas9 derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT (SEQ ID NO:160), but has activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM (SEQ ID NO:212) (Esvelt et al. *Nature Methods* (2013) doi:10.1038/nmeth.2681). In another example, the Cas9 derived from *Campylobacter jejuni* typically uses 5'-NNNNACAC-3' (SEQ ID NO:216) or 5'-NNNNRYAC-3' (SEQ ID NO:161) PAM sequences, where “N” can be any nucleotide, “R” can be either guanine (G) or adenine (A), and “Y” can be either cytosine (C) or thymine (T). In some aspects, the PAM sequences for spacer targeting depends on the type, ortholog, variant or species of the Cas protein.

**[0263]** In some embodiments, the Cas9 protein comprises a sequence from a Cas9 molecule of *S. aureus*. In some embodiments, the Cas9 protein comprises a sequence set forth in SEQ ID NO:99 or SEQ ID NO:113, or a variant thereof, such as an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:99 or SEQ ID NO:113. In some embodiments, the Cas9 protein comprises a sequence from a Cas9 molecule of *S. pyogenes*. In some embodiments, the Cas9 protein comprises a sequence set forth in SEQ ID NO:96 or SEQ ID NO:112, or a variant thereof, such as an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:96 or SEQ ID NO:112.

**[0264]** In Type III systems, the RNP complex is multimeric with a helicoid structure similar to Cascade. In contrast to Type I CRISPR/Cas systems, the Type III RNP complex recognizes complementary RNA sequences instead of dsDNA. RNA recognition stimulates a nonspecific DNA cleavage activity of the exemplary Type III Cas10 nuclease that is part of the RNP complex, such that DNA cleavage is achieved cotranscriptionally.

**[0265]** In some embodiments, the Cas protein is from a Type V CRISPR system. Exemplary Cas proteins of a Type V CRISPR system include Cas12a (also known as Cpf1), Cas12b (also known as C2c1), Cas12e (also known as CasX), Cas12k (also known as C2c5), Cas14a, and Cas14b. In some embodiments, the Cas protein is from a Cas12 protein (i.e. Cpf1) or variant thereof, for example as described in WO 2017/189308, WO2019/232069 and Zetsche et al. *Cell*. 163(3):759-71 (2015).

**[0266]** Exemplary Type V systems include those based on a Cas12 effector, and the C-

terminus with only one RuvC endonuclease domain is the defining characteristic of the Type V systems. The RuvC nuclease domain cleaves dsDNA adjacent to protospacer adjacent motif (PAM) sequences and single-stranded DNA (ssDNA) nonspecifically. The Type V systems can be further divided into subtypes, each characterized by different signature proteins, PAM sequences, and properties. Non-limiting exemplary Cas proteins derived from Type V CRISPR systems include Cas12a (Cpf1), Un1Cas12f1, Cas12j (CasPhi, such as CasPhi-2), Cas12k, and CasMini. For example, Type V-A includes, for example, Cas12a, which uses “TTTV” (SEQ ID NO:164) PAM sequence, where “V” is adenine (A), cytosine (C), or guanine (G). Type V-F includes, for example, Cas12f, which can use “TTTR” (SEQ ID NO:222), where “R” is G or A, or “TTTN” (SEQ ID NO:215), where “N” is any nucleotide. Type V-K includes, for example, Cas12k, which uses “GGTT” (SEQ ID NO:217) PAM sequence.

**[0267]** In some embodiments, the Cas12a protein comprises a sequence from a Cas12a molecule of *Acidaminococcus sp.*, such as an AsCas12a set forth in SEQ ID NO:183 or SEQ ID NO:184, or a variant thereof, such as an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:183 or SEQ ID NO:184.

**[0268]** Non-limiting examples of Cas proteins or Cas orthologs, such as Cas9 orthologs, from other bacterial strains include but are not limited to, Cas proteins identified in *Acaryochloris marina* MBIC11017; *Acetohalobium arabaticum* DSM 5501; *Acidaminococcus sp.*; *Acidithiobacillus caldus*; *Acidithiobacillus ferrooxidans* ATCC 23270; *Alicyclobacillus acidocaldarius* LAA1; *Alicyclobacillus acidocaldarius* subsp. *acidocaldarius* DSM 446; *Allochromatium vinosum* DSM 180; *Ammonifex degensii* KC4; *Anabaena variabilis* ATCC 29413; *Arthrospira maxima* CS-328; *Arthrospira platensis* str. Paraca; *Arthrospira sp.* PCC 8005; *Bacillus pseudomycooides* DSM 12442; *Bacillus selenitireducens* MLS10; *Burkholderiales bacterium 1\_1\_47*; *Caldicelulosiruptor beccii* DSM 6725; *Campylobacter jejuni*; *Candidatus Desulforudis audaxviator* MP104C; *Caldicelulosiruptor hydrothermalis* 108; *Clostridium phage c-st*; *Clostridium botulinum* A3 str. Loch Maree; *Clostridium botulinum* Ba4 str. 657; *Clostridium difficile* QCD-63q42; *Crocospaera watsonii* WH 8501; *Cyanothece sp.* ATCC 51142; *Cyanothece sp.* CCY0110; *Cyanothece sp.* PCC 7424; *Cyanothece sp.* PCC 7822; *Exiguobacterium sibiricum* 255-15; *Finegoldia magna* ATCC 29328; *Ktedonobacter racemifer* DSM 44963; *Lactobacillus delbrueckii* subsp. *bulgaricus* PB2003/044-T3-4; *Lactobacillus salivarius* ATCC 11741; *Listeria innocua*; *Lyngbya sp.* PCC 8106; *Marinobacter sp.* ELB17; *Methanohalobium evestigatum* Z-7303; *Microcystis phage Ma-LMM01*; *Microcystis aeruginosa* NIES-843; *Microscilla marina* ATCC 23134; *Microcoleus chthonoplastes* PCC 7420; *Neisseria*

meningitidis; Nitrosococcus halophilus Nc4; Nocardiosis dassonvillei subsp. dassonvillei DSM 43111; Nodularia spumigena CCY9414; Nostoc sp. PCC 7120; Oscillatoria sp. PCC 6506; Pelotomaculum\_thermopropionicum SI; Petrotoga mobilis SJ95; Polaromonas naphthalenivorans CJ2; Polaromonas sp. JS666; Pseudoalteromonas haloplanktis TAC125; Streptomyces pristinaespiralis ATCC 25486; Streptomyces pristinaespiralis ATCC 25486; Streptococcus thermophilus; Streptomyces viridochromogenes DSM 40736; Streptosporangium roseum DSM 43021; Synechococcus sp. PCC 7335; and Thermosiphon africanus TCF52B (Chylinski et al., RNA Biol., 2013; 10(5): 726-737).

**[0269]** In some embodiments, the DNA-targeting systems or fusion proteins comprise a Cas protein, such as a Cas protein set forth in any one of SEQ ID NOS:96, 99, 112, 113, 183, 184, 187-190, and 195-198, or a variant thereof, such as an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOS:96, 99, 112, 113, 183, 184, 187-190, and 195-198. In some embodiments, the Cas protein of any of the DNA-targeting systems or fusion proteins provided herein comprise a sequence set forth in any one of SEQ ID NOS:96, 99, 112, 113, 183, 184, 187-190, and 195-198, or a variant thereof, such as an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOS:96, 99, 112, 113, 183, 184, 187-190, and 195-198. In some aspects, the Cas protein lacks an initial methionine residue. In some aspects, the Cas protein comprises an initial methionine residue.

**[0270]** In some aspects, in the provided DNA-targeting systems and fusion proteins, the DNA-targeting domain, e.g., Cas, is a deactivated Cas (dCas), or a nuclease-inactive Cas (iCas). In some embodiments, the component of the DNA-targeting domain, such as a protein component, comprises a Cas9 variant such as a deactivated Cas9 or inactivated Cas9. In some embodiments, the component of the DNA-targeting domain, such as a protein component, comprises a Cas12a variant such as a deactivated Cas12a (Cpf1) or inactivated Cas12a (Cpf1). In some aspects, the Cas9 protein may be mutated so that the nuclease activity is deactivated or inactivated (also referred to as dCas9 or iCas9). In some aspects, the Cas protein is a variant that lacks nuclease activity (i.e. is a dCas protein). In some embodiments, the Cas protein is mutated so that nuclease activity is reduced or eliminated. Such Cas proteins are referred to as deactivated Cas or dead Cas (dCas) or nuclease-inactive Cas (iCas) proteins, as referred to interchangeably herein. In some embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9, or iCas9) protein. In some embodiments, the variant Cas protein is a variant Cpf1 protein that lacks nuclease activity or that is a deactivated Cas12a (dCas12a, or iCas12a) protein.

[0271] In some embodiments, Cas proteins are engineered to be catalytically inactivated or nuclease inactive to allow targeting of Cas/gRNA RNPs without inducing cleavage at the target site. Mutations in Cas proteins can reduce or abolish nuclease activity of the Cas protein, rendering the Cas protein catalytically inactive. Cas proteins with reduced or abolished nuclease activity are referred to as deactivated Cas (dCas), or nuclease-inactive Cas (iCas) proteins, as referred to interchangeably herein. In some aspects, the dCas or iCas can still bind to target site in the DNA in a site- and/or sequence-specific manner, as long as it retains the ability to interact with the guide RNA (gRNA) which directs the Cas-gRNA combination to the target site.

[0272] In some aspects, the dCas or iCas exhibits reduced or no endodeoxyribonuclease activity. For example, an exemplary dCas or iCas, for example dCas9 or iCas9, exhibits less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 1%, or less than about 0.1%, of the endodeoxyribonuclease activity of a wild-type Cas protein, e.g., a wild-type Cas9 protein. In some embodiments, the dCas or iCas, for example dCas9 or iCas9, exhibits substantially no detectable endodeoxyribonuclease activity. In some embodiments, an exemplary dCas or iCas, for example dCas9 or iCas9, comprises one or more amino acid mutations, substitutions, deletions or insertions at a position corresponding to a position selected from D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987, with reference to a wild-type *Streptococcus pyogenes* Cas9 (SpCas9), for example, with reference to numbering of positions of a SpCas9 sequence set forth in SEQ ID NO:112. In some aspects, the dCas9 or iCas9 comprises one or more amino acid mutations, substitutions, deletions or insertions corresponding to D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A, with reference to a wild-type *Streptococcus pyogenes* Cas9 (SpCas9), for example, with reference to numbering of positions of a SpCas9 sequence set forth in SEQ ID NO:112. Corresponding positions for mutations can be determined based on sequence alignments and determination of sequence conservation, for example, as described in WO 2013/171772 for Cas9 proteins from various species. In some aspects, the Cas protein lacks an initial methionine residue. In some aspects, the Cas protein comprises an initial methionine residue.

[0273] In some embodiments, the dCas9 protein can comprise a sequence from a Cas9 molecule, or variant thereof. In some embodiments, the dCas9 protein can comprise a sequence derived from a Cas9 molecule of *S. pyogenes*, *S. thermophilus*, *S. aureus*, *N. meningitidis*, *F. novicida*, *S. canis*, *S. auricularis*, or variant thereof. In some embodiments, the dCas9 protein comprises a sequence from a Cas9 molecule of *S. aureus*. In some embodiments, the dCas9 protein comprises a sequence from a Cas9 molecule of *S. pyogenes*. In some embodiments, the

dCas9 protein comprises a sequence from a Cas9 molecule of *C. jejuni*.

[0274] Exemplary deactivated Cas9 (dCas9) derived from *S. pyogenes* contains silencing mutations of the RuvC and HNH nuclease domains (D10A and H840A), for example as described in WO 2013/176772, WO 2014/093661, Jinek et al. *Science* 337(6096):816-21 (2012), and Qi et al. *Cell* 152(5):1173-83 (2013). Exemplary dCas variants derived from the Cas12 system (i.e. Cpf1) are described, for example in WO 2017/189308 and Zetsche et al. *Cell* 163(3):759-71 (2015). Conserved domains that mediate nucleic acid cleavage, such as RuvC and HNH endonuclease domains, are readily identifiable in Cas orthologs, and can be mutated to produce inactive variants, for example as described in Zetsche et al. *Cell* 163(3):759-71 (2015). Other exemplary Cas orthologs or variants include engineered variants based on a Cas12f (also known as Cas14), including those described in Xu et al., *Mol. Cell* 81(20):4333-4345 (2021).

[0275] In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA. In some embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein. In some embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some embodiments, the gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site (e.g., in a MeCP2 locus).

[0276] In some embodiments, the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof. In some embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, which lacks an initial methionine residue. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:190, which includes an initial methionine residue.

[0277] In some embodiments, the Cas protein or a variant thereof is a Cas9 protein or a variant thereof. In some embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein. In some embodiments, the Cas9 protein or a variant thereof is a *Staphylococcus aureus* Cas9 (SaCas9) protein or a variant thereof. In some embodiments, the variant Cas9 is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A,

with reference to numbering of positions of SEQ ID NO:99. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, which lacks an initial methionine residue. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:179, which includes an initial methionine residue.

**[0278]** In some embodiments, the Cas9 protein or variant thereof is a *Campylobacter jejuni* Cas9 (CjCas9) protein or a variant thereof. In some embodiments, the variant Cas9 comprises at least one amino acid mutation compared to the sequence set forth in SEQ ID NO:195 or 196. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:193, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:194, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0279]** In some embodiments, the Cas protein or a variant thereof is a Cas12a protein or a variant thereof. In some embodiments, the variant Cas protein is a variant Cas12a protein that lacks nuclease activity or that is a deactivated Cas12a (dCas12a) protein. In some embodiments, the Cas12a protein or variant thereof is a *Acidaminococcus sp.* Cas12a (AsCas12a) protein or a variant thereof. In some embodiments, the variant Cas12a is a *Acidaminococcus sp.* dCas12a (dAsCas12a) protein that comprises at least one amino acid mutation compared to the sequence set forth in SEQ ID NO:183 or 184. In some embodiments, the variant Cas12a protein comprises the sequence set forth in SEQ ID NO:181, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Cas12a protein comprises the sequence set forth in SEQ ID NO:182, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Cas12a protein comprises the sequence set forth in SEQ ID NO:182, which lacks an initial methionine residue. In some embodiments, the variant Cas12a protein comprises the sequence set forth in SEQ ID NO:181, which includes an initial methionine residue.

**[0280]** In some embodiments, the Cas protein or a variant thereof is a CasPhi-2 protein or a variant thereof. In some embodiments, the variant Cas protein is a variant CasPhi-2 protein that lacks nuclease activity or that is a deactivated CasPhi-2 (dCasPhi-2) protein. In some embodiments, the variant CasPhi-2 comprises at least one amino acid mutation compared to the

sequence set forth in SEQ ID NO:187 or 188. In some embodiments, the variant CasPhi-2 protein comprises the sequence set forth in SEQ ID NO:185, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant CasPhi-2 protein comprises the sequence set forth in SEQ ID NO:186, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant CasPhi-2 protein comprises the sequence set forth in SEQ ID NO:186, which lacks an initial methionine residue. In some embodiments, the variant CasPhi-2 protein comprises the sequence set forth in SEQ ID NO:185, which includes an initial methionine residue.

**[0281]** In some embodiments, the Cas protein or a variant thereof is a Un1Cas12f1 protein or a variant thereof. In some embodiments, the variant Cas protein is a variant Un1Cas12f1 protein that lacks nuclease activity or that is a deactivated Un1Cas12f1 (dUn1Cas12f1) protein. In some embodiments, the variant Un1Cas12f1 comprises at least one amino acid mutation compared to the sequence set forth in SEQ ID NO:189 or 190. In some embodiments, the variant Un1Cas12f1 protein comprises the sequence set forth in SEQ ID NO:191, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Un1Cas12f1 protein comprises the sequence set forth in SEQ ID NO:192, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the variant Un1Cas12f1 protein comprises the sequence set forth in SEQ ID NO:192, which lacks an initial methionine residue. In some embodiments, the variant Un1Cas12f1 protein comprises the sequence set forth in SEQ ID NO:191, which includes an initial methionine residue.

**[0282]** In some embodiments, the Cas protein or a variant thereof is a Cas12k protein or a variant thereof. In some embodiments, the Cas12k protein comprises the sequence set forth in SEQ ID NO:197, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the Cas12k protein comprises the sequence set forth in SEQ ID NO:198, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the Cas12k protein comprises the sequence set forth in SEQ ID NO:198, which lacks an initial methionine residue. In some embodiments, the Cas12k protein comprises the sequence set forth in SEQ ID NO:197, which includes an initial methionine residue.

**[0283]** In some embodiments, the Cas protein or a variant thereof is a CasMini protein or a variant thereof, such as an engineered Cas protein or variant based on a Cas12f (also known as Cas14), including those described in Xu et al., Mol. Cell 81(20):4333-4345 (2021) or set forth in

SEQ ID NO:213. In some embodiments, the variant Cas protein is a variant CasMini protein that lacks nuclease activity or that is a deactivated CasMini (dCasMini) protein. In some embodiments, the variant CasMini comprises at least one amino acid mutation compared to the sequence set forth in SEQ ID NO:213. In some embodiments, the variant CasMini protein comprises the sequence set forth in SEQ ID NO:213, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the CasMini protein comprises the sequence set forth in SEQ ID NO:213. In some embodiments, the variant CasMini protein comprises the sequence set forth in SEQ ID NO:199 or 200, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the CasMini protein comprises the sequence set forth in SEQ ID NO:199, which lacks an initial methionine residue. In some embodiments, the CasMini protein comprises the sequence set forth in SEQ ID NO:200, which includes an initial methionine residue.

**[0284]** DNA-targeting systems, in some cases comprising a fusion protein, such as dCas-fusion proteins include fusion of the Cas with an effector domain, such as a TET domain. Any of a variety of effector domains, for example those that increase, re-activate or de-repress transcription from the target locus, e.g., MeCP2 locus, including any described herein, for example, in Section II.D, can be used.

**[0285]** In some aspects, provided is a DNA-targeting system comprising a fusion protein comprising a DNA-targeting domain comprising a nuclease-inactive Cas protein or variant thereof, and an effector domain for increasing or inducing transcriptional de-repression or re-activation (e.g., TET domain) when targeted to a target site in a MeCP2 gene or regulatory element thereof. In some aspects, the DNA-targeting system also includes one or more gRNA, provided in combination or as a complex with the dCas protein or variant thereof, for targeting of the DNA-targeting system to the target site. In some embodiments, the fusion protein is guided to a specific target site sequence of the target gene by the guide RNA, wherein the effector domain mediates targeted epigenetic modification to increase, de-repress or promote transcription of the target gene.

## **2. Other Domains**

**[0286]** In some of any of the provided embodiments, the DNA-targeting domain comprises a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof. In some embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.



[0287] In some aspects, types of DNA-targeting domains include domains from proteins that can recognize nucleic acid sequences (e.g., target site) in a sequence-specific manner.

[0288] In some embodiments, a “zinc finger DNA binding protein” (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP. Among the ZFPs are artificial, or engineered, ZFPs, comprising ZFP domains targeting specific DNA sequences, typically 9-18 nucleotides long, generated by assembly of individual fingers. ZFPs include those in which a single finger domain is approximately 30 amino acids in length and contains an alpha helix containing two invariant histidine residues coordinated through zinc with two cysteines of a single beta turn, and having two, three, four, five, or six fingers. Generally, sequence-specificity of a ZFP may be altered by making amino acid substitutions at the four helix positions (–1, 2, 3, and 6) on a zinc finger recognition helix. Thus, for example, the ZFP or ZFP-containing molecule is non-naturally occurring, *e.g.*, is engineered to bind to a target site of choice.

[0289] In some cases, the DNA-targeting system is or comprises a zinc-finger DNA binding domain fused to an effector domain. In some embodiments, zinc fingers are custom-designed (i.e. designed by the user), or obtained from a commercial source. Various methods for designing zinc finger proteins are available. For example, methods for designing zinc finger proteins to bind to a target DNA sequence of interest are described, for example in Liu, Q. et al., PNAS, 94(11):5525-30 (1997); Wright, D.A. et al., Nat. Protoc., 1(3):1637-52 (2006); Gersbach, C.A. et al., Acc. Chem. Res., 47(8):2309-18 (2014); Bhakta M.S. et al., Methods Mol. Biol., 649:3-30 (2010); and Gaj et al., Trends Biotechnol, 31(7):397-405 (2013). In addition, various web-based tools for designing zinc finger proteins to bind to a DNA target sequence of interest are publicly available. See, for example, the Zinc Finger Tools design web site from Scripps available on the world wide web at [scripps.edu/barbas/zfdesign/zfdesignhome.php](http://scripps.edu/barbas/zfdesign/zfdesignhome.php). Various commercial services for designing zinc finger proteins to bind to a DNA target sequence of interest are also available. See, for example, the commercially available services or kits offered by Creative Biolabs (world wide web at [creative-biolabs.com/Design-and-Synthesis-of-Artificial-Zinc-Finger-Proteins.html](http://creative-biolabs.com/Design-and-Synthesis-of-Artificial-Zinc-Finger-Proteins.html)), the Zinc Finger Consortium Modular Assembly Kit available from Addgene (world wide web at [addgene.org/kits/zfc-modular-assembly/](http://addgene.org/kits/zfc-modular-assembly/)), or the CompoZr Custom ZFN Service from Sigma Aldrich (world wide web at [sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/custom-zfn.html](http://sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/custom-zfn.html)). For example, platforms for zinc-

finger construction are available that provide specifically targeted zinc fingers for thousands of targets. See, e.g., Gaj *et al.*, *Trends in Biotechnology*, 2013, 31(7), 397-405. Some gene-specific engineered zinc fingers are available commercially. In some cases, commercially available zinc fingers are used or are custom designed.

**[0290]** In some aspects, the DNA-targeting domain is a domain from Transcription activator-like effectors (TALEs). TALEs are proteins found in *Xanthomonas* bacteria. TALEs comprise a plurality of repeated amino acid sequences, each repeat having binding specificity for one base in a target sequence. Each repeat comprises a pair of variable residues in position 12 and 13 (repeat variable diresidue; RVD) that determine the nucleotide specificity of the repeat. In some embodiments, RVDs associated with recognition of the different nucleotides are HD for recognizing C, NG for recognizing T, NI for recognizing A, NN for recognizing G or A, NS for recognizing A, C, G or T, HG for recognizing T, IG for recognizing T, NK for recognizing G, HA for recognizing C, ND for recognizing C, HI for recognizing C, HN for recognizing G, NA for recognizing G, SN for recognizing G or A and YG for recognizing T, TL for recognizing A, VT for recognizing A or G and SW for recognizing A. In some embodiments, RVDs can be mutated towards other amino acid residues in order to modulate their specificity towards nucleotides A, T, C and G and in particular to enhance this specificity. Binding domains with similar modular base-per-base nucleic acid binding properties can also be derived from different bacterial species. These alternative modular proteins may exhibit more sequence variability than TALE repeats.

**[0291]** In some embodiments, a “TALE DNA binding domain” or “TALE” is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains, each comprising a repeat variable diresidue (RVD), are involved in binding of the TALE to its cognate target DNA sequence. A single “repeat unit” (also referred to as a “repeat”) is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a TALE protein. TALE proteins may be designed to bind to a target site using canonical or non-canonical RVDs within the repeat units. See, e.g., U.S. Pat. Nos. 8,586,526 and 9,458,205.

**[0292]** In some embodiments, a TALE is a fusion protein comprising a nucleic acid binding domain derived from a TALE and an effector domain. In some embodiments, one or more sites in the *MeCP2* locus can be targeted by engineered TALEs.

**[0293]** Zinc finger and TALE DNA-binding domains can be engineered to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a zinc finger protein, by engineering of the amino acids in a TALE repeat involved in DNA binding (the repeat variable diresidue or RVD region), or by

systematic ordering of modular DNA-binding domains, such as TALE repeats or ZFP domains. Therefore, engineered zinc finger proteins or TALE proteins are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering zinc finger proteins and TALEs are design and selection. A designed protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP or TALE designs (canonical and non-canonical RVDs) and binding data. See, for example, U.S. Pat. Nos. 9,458,205; 8,586,526; 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

#### **E. Effector Domains**

**[0294]** In some embodiments, the DNA-targeting system also includes at least one effector domain. In some embodiments, the DNA-targeting domain or a component thereof is fused to the at least one effector domain. In some embodiments, provided herein is a DNA-targeting system comprising a fusion protein comprising: (a) a DNA-targeting domain capable of being targeted to a target site at a MeCP2 locus or a regulatory element thereof, such as any described herein, and (b) at least one effector domain. In some aspects, the effector domain leads to an increase in transcription of MeCP2, or is capable of increasing transcription of MeCP2. In some aspects, the effector domain comprises a transcription activation domain. In some aspects, the effector domain comprises a domain that induces an epigenetic modification, such as demethylation.

**[0295]** In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA, and the component thereof fused to the at least one effector domain is the Cas protein or a variant thereof.

**[0296]** In some aspects, the effector domain activates, induces, catalyzes, or leads to demethylation, de-repression and/or increased transcription of MeCP2 when ectopically recruited to MeCP2 or a DNA regulatory element thereof. Exemplary fusion of DNA-targeting domain and at least one effector domain include fusing dCas9 with TET1 can result in robust induction of gene expression.

**[0297]** In some aspects, the effector domain activates, induces, catalyzes, or leads to demethylation and/or increased transcription of MeCP2 when ectopically recruited to MeCP2 or a DNA regulatory element thereof. In some embodiments, the effector domain induces,

catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. In some embodiments, the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation. In some embodiments, the effector domain induces transcription de-repression.

**[0298]** In some embodiments, the effector domain induces transcription activation. In some embodiments, the effector domain has one of the aforementioned activities itself (i.e. acts directly). In some embodiments, the effector domain recruits and/or interacts with a polypeptide domain that has one of the aforementioned activities (i.e. acts indirectly).

### *I. Exemplary Effector Domains*

**[0299]** In some embodiments, the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation. In some embodiments, the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, or transcription elongation. In some embodiments, the effector domain induces transcription de-repression. In some embodiments, the effector domain activates transcription from one or more regulatory elements (e.g., promoters and/or enhancers) from the target locus, e.g., MeCP2. In some embodiments, the effector domain induces transcription activation. In some embodiments, the effector domain has one of the aforementioned activities itself (i.e. acts or catalyzes directly). In some embodiments, the effector domain recruits and/or interacts with another cellular component (e.g., transcription factor) that has one of the aforementioned activities (i.e. acts or catalyzes indirectly).

**[0300]** Gene expression of endogenous mammalian genes, such as human genes, can be achieved by targeting a fusion protein comprising a DNA-targeting domain, such as a dCas9, and an effector domain, such as a transcription activation domain, to mammalian genes or regulatory DNA elements thereof (e.g. a promoter or enhancer), e.g. via one or more gRNAs. Any of a variety of effector domains for transcriptional activation (e.g. transcription activation domains) are known and can be used in accord with the provided embodiments. Transcription

activation domains, as well as activation of target genes by Cas fusion proteins (with a variety of Cas molecules) and the transcription activation domains, are described, for example, in WO 2014/197748, WO 2016/130600, WO 2017/180915, WO 2021/226555, WO 2021/226077, WO 2013/176772, WO 2014/152432, WO 2014/093661, Adli, M. *Nat. Commun.* 9, 1911 (2018), Perez-Pinera et al. *Nat. Methods* 10, 973–976 (2013), Mali et al. *Nat. Biotechnol.* 31, 833–838 (2013), and Maeder et al. *Nat. Methods* 10, 977–979 (2013).

**[0301]** In some embodiments, the effector domain comprises a transcriptional activator domain described in WO 2021/226077.

**[0302]** In some aspects, de-repression, activation or increase in gene expression of MeCP2 is achieved by targeting a fusion protein comprising a DNA-targeting domain, such as a dCas9, and an effector domain, such as a transcription activation domain, to a MeCP2 locus or regulatory DNA elements thereof (e.g. a promoter or enhancer) via one or more gRNAs. In some aspects, the one or more target sites of the one or more gRNA is at a MeCP2 locus or regulatory DNA elements thereof (e.g., a promoter or enhancer), for example, as described herein, for example, in Section II.A and II.B. Any of a variety of effector domains for transcriptional activation (e.g. transcription activation domains) are known and can be used in accord with the provided embodiments.

**[0303]** In some embodiments, the effector domain may comprise a TET protein (e.g. TET1, TET2, TET3), VP64, p65, Rta, p300, CBP, HSF1, VPR, VPH, SunTag, a partially or fully functional fragment or domain thereof, or a combination of any of the foregoing. In some embodiments, the effector domain comprises a catalytic domain of TET1.

**[0304]** In some embodiments, the effector domain may have demethylase activity. The effector domain can include an enzyme that removes methyl (CH<sub>3</sub>-) groups from nucleic acids, proteins (in particular histones), and other molecules. Alternatively, the effector can convert the methyl group to hydroxymethylcytosine in a mechanism for demethylating DNA. The effector domain can catalyze this reaction. For example, the effector domain that catalyzes this reaction may comprise a domain from a TET protein, for example TET1 (Ten-eleven translocation methylcytosine dioxygenase 1). TET1, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2021/226555.

**[0305]** In some embodiments, the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof. In some embodiments, the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof. An exemplary TET1 catalytic domain is set forth in SEQ ID NO:93. In some embodiments, the

effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0306]** In some embodiments, the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 2 (TET2) or a portion or a variant thereof. An exemplary TET2 protein is set forth in SEQ ID NO:169. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:169, or a portion thereof (such as a catalytic domain), or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0307]** In some embodiments, the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 3 (TET3) or a portion or a variant thereof. An exemplary TET3 protein is set forth in SEQ ID NO:170. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:170, or a portion thereof (such as a catalytic domain), or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0308]** In some embodiments, the effector domain may comprise a VP64 domain. For example, dCas9-VP64 can be targeted to a target site by one or more gRNAs to activate a gene. VP64 is a polypeptide composed of four tandem copies of VP16, a 16 amino acid transactivation domain of the Herpes simplex virus. VP64 domains, including in dCas fusion proteins, have been described, for example, in WO 2014/197748, WO 2013/176772, WO 2014/152432, and WO 2014/093661. In some embodiments, the effector domain comprises at least one VP16 domain, or a VP16 tetramer (“VP64”) or a variant thereof. An exemplary VP64 domain is set forth in SEQ ID NO:171. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:171, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0309]** In some embodiments, the effector domain may comprise a p65 activation domain (p65AD). p65AD is the principal transactivation domain of the 65kDa polypeptide of the nuclear form of the NF-KB transcription factor. An exemplary sequence of human transcription factor p65 is available at the Uniprot database under accession number Q04206. p65 domains, including in dCas fusion proteins, have been described, for example in WO 2017/180915 and Chavez, A. et al. Nat. Methods 12, 326–328 (2015). An exemplary p65 activation domain is set forth in SEQ ID NO:172. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:172, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0310]** In some embodiments, the effector domain may comprise a R transactivator (Rta) domain. Rta is an immediate-early protein of Epstein-Barr virus (EBV), and is a transcriptional activator that induces lytic gene expression and triggers virus reactivation. The Rta domain, including in dCas fusion proteins, has been described, for example in WO 2017/180915 and Chavez, A. et al. *Nat. Methods* 12, 326–328 (2015). An exemplary Rta domain is set forth in SEQ ID NO:173. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:173, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0311]** In some embodiments, the effector domain may have histone acetyltransferase activity. For example, the effector domain may comprise a domain from p300 or CREB-binding protein (CBP) protein. The effector domain may comprise a p300 domain. p300 functions as a histone acetyltransferase that regulates transcription via chromatin remodeling and is involved with the processes of cell proliferation and differentiation. The p300 domain, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2016/130600 and WO 2017/180915. An exemplary p300 domain is set forth in SEQ ID NO:174. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:174, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0312]** “p300 protein,” “EP300,” or “E1A binding protein p300” as used interchangeably herein refers to the adenovirus E1A-associated cellular p300 transcriptional co-activator protein encoded by the EP300 gene. p300 is a highly conserved acetyltransferase involved in a wide range of cellular processes. p300 functions as a histone acetyltransferase that regulates transcription via chromatin remodeling and is involved with the processes of cell proliferation and differentiation.

**[0313]** The p300 domain, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2016/130600 and WO 2017/180915. An exemplary p300 domain sequence is set forth in SEQ ID NO:174. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO: 174, a domain thereof, a portion thereof, or a variant thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing. In some embodiments, the effector domain comprises p300 or a domain thereof, a portion thereof, or a variant thereof. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:174, or a domain thereof, a portion thereof, or a variant thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the

foregoing.

**[0314]** In some embodiments, the effector domain may comprise a HSF1 domain. HSF1 is a gene that encodes Heat shock factor protein 1. HSF1, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2021/226555, WO 2015/089427, and Konermann et al. *Nature* 517(7536):583-8 (2015). An exemplary HSF1 domain is set forth in SEQ ID NO:175. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:175, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0315]** In some embodiments, the effector domain may comprise a eukaryotic release factor domain, for example from eukaryotic release factor 1 (ERF1) or eukaryotic release factor 3 (ERF3). The effector domain may have transcription release factor activity. The effector domain may have eukaryotic release factor 1 (ERF1) activity or eukaryotic release factor 3 (ERF3) activity.

**[0316]** In some embodiments, the effector domain may comprise the tripartite activator VP64-p65-Rta (also known as VPR). VPR comprises three transcription activation domains (VP64, p65, and Rta) fused by short amino acid linkers, and can effectively upregulate target gene expression. VPR, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2021/226555 and Chavez, A. et al. *Nat. Methods* 12, 326–328 (2015). An exemplary VPR polypeptide is set forth in SEQ ID NO:176. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:176, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0317]** In some embodiments, the effector domain may comprise VPH. VPH is a polypeptide comprising VP64, mouse p65, and HSF1. VPH, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2021/226555. An exemplary VPH polypeptide is set forth in SEQ ID NO:136. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:136, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0318]** In some embodiments, the effector domain may comprise a LSD1 domain. LSD1 (also known as Lysine-specific histone demethylase 1A) is a histone demethylase that can demethylate lysine residues of histone H3, thereby acting as a coactivator or a corepressor, depending on the context. LSD1, including in dCas fusion proteins, has been described, for example, in WO 2013/176772, WO 2014/152432, and Kearns, N. A. et al. *Nat. Methods*.



12(5):401–403 (2015). An exemplary LSD1 polypeptide is set forth in SEQ ID NO:123. In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:123, a domain thereof, a portion thereof, or a variant thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0319]** In some embodiments, the effector domain may comprise a SunTag domain. SunTag is a repeating peptide array, which can recruit multiple copies of an antibody-fusion protein that binds the repeating peptide. The antibody-fusion protein may comprise an additional effector domain, (e.g. TET1, VP64), to induce increased transcription of the target gene. SunTag, including in dCas fusion proteins for gene activation, has been described, for example, in WO 2016/011070 and Tancnbaum, M. et al. *Cell*. 159(3):635–646 (2014). An exemplary SunTag effector domain includes a repeating GCN4 peptide having the amino acid sequence LLPKNYHLENEVARLKKLVGGER (SEQ ID NO:137) separated by linkers having the amino acid sequence GGSGG (SEQ ID NO:138). In some embodiments, the effector domain comprises the sequence set forth in SEQ ID NO:137, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing. In some embodiments, the SunTag effector domain recruits an antibody-fusion protein that comprises a TET protein (e.g. TET1) and binds the GCN4 peptide. In some embodiments, the SunTag effector domain recruits an antibody-fusion protein that comprises a transcriptional activator (e.g. VP64) and binds the GCN4 peptide.

#### **F. Fusion Protein**

**[0320]** Provided are fusion proteins that include (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein: the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. In some embodiments, the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation. In some embodiments, the effector domain induces transcription de-repression. In some embodiments, the fusion protein comprises any of the effector domains described herein.

**[0321]** In some aspects, the effector domain comprises any one of the effector domains

described herein.

**[0322]** In some embodiments, the fusion protein comprises a DNA-targeting domain or a protein component of the DNA-targeting domain, e.g., a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas); a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof, such as a catalytically inactive variant of any of the foregoing; and an effector domain, such as any of the effector domains described herein.

**[0323]** In some embodiments, binding of the DNA-targeting domain or a component thereof to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

**[0324]** In some embodiments, the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination that includes (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof, such as a catalytically inactive variant thereof. In some embodiments, the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

**[0325]** In some embodiments, the DNA-targeting domain comprises a Cas-gRNA combination that includes a Cas protein or a variant thereof (e.g., protein component) and at least one gRNA, and the component of the DNA-targeting domain is a Cas protein or a variant thereof. In some embodiments, the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein. In some embodiments, the gRNA is capable of complexing with the Cas protein or variant thereof. In some embodiments, the Cas protein or a variant thereof is a Cas9 protein or a variant thereof. In some embodiments, the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein or a nuclease-inactive Cas9 (iCas9) protein. In some aspects, the dCas9 or iCas9 component of the fusion protein includes any described herein, for example, in Section II.C.1.

**[0326]** In some embodiments, the Cas9 protein or a variant thereof is a *Staphylococcus aureus* Cas9 (SaCas9) protein or a variant thereof. In some embodiments, the variant Cas9 is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:179. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0327]** In some embodiments, the Cas9 protein or variant thereof is a *Streptococcus*

*pyogenes* Cas9 (SpCas9) protein or a variant thereof. In some embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0328]** In some embodiments, the Cas9 protein or a variant thereof is a *Streptococcus pyogenes* Cas9 (SaCas9) protein or a variant thereof. In some embodiments, the variant Cas9 is a *Streptococcus pyogenes* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99. In some embodiments, the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0329]** In some embodiments, the DNA-targeting domain of the fusion protein is a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof, such as a catalytically inactive variant thereof. In some aspects, the DNA-targeting domain of the fusion protein is targeted to one or more target sites at a MeCP2 locus, such as one or more target sites described herein, for example, in Section II.A. In some aspects, the DNA-targeting domain of the fusion protein is a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or a I-SceI enzymes or a variant thereof that is capable of binding to a target site at a MeCP2 locus described herein, in a sequence-specific manner.

**[0330]** In some embodiments, the DNA-binding domain or component thereof targets a target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158, such as any target site in the MeCP2 locus described herein.

**[0331]** In some embodiments, the fusion protein comprises the sequence set forth in SEQ ID NO:101, 103, 139-152, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the NLS comprises the sequence set forth in SEQ ID NO:101, 103, 139-152, or a portion thereof. In some embodiments, the NLS comprises the sequence set forth in SEQ ID NO:85 or a portion thereof.

**[0332]** In some embodiments, the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0333]** In some embodiments, the fusion protein further comprises one or more linkers

connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprises one or more nuclear localization signals (NLS).

**[0334]** In some embodiments, the fusion protein includes at least one linker. A linker may be included anywhere in the polypeptide sequence of the fusion protein, for example, between the effector domain and the DNA-targeting domain or a component thereof. A linker may be of any length and designed to promote or restrict the mobility of components in the fusion protein.

**[0335]** A linker may comprise any amino acid sequence of about 2 to about 100, about 5 to about 80, about 10 to about 60, or about 20 to about 50 amino acids. A linker may comprise an amino acid sequence of at least about 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 or 85 amino acids. A linker may comprise an amino acid sequence of less than about 100, 90, 80, 70, 60, 50, or 40 amino acids. A linker may include sequential or tandem repeats of an amino acid sequence that is 2 to 20 amino acids in length. Linkers may be rich in amino acids glycine (G), serine (S), and/or alanine (A). Linkers may include, for example, a GS linker such as (Gly-Gly-Gly-Gly-Ser)<sub>n</sub>. An exemplary GS linker is represented by the sequence GGGGS (SEQ ID NO:157). A linker may comprise repeats of a sequence, for example as represented by the formula (GGGGS)<sub>n</sub>, wherein n is an integer that represents the number of times the GGGGS sequence is repeated (e.g. between 1 and 10 times). The number of times a linker sequence is repeated, for example n in a GS linker, can be adjusted to optimize the linker length and achieve appropriate separation of the functional domains. Other examples of linkers may include, for example, Gly-Gly-Gly-Gly-Gly (SEQ ID NO:153), Gly-Gly-Ala-Gly-Gly (SEQ ID NO:154), Gly/Ser rich linkers such as Gly-Gly-Gly-Gly-Ser-Ser-Ser (SEQ ID NO:155), or Gly/Ala rich linkers such as Gly-Gly-Gly-Gly-Ala-Ala-Ala (SEQ ID NO:156) or Gly-Ser-Gly-Ser-Gly (SEQ ID NO:206).

**[0336]** In some embodiments, the linker is an XTEN linker. In some aspects, an XTEN linker is a recombinant polypeptide (e.g., an unstructured recombinant peptide) lacking hydrophobic amino acid residues. Exemplary XTEN linkers are described in, for example, Schellenberger et al., *Nature Biotechnology* 27, 1186-1178 (2009) or WO 2021/247570. In some embodiments, an exemplary linker comprises a linker described in WO 2021/247570. In some aspects, the linker is or comprises the sequence set forth in SEQ ID NO:117 or SEQ ID NO:178, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing. In some embodiments, the linker comprises the sequence set forth in SEQ ID NO:117, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing. In some aspects, the linker comprises the sequence set

forth in SEQ ID NO:117, or a contiguous portion of SEQ ID NO:117 of at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 or 75 amino acids. In some aspects, the linker consists of the sequence set forth in SEQ ID NO:117, or a contiguous portion of SEQ ID NO:117 of at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 or 75 amino acids. In some embodiments, the linker comprises the sequence set forth in SEQ ID NO:117. In some embodiments, the linker consist of the sequence set forth in SEQ ID NO:117. In some embodiments, the linker comprises the sequence set forth in SEQ ID NO:178, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing. In some aspects, the linker comprises the sequence set forth in SEQ ID NO:178, or a contiguous portion of SEQ ID NO:178 of at least 5, 10, or 15 amino acids. In some aspects, the linker consists of the sequence set forth in SEQ ID NO:178, or a contiguous portion of SEQ ID NO:178 of at least 5, 10, or 15 amino acids. In some embodiments, the linker comprises the sequence set forth in SEQ ID NO:178. In some embodiments, the linker consist of the sequence set forth in SEQ ID NO:178. Appropriate linkers may be selected or designed based rational criteria known in the art, for example as described in Chen et al. *Adv. Drug Deliv. Rev.* 65(10):1357-1369 (2013). In some embodiments, a linker comprises the sequence set forth in SEQ ID NO:119, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

**[0337]** In some embodiments, a fusion protein described herein comprises one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the sequence PKKKRKV (SEQ ID NO:103); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS) having the sequence KRPAATKKAGQAKKKK (SEQ ID NO:105); the c-myc NLS having the sequence PAAKRVKLD (SEQ ID NO:139) or RQRRNELKRSP (SEQ ID NO:140); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:141); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:142) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO:143) and PPKKARED (SEQ ID NO:144) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO:145) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO:146) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO:147) and PKQKKRK (SEQ ID NO:148) of the

influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO:149) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO:150) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKSKK (SEQ ID NO:151) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO:152) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the fusion protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the fusion protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the fusion protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of the fusion protein (e.g. an assay for altered gene expression activity in a cell transformed with the DNA-targeting system comprising the fusion protein), as compared to a control condition (e.g. an untransformed cell).

**[0338]** In some embodiments, the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0339]** In some embodiments, the fusion protein comprises the sequence set forth in SEQ ID NO:115, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

### **G. Split Fusion Proteins**

**[0340]** In some embodiments, the fusion protein is a split protein, i.e. comprises two or more separate polypeptide domains that interact or self-assemble to form a functional fusion protein. In some aspects, the split fusion protein comprises a dCas9 and an effector domain. In some aspects, the fusion protein comprises a split dCas9-TET1 fusion protein.

**[0341]** In some embodiments, the split fusion protein is assembled from separate polypeptide domains comprising trans-splicing inteins. Inteins are internal protein elements that self-excite from their host protein and catalyze ligation of flanking sequences with a peptide bond. In some embodiments, the split fusion protein is assembled from a first polypeptide

comprising an N-terminal intein and a second polypeptide comprising a C-terminal intein. In some embodiments, the N terminal intein is the N terminal Npu Intein set forth in SEQ ID NO:129. In some embodiments, the C terminal intein is the C terminal Npu intein set forth in SEQ ID NO:133.

**[0342]** Also provided are fusion proteins comprising a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. Also provided are fusion proteins comprising a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus. In some aspects, the first polypeptide of the split variant Cas protein, and a second polypeptide of the split variant Cas protein comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

**[0343]** Also provided are fusion proteins comprising a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation. Also provided are fusion proteins comprising a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus. In some aspects, the second polypeptide of the split variant Cas protein, and a first polypeptide of the split variant Cas protein comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

**[0344]** In some embodiments, the split fusion protein comprises a split dCas9-TET1 fusion protein assembled from two polypeptides. In an exemplary embodiment, the first polypeptide

comprises a TET1 catalytic domain and an N-terminal fragment of dSpCas9, followed by an N terminal Npu Intein (TET1-dSpCas9-573N; set forth in SEQ ID NO:121), and the second polypeptide comprises a C terminal Npu Intein, followed by a C-terminal fragment of dSpCas9 (dSpCas9-573C; set forth in SEQ ID NO:131). The N- and C-terminal fragments of the fusion protein are split at position 573Glu of the dSpCas9 molecule, with reference to SEQ ID NO:96. In some aspects, the N-terminal Npu Intein (SEQ ID NO:129) and C-terminal Npu Intein (set forth in SEQ ID NO:133) may self-excise and ligate the two fragments, thereby forming the full-length dSpCas9-TET1 fusion protein when expressed in a cell.

**[0345]** In some embodiments, the polypeptides of a split protein may interact non-covalently to form a complex that recapitulates the activity of the non-split protein. For example, two domains of a Cas enzyme expressed as separate polypeptides may be recruited by a gRNA to form a ternary complex that recapitulates the activity of the full-length Cas enzyme in complex with the gRNA, for example as described in Wright et al. PNAS 112(10):2984-2989 (2015). In some embodiments, assembly of the split protein is inducible (e.g. light inducible, chemically inducible, small-molecule inducible).

**[0346]** In some aspects, the two polypeptides of a split fusion protein may be delivered and/or expressed from separate vectors, such as any of the vectors described herein. In some embodiments, the two polypeptides of a split fusion protein may be delivered to a cell and/or expressed from two separate AAV vectors, i.e. using a split AAV-based approach, for example as described in WO 2017/197238.

**[0347]** Approaches for the rationale design of split proteins and their delivery, including Cas proteins and fusions thereof, are described, for example, in WO 2016/114972, WO 2017/197238, Zetsche. et al. Nat. Biotechnol. 33(2):139-42 (2015), Wright et al. PNAS 112(10):2984-2989 (2015), Truong. et al. Nucleic Acids Res. 43, 6450–6458 (2015), and Fine et al. Sci. Rep. 5, 10777 (2015).

## **H. Exemplary Fusion Proteins**

**[0348]** In some aspects, provided are DNA-targeting systems or fusion proteins that comprise a Cas protein or a variant thereof and at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

**[0349]** In some embodiments, the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof (such as a protein or polypeptide component thereof, for example, a Cas component of a Cas-gRNA combination). In some embodiments, the DNA-targeting system also



includes one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

**[0350]** In some aspects, the DNA-targeting system or fusion protein comprises one or more tags, linkers and/or NLS sequences. In some embodiments, exemplary tags, linkers and/or NLS sequences can be any described herein.

**[0351]** In some cases, sequences provided herein, including amino acid sequences for the DNA-targeting systems or fusion proteins provided herein, contain sequences of one or more tags, linkers and/or NLS sequences. In some aspects, it is understood that the exemplary tags, linkers and/or NLS sequences are not required or are not the sole or exclusive tags, linkers and/or NLS sequences that can be employed in the DNA-targeting systems or fusion proteins. In some aspects, sequences containing tags, linkers and/or NLS sequences are exemplary, and are not limited to the specific tags, linkers and/or NLS sequences contained in the described sequences. In some aspects, alternative tags, linkers and/or NLS sequences can be employed in the DNA-targeting systems or fusion proteins, or the DNA-targeting system or fusion protein in some cases does not contain or lacks a tag, linker and/or NLS. In some aspects, alternative tags, linkers and/or NLS sequences include other known tags, linkers and/or NLS sequences that have similar function or serve similar purposes.

**[0352]** In some embodiments, the DNA-targeting system or the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the DNA-targeting system or the fusion protein comprises the sequence set forth in SEQ ID NO:115, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

#### **I. Combinations of Fusion Proteins and/or DNA-targeting Systems**

**[0353]** Also provided are combinations, such as combinations of two or more DNA-targeting systems or components thereof. In some aspects, provided herein are combinations of two or more DNA-targeting systems that independently target different target sites at a MeCP2 locus. In some aspects, the two or more DNA-targeting systems each comprise any of the DNA-targeting systems described herein.

**[0354]** In some embodiments, the DNA-targeting domain is a first DNA-targeting domain, and the DNA-targeting system further comprises one or more second DNA-targeting domain.

**[0355]** In some embodiments, the first DNA-targeting domain binds a first target site in a

MeCP2 locus; and the second DNA-targeting domain binds a second target site in a MeCP2 locus.

**[0356]** Also provided herein are DNA-targeting systems that binds to one or more target sites in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, the DNA-targeting system comprising: a first DNA-targeting domain that binds a first target site in a MeCP2 locus; and a second DNA-targeting domain that binds a second target site in a MeCP2 locus.

**[0357]** Also provided are combinations, such as combinations of two or more DNA-targeting domains or fusion proteins or components thereof. In some aspects, provided herein are combinations of two or more DNA-targeting domains or fusion proteins that independently target different target sites at a MeCP2 locus. In some aspects, the two or more DNA-targeting domains or fusion proteins each comprise any of the DNA-targeting domains or fusion proteins described herein.

**[0358]** In some embodiments, the DNA-targeting domain is a first DNA-targeting domain, and the DNA-targeting domain or fusion protein further comprises one or more second DNA-targeting domains. In some embodiments, the first DNA-targeting domain binds a first target site in the MECP2 locus, and the second DNA-targeting domain binds a second target site in the MECP2 locus.

**[0359]** In some aspects, the provided combination of DNA-targeting domains or fusion proteins include two or more DNA-targeting domains or fusion proteins, each of which target particular regions of a MeCP2 locus.

**[0360]** Also provided herein is a combination, comprising a first DNA-targeting domain or fusion protein comprising any of the DNA-targeting domains or fusion proteins described herein, and one or more second DNA-targeting domains or fusion proteins that binds to a second target site in a regulatory DNA element of a MeCP2 locus. In some embodiments, the second DNA-targeting domain or fusion protein comprises any of the DNA-targeting domains or fusion proteins described herein.

**[0361]** In some embodiments, the first target site is any described herein, such as in Section II.A. In some embodiments, the second target site is any described herein, such as in Section II.A. In some embodiments, the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158. In some embodiments, the second target site is located within the genomic coordinates hg38 chrX:154,097,151-154,098,158. In some embodiments, the first target site and the second target site independently are located within the genomic coordinates hg38 chrX:154,097,151-154,098,158. In some

embodiments, the first target site and the second target site are different.

**[0362]** In some embodiments, the first DNA-targeting domain comprises a first Cas-gRNA combination that includes (a) a first Cas protein or a variant thereof and (b) a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site; and the second DNA-targeting domain comprises a second Cas-gRNA combination that includes (a) a second Cas protein or a variant thereof and (b) a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site.

**[0363]** In some embodiments, the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein.

**[0364]** In some embodiments, the first variant Cas protein and/or the second variant Cas protein is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96; or comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0365]** In some embodiments, the first variant Cas protein and/or the second variant Cas protein is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; or comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

**[0366]** In some embodiments, the first Cas protein and the second Cas protein are the same. In some embodiments, the first Cas protein and the second Cas protein are different.

**[0367]** In some embodiments, the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is fused to at least one effector domain.

**[0368]** In some embodiments, the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation. In some embodiments, the effector domain induces transcription activation.

**[0369]** In some aspects, exemplary combination of DNA-targeting systems include: (a) a fusion protein comprising a Cas protein or a variant thereof and (b) a combination of gRNAs, such as a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site and a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site. In some aspects, also provided herein are combinations of DNA-targeting systems comprising one type of Cas protein or variant thereof, such as a dCas9 protein or variant thereof, and two or more different gRNAs, such as a combination of gRNAs, such as any combination of gRNAs described herein. In some aspects, also provided herein are combinations of DNA-targeting systems comprising one type of Cas protein or variant thereof, such as a dCas9 protein or variant thereof, two or more different types of effector domains, and two or more different gRNAs, such as a combination of gRNAs, such as any combination of gRNAs described herein. In some aspects, also provided herein are combinations of DNA-targeting systems comprising two or more different type of Cas protein or variant thereof, such as a dCas9 protein or variant thereof, and two or more different gRNAs, such as a combination of gRNAs, such as any combination of gRNAs described herein. In some aspects, also provided herein are combinations of DNA-targeting systems comprising two or more different types of DNA-targeting domains and one type of effector domain. In some aspects, also provided herein are combinations of DNA-targeting systems comprising two or more different types of DNA-targeting domains and two or more different types of effector domain.

**[0370]** In some embodiments, the first DNA-targeting domain comprises a first Cas-gRNA combination comprising (a) a first Cas protein or a variant thereof and (b) a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site; and the second DNA-targeting domain comprises a second Cas-gRNA combination comprising (a) a second Cas protein or a variant thereof and (b) a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site. In some embodiments, the first DNA-targeting domain comprises a first Cas-gRNA combination comprising (a) a first Cas protein or a variant thereof and (b) a first gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:22 or a contiguous portion thereof of at least 14 nt. In some embodiments, the second DNA-targeting domain comprises a second Cas-gRNA combination comprising (a) a second Cas protein or a variant thereof and (b) a second gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:28 or a contiguous portion thereof of at least 14 nt.

**[0371]** In some embodiments, the first Cas-gRNA combination comprises (a) a first Cas

protein or a variant thereof and (b) a first gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:9 or a contiguous portion thereof of at least 14 nt; and the second Cas-gRNA combination comprises (a) a second Cas protein or a variant thereof and (b) a second gRNA comprising at least one gRNA spacer sequence set forth in SEQ ID NO:27 or a contiguous portion thereof of at least 14 nt.

**[0372]** In some embodiments, all of the components of the combination of DNA-targeting systems, DNA-targeting domains or fusion proteins provided herein are encoded in one polynucleotide. In some embodiments, all of the components of the combination of DNA-targeting systems, DNA-targeting domains or fusion proteins provided herein are encoded in multiple individual polynucleotides, such as a first polynucleotide and a second polynucleotide. In some aspects, first DNA-targeting system, DNA-targeting domain or fusion protein and the second DNA-targeting system, DNA-targeting domain or fusion protein are encoded in one polynucleotide, such as a first polynucleotide. In some embodiments, the first DNA-targeting system, domain or fusion protein and the second DNA-targeting system, domain or fusion protein are encoded in one polynucleotide, such as a first polynucleotide. In some embodiments, the first Cas protein and the second Cas protein are encoded in a first polynucleotide. In some embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence. In some embodiments, the first gRNA and the second gRNA are encoded in a first polynucleotide. In some embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence, and the Cas protein, the first gRNA, and the second gRNA are encoded in a first polynucleotide. In some embodiments, the first DNA-targeting domain is encoded in a first polynucleotide and the second DNA-targeting domain is encoded in a second polynucleotide. In some embodiments, the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide. In some embodiments, the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide. In some embodiments, the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

### **III. POLYNUCLEOTIDES, VECTORS AND DELIVERY OF DNA-TARGETING SYSTEMS**

**[0373]** Provided are polynucleotides encoding any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing. In

some of any embodiments, provided are polynucleotides encoding any of the fusion proteins described herein. Also provided herein are polynucleotides encoding any of the gRNAs or combinations of gRNAs described herein.

**[0374]** The polynucleotides can encode any of the components of the DNA-targeting systems, and/or any nucleic acid or proteinaceous molecule necessary to carry out the aspects of the methods of the disclosure can comprise a vector (e.g., a recombinant expression vector).

#### **A. Nucleic Acids**

**[0375]** Provided are polynucleotides encoding any of the DNA-targeting systems described herein, including a protein component of the DNA-targeting system (e.g., Cas protein or a variant thereof) and the at least one gRNA, such as one or more RNAs.

**[0376]** In some embodiments, provided are polynucleotides comprising the gRNAs described herein. In some embodiments, the gRNA is transcribed from a genetic construct (i.e. vector or plasmid) in the target cell. In some embodiments, the gRNA is produced by in vitro transcription and delivered to the target cell. In some embodiments, the gRNA comprises one or more modified nucleotides for increased stability. In some embodiments, the gRNA is delivered to the target cell pre-complexed as a RNP with the fusion protein.

**[0377]** In some embodiments, a provided polynucleotide encodes a fusion protein as described herein that includes (a) a DNA-targeting domain capable of being targeted to a target site of a target gene as described; and (b) at least one effector domain capable of reducing transcription of the gene. In some embodiments, the fusion protein includes a fusion protein of a Cas protein or variant thereof and at least one effector domain capable of reducing transcription of a gene. In a particular example, the Cas is a deactivated Cas (dCas), such as dCas9. In some embodiments, the dCas9 is a dSpCas9. Examples of such domains and fusion proteins include any as described in Section I.

**[0378]** In some embodiments, the polynucleotide, such as a polynucleotide encoding any of the components of the DNA targeting system, fusion protein and/or gRNA, is DNA. In some embodiments, the polynucleotide, such as a polynucleotide encoding any of the components of the DNA targeting system, fusion protein and/or gRNA, is RNA. In some embodiments, the polynucleotide is mRNA. In some embodiments, the gRNA is provided as RNA and a polynucleotide encoding the fusion protein is mRNA. In some aspects, the mRNA is 5' capped and/or 3' polyadenylated. In some embodiments, a polynucleotide provided herein is DNA. In some aspects, the DNA is present in a vector.

**[0379]** In some embodiments, the polynucleotide encodes the fusion protein and one or more

gRNAs or a combination of gRNAs.

**[0380]** In some embodiments, the polynucleotide as provided herein can be codon optimized for efficient translation into protein in the eukaryotic cell or animal of interest. For example, codons can be optimized for expression in humans, mice, rats, hamsters, cows, pigs, cats, dogs, fish, amphibians, plants, yeast, insects, and others.

**[0381]** In some embodiments, the polynucleotide comprises the sequence set forth in SEQ ID NO:90, or a sequence having at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the polynucleotide comprises the sequence set forth in SEQ ID NO:90.

**[0382]** Also provided are polynucleotides encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of any of the DNA-targeting systems described herein or any of the combinations described herein.

**[0383]** Also provided herein are pluralities of polynucleotides, comprising: (a) a polynucleotide encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of any of the embodiments disclosed herein or any of the combinations of gRNAs disclosed herein, and (b) a polynucleotide encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of any of the embodiments disclosed herein or any of the combinations of gRNAs disclosed herein.

**[0384]** Provided are polynucleotides encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of any of the DNA-targeting systems described herein or any of the combinations described herein.

**[0385]** Provided are polynucleotides that include any of the polynucleotides described herein, and one or more additional polynucleotides encoding an additional portion or an additional component of any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing.

**[0386]** Provided are pluralities of polynucleotides, that includes a first polynucleotide comprising any of the polynucleotides described herein; and a second polynucleotide comprising any of the polynucleotides described herein.

**[0387]** In some embodiments, the first DNA-targeting domain and the second DNA-targeting domain are encoded in a first polynucleotide. In some embodiments, the first Cas protein and the second Cas protein are encoded in a first polynucleotide. In some embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence. In some embodiments, the first gRNA and the second gRNA are encoded in a first polynucleotide.

In some embodiments, the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence, and the Cas protein, the first gRNA, and the second gRNA are encoded in a first polynucleotide.

**[0388]** In some embodiments, the first DNA-targeting domain is encoded in a first polynucleotide and the second DNA-targeting domain is encoded in a second polynucleotide. In some embodiments, the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide. In some embodiments, the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide. In some embodiments, the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

## **B. Vectors**

**[0389]** Provided are vectors that include any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, or a first polynucleotide or a second polynucleotide of any of the pluralities of polynucleotides described herein, or a portion or a component of any of the foregoing. Also provided herein is a vector that comprises or contains any of the provided polynucleotides. In some embodiments, the vector comprises a genetic construct, such as a plasmid or an expression vector. The vector can be a self-inactivating vector that either inactivates the viral sequences or the components of the CRISPR machinery or other elements.

**[0390]** In some embodiments, the expression vector comprising the sequence encoding the fusion protein of a DNA-targeting system provided herein further comprises a nucleic acid sequence encoding at least one gRNA. In some embodiments, the expression vector comprises a nucleic acid sequence or combination of nucleic acid sequences encoding two or more gRNAs, such as two gRNAs. In some embodiments, the expression vector comprises a nucleic acid sequence or combination of nucleic acid sequences encoding three gRNAs. In some cases, the sequence encoding the gRNA is operably linked to at least one transcriptional control sequence or transcriptional regulatory sequence (e.g., cis-regulatory sequence) for expression of the gRNA in the cell. In some aspects, DNA encoding the gRNA can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters, or variants thereof. In some aspects, if the expression vector comprises nucleic acid sequences encoding two or more gRNAs, each gRNA is operably linked to an identical Pol III promoter, or different Pol III promoters.



**[0391]** In some embodiments, provided is a vector containing a polynucleotide that encodes a fusion protein comprising a DNA-targeting domain comprising a dCas and at least one effector domain capable of increasing transcription of a gene, and a polynucleotide or combination of polynucleotides encoding a gRNA, or a plurality of gRNAs, such as two, three, or four or more gRNAs, or such as two, three, or four or more different gRNAs. In some embodiments, the dCas is a dCas9, such as dSaCas9 or dSpCas9. In some embodiments, the polynucleotide encodes a fusion protein that includes a dSaCas9 set forth in SEQ ID NO:72. In some embodiments, the polynucleotide encodes a fusion protein that includes a dSpCas9 set forth in SEQ ID NO:78. In some embodiments, the polynucleotide(s) encodes one or more a gRNAs described herein, for example in or a plurality of gRNAs, each gRNA as described in Section II.B.

**[0392]** In some examples, a polynucleotide and/or a vector described herein can comprise one or more transcription and/or translation control elements. Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. can be used in the expression vector. The vector can be a self-inactivating vector that either inactivates the viral sequences or the components of the CRISPR machinery or other elements.

**[0393]** Non-limiting examples of suitable eukaryotic promoters (i.e., promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, human elongation factor-1 promoter (EF1), a hybrid construct comprising the cytomegalovirus (CMV) enhancer fused to the chicken beta-actin promoter (CAG), murine stem cell virus promoter (MSCV), phosphoglycerate kinase-1 locus promoter (PGK), and mouse metallothionein-1.

**[0394]** For expressing small RNAs, including guide RNAs used in connection with the DNA-targeting systems, various promoters such as RNA polymerase III promoters, including for example U6 and H1, can be advantageous. Descriptions of and parameters for enhancing the use of such promoters are known in art, and additional information and approaches are regularly being described; see, e.g., Ma, H. et al., *Molecular Therapy—Nucleic Acids* 3, e161 (2014) doi:10.1038/mtna.2014.12.

**[0395]** The expression vector can also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector can also comprise appropriate sequences for amplifying expression. The expression vector can also include nucleotide

sequences encoding non-native tags (e.g., histidine tag, hemagglutinin tag, green fluorescent protein, etc.) that are fused to the site-directed polypeptide, thus resulting in a fusion protein.

**[0396]** A promoter can be an inducible promoter (e.g., a heat shock promoter, tetracycline-regulated promoter, steroid-regulated promoter, metal-regulated promoter, estrogen receptor-regulated promoter, etc.). The promoter can be a constitutive promoter (e.g., CMV promoter, UBC promoter). In some cases, the promoter can be a spatially restricted and/or temporally restricted promoter (e.g., a tissue specific promoter, a cell type specific promoter (e.g. nervous system specific promoter), etc.).

**[0397]** In some examples, vectors can be capable of directing the expression of nucleic acids to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors”, or more simply “expression vectors”, which serve equivalent functions.

**[0398]** Exemplary expression vectors contemplated include, but are not limited to, viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, human immunodeficiency virus, retrovirus (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus) and other recombinant vectors. Other vectors contemplated for eukaryotic target cells include, but are not limited to, the vectors pXT1, pSG5, pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). Other vectors can be used so long as they are compatible with the host cell.

**[0399]** In some embodiments, the vector is a viral vector, such as an adeno-associated virus (AAV) vector, a retroviral vector, a lentiviral vector, or a gammaretroviral vector. In some embodiments, the viral vector is an adeno-associated virus (AAV) vector. In some embodiments, the AAV vector is selected from among an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 vector. In some embodiments, the vector is a lentiviral vector. In some embodiments, the vector is a non-viral vector, for example a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide.

**[0400]** In some embodiments, the vector comprises one vector, or two or more vectors.

**[0401]** In some aspects, provided herein are pluralities of vectors that comprise any of the vectors described herein, and one or more additional vectors comprising one or more additional polynucleotides encoding an additional portion or an additional component of any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, or any of the fusion proteins described herein, or a portion or a component of any of the foregoing.

**[0402]** Provided are pluralities of vectors, that include: a first vector comprising any of the polynucleotides described herein; and a second vector comprising any of the polynucleotides described herein. Also provided herein are pluralities of vectors, comprising: a first vector comprising a polynucleotide encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of any of the embodiments of a DNA-targeting system described herein or any of the combinations of gRNAs described herein; and; a second vector comprising a polynucleotide encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of any of the embodiments of a DNA-targeting system described herein or any of the combinations of gRNAs described herein.

**[0403]** In some embodiments, polynucleotides can be cloned into a suitable vector, such as an expression vector or vectors. The expression vector can be any suitable recombinant expression vector, and can be used to transform or transfect any suitable cell. Suitable vectors include those designed for propagation and expansion or for expression or both, such as plasmids and viruses.

**[0404]** In some embodiments, the vector can be a vector of the pUC series (Fermentas Life Sciences), the pBluescript series (Stratagene, LaJolla, Calif.), the pET series (Novagen, Madison, Wis.), the pGEX series (Pharmacia Biotech, Uppsala, Sweden), or the pEX series (Clontech, Palo Alto, Calif.). In some embodiments, animal expression vectors include pEUK-CI, pMAM and pMAMneo (Clontech). In some embodiments, a viral vector is used, such as a lentiviral or retroviral vector. In some embodiments, the recombinant expression vectors can be prepared using standard recombinant DNA techniques. In some embodiments, vectors can contain regulatory sequences, such as transcription and translation initiation and termination codons, which are specific to the type of host into which the vector is to be introduced, as appropriate and taking into consideration whether the vector is DNA- or RNA- based. In some embodiments, the vector can contain a nonnative promoter operably linked to the nucleotide sequence encoding the recombinant receptor. In some embodiments, the promoter can be a non-viral promoter or a viral promoter, such as a cytomegalovirus (CMV) promoter, an SV40 promoter, an RSV promoter, and a promoter found in the long-terminal repeat of the murine stem cell virus. Other promoters known to a skilled artisan also are contemplated.

**[0405]** In some embodiments, recombinant nucleic acids are transferred into cells using recombinant infectious virus particles, such as, e.g., vectors derived from simian virus 40 (SV40), adenoviruses, or adeno-associated virus (AAV). In some embodiments, recombinant nucleic acids are transferred into cells (e.g. central nervous system cells, such as neurons) using recombinant lentiviral vectors or retroviral vectors, such as gamma-retroviral vectors (see, e.g.,

Koste et al. (2014) *Gene Therapy* 2014 Apr 3. doi: 10.1038/gt.2014.25; Carlens et al. (2000) *Exp Hematol* 28(10): 1137-46; Alonso-Camino et al. (2013) *Mol Ther Nucl Acids* 2, e93; Park et al., *Trends Biotechnol.* 2011 November 29(11): 550-557.

**[0406]** In some embodiments, the retroviral vector has a long terminal repeat sequence (LTR), e.g., a retroviral vector derived from the Moloney murine leukemia virus (MoMLV), myeloproliferative sarcoma virus (MPSV), murine embryonic stem cell virus (MESV), murine stem cell virus (MSCV), spleen focus forming virus (SFFV), or adeno-associated virus (AAV). Most retroviral vectors are derived from murine retroviruses. In some embodiments, the retroviruses include those derived from any avian or mammalian cell source. The retroviruses typically are amphotropic, meaning that they are capable of infecting host cells of several species, including humans. In one embodiment, the gene to be expressed replaces the retroviral gag, pol and/or env sequences. A number of illustrative retroviral systems have been described (e.g., U.S. Pat. Nos. 5,219,740; 6,207,453; 5,219,740; Miller and Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa et al. (1991) *Virology* 180:849-852; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; and Boris-Lawrie and Temin (1993) *Cur. Opin. Genet. Develop.* 3: 102-109.

**[0407]** In some embodiments, the vector is a lentiviral vector. In some embodiments, the lentiviral vector is an integrase-deficient lentiviral vector. In some embodiments, the lentiviral vector is a recombinant lentiviral vector. In some embodiments, the lentivirus is selected or engineered for a desired tropism (e.g. for central nervous system tropism, or tropism for a heart cell, such as a cardiomyocyte, a skeletal muscle cell, a nervous system cell, such as a neuron, a fibroblast, or an induced pluripotent stem cell). In some embodiments, the cell for any of the provided compositions, such as DNA-targeting systems, fusion proteins, gRNAs, polynucleotides and/or vectors to be delivered is a heart cell, a skeletal muscle cell, a nervous system cell, or an induced pluripotent stem cell. Methods of lentiviral production, transduction, and engineering are known, for example as described in Kasaraneni, N. et al. *Sci. Rep.* 8(1):10990 (2018), Ghaleh, H.E.G. et al. *Biomed. Pharmacother.* 128:110276 (2020), and Milone, M.C. et al. *Leukemia.* 32(7):1529-1541 (2018). Additional methods for lentiviral transduction are described, for example in Wang et al. (2012) *J. Immunother.* 35(9): 689-701; Cooper et al. (2003) *Blood.* 101: 1637- 1644; Verhoeyen et al. (2009) *Methods Mol Biol.* 506: 97-114; and Cavalieri et al. (2003) *Blood.* 102(2): 497-505.

**[0408]** In some embodiments, recombinant nucleic acids are transferred into cells (e.g. central nervous system cells, such as neurons, or a heart cell, a skeletal muscle cell, a nervous system cell, or an induced pluripotent stem cell) via electroporation (see, e.g., Chicaybam et al,

(2013) PLoS ONE 8(3): e60298 and Van Tedeloo et al. (2000) Gene Therapy 7(16): 1431-1437). In some embodiments, recombinant nucleic acids are transferred into cells via transposition (see, e.g., Manuri et al. (2010) Hum Gene Ther 21(4): 427-437; Sharma et al. (2013) Molec Ther Nucl Acids 2, e74; and Huang et al. (2009) Methods Mol Biol 506: 115-126). Other methods of introducing and expressing genetic material into immune cells include calcium phosphate transfection (e.g., as described in Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.), protoplast fusion, cationic liposome-mediated transfection; tungsten particle-facilitated microparticle bombardment (Johnston, Nature, 346: 776-777 (1990)); and strontium phosphate DNA co-precipitation (Brash et al., Mol. Cell Biol., 7: 2031-2034 (1987)).

### **I. AAV vectors**

**[0409]** In some embodiments, the viral vector is an AAV vector. In some embodiments, the AAV vector is selected from among an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or an AAV-DJ vector. In some embodiments, the AAV vector is an AAV vector engineered for central nervous system (CNS) tropism. In some embodiments, the AAV vector is selected from among an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 vector. In some embodiments, the AAV vector is an AAV5 vector or an AAV9 vector. In some aspects, the AAV vector is an AAV9 vector. In some aspects, the AAV vector is an AAV5 vector. In some aspects, the AAV vector is an AAV-DJ vector.

**[0410]** In some embodiments, the AAV is selected or engineered for a desired tropism (e.g. for central nervous system tropism, or tropism for a heart cell, such as a cardiomyocyte, a skeletal muscle cell, a nervous system cell, such as a neuron, a fibroblast, or an induced pluripotent stem cell (iPSC)). In some embodiments, the AAV is exhibits tropism for a cardiomyocyte. In some embodiments, the AAV is exhibits tropism for a nervous system cell. In some embodiments, the AAV is exhibits tropism for a cell of the central nervous system (CNS). In some embodiments, the AAV is exhibits tropism for a neuron. In some embodiments, the AAV is exhibits tropism for a fibroblast. In some embodiments, the AAV is exhibits tropism for an iPSC.

**[0411]** In some aspects, nucleic acids or polynucleotides encoding any of the DNA-targeting systems, guide RNAs, fusion proteins, or components, portions or combinations thereof can be delivered to cells or subjects using gene delivery vectors, such as viral vectors. In some aspects, provided herein are viral vectors that comprise any of the nucleic acids or polynucleotides

described herein, any of the pluralities of nucleic acids or polynucleotides described herein, or a first polynucleotide or a second polynucleotide of any of the pluralities of polynucleotides described herein, or a portion or a component of any of the foregoing.

**[0412]** Examples of virions that can be employed to deliver any of the nucleic acids or polynucleotides provided herein include but are not limited to retroviral virions, lentiviral virions, adenovirus virions, herpes virus virions, alphavirus virions, and adeno-associated virus (AAV) virions. AAV is a 4.7 kb, single-stranded DNA virus. Recombinant virions based on AAV (rAAV virions) are associated with excellent clinical safety, since wild-type AAV is nonpathogenic and has no etiologic association with any known diseases. In addition, AAV offers the capability for highly efficient delivery and sustained expression of the delivered nucleic acid, composition or component thereof, in numerous tissues, including the nervous system, eye, muscle, lung and brain.

**[0413]** A “recombinant AAV vector (recombinant adeno-associated viral vector)” in some aspects refers to a polynucleotide vector comprising one or more heterologous sequences (i.e., nucleic acid sequence not of AAV origin) that are flanked by at least one AAV inverted terminal repeat sequences (ITR). In some aspects, the recombinant nucleic acid is flanked by two inverted terminal repeat sequences (ITRs). Such recombinant viral vectors can be replicated and packaged into infectious viral particles when present in a host cell that has been infected with a suitable helper virus (or that is expressing suitable helper functions) and that is expressing AAV rep and cap gene products (i.e., AAV Rep and Cap proteins). When a recombinant viral vector is incorporated into a larger polynucleotide (e.g., in a chromosome or in another vector such as a plasmid used for cloning or transfection), then the recombinant viral vector may be referred to as a “pro-vector” which can be “rescued” by replication and encapsidation in the presence of AAV packaging functions and suitable helper functions. A recombinant viral vector can be in any of a number of forms, including, but not limited to, plasmids, linear artificial chromosomes, complexed with lipids, encapsulated within liposomes, and encapsidated in a viral particle, for example, an AAV particle. A recombinant viral vector can be packaged into an AAV virus capsid to generate a “recombinant adeno-associated viral particle (recombinant viral particle)”.

**[0414]** An “rAAV virus” or “rAAV viral particle” refers to a viral particle composed of at least one AAV capsid protein and an encapsidated rAAV vector genome.

**[0415]** “AAV helper functions” refer to functions that allow AAV to be replicated and packaged by a host cell for producing viruses. AAV helper functions can be provided in any of a number of forms, including, but not limited to, helper virus or helper virus genes which aid in AAV replication and packaging. Other AAV helper functions are known, such as genotoxic

agents.

**[0416]** A “helper virus” for AAV refers to a virus that allows AAV (which is a defective parvovirus) to be replicated and packaged by a host cell for producing viruses. A helper virus provides “helper functions” which allow for the replication of AAV. A number of such helper viruses have been identified, including adenoviruses, herpesviruses, poxviruses such as vaccinia and baculovirus. The adenoviruses encompass a number of different subgroups, although Adenovirus type 5 of subgroup C (Ad5) is most commonly used. Numerous adenoviruses of human, non-human mammalian and avian origin are known and are available from depositories such as the ATCC. Viruses of the herpes family, which are also available from depositories such as ATCC, include, for example, herpes simplex viruses (HSV), Epstein-Barr viruses (EBV), cytomegaloviruses (CMV) and pseudorabies viruses (PRV). Examples of adenovirus helper functions for the replication of AAV include E1A functions, E1B functions, E2A functions, VA functions and E4orf6 functions. Baculoviruses available from depositories include *Autographa californica* nuclear polyhedrosis virus.

**[0417]** A preparation of rAAV is said to be “substantially free” of helper virus if the ratio of infectious AAV particles to infectious helper virus particles is at least about 102:1; at least about 104:1, at least about 106:1; or at least about 108:1 or more. In some aspects, preparations are also free of equivalent amounts of helper virus proteins (i.e., proteins as would be present as a result of such a level of helper virus if the helper virus particle impurities noted above were present in disrupted form). Viral and/or cellular protein contamination can generally be observed as the presence of Coomassie staining bands on SDS gels (e.g., the appearance of bands other than those corresponding to the AAV capsid proteins VP1, VP2 and VP3).

**[0418]** In some aspects, the recombinant viral particles for delivery of any of the provided nucleic acids, compositions or components thereof comprise a self-complementary AAV (scAAV) genome. In some aspects, the recombinant AAV genome comprises a first heterologous polynucleotide sequence (e.g., coding strand) and a second heterologous polynucleotide sequence (e.g., the noncoding or antisense strand) wherein the first heterologous polynucleotide sequence can form intrastrand base pairs with the second polynucleotide sequence along most or all of its length. In some aspects, the first heterologous polynucleotide sequence and a second heterologous polynucleotide sequence are linked by a sequence that facilitates intrastrand base-pairing; e.g., a hairpin DNA structure. Hairpin structures are known, for example in siRNA molecules. In some aspects, the first heterologous polynucleotide sequence and a second heterologous polynucleotide sequence are linked by a mutated ITR. In some aspects, the scAAV viral particles comprise a monomeric form of an scAAV genome. In

some aspects, the scAAV viral particles comprise the dimeric form of and scAAV genome. In some aspects, AUC as described herein is used to detect the presence of rAAV particles comprising the monomeric form of an scAAV genome. In some aspects, AUC as described herein is used to detect the presence of rAAV particles comprising the dimeric form of an scAAV genome. In some aspects, the packaging of scAAV genomes into capsid is monitored by AUC.

**[0419]** In some aspects, the rAAV particles comprise an AAV1 capsid, an AAV2 capsid, an AAV3 capsid, an AAV4 capsid, an AAV5 capsid, an AAV6 capsid (e.g., a wild-type AAV6 capsid, or a variant AAV6 capsid such as ShH10, as described in US 2012/0164106), an AAV7 capsid, an AAV8 capsid, an AAVrh8 capsid, an AAVrh8R, an AAV9 capsid (e.g., a wild-type AAV9 capsid, or a modified AAV9 capsid as described in US 2013/0323226), an AAV10 capsid, an AAVrh10 capsid, an AAV11 capsid, an AAV12 capsid, a tyrosine capsid mutant, a heparin binding capsid mutant, an AAV2R471A capsid, an AAVAAV2/2-7m8 capsid, an AAV DJ capsid (e.g., an AAV-DJ/8 capsid, an AAV-DJ/9 capsid, or any other AAV-DJ capsid, such as any of the capsids described, for example, in US 2012/0066783 or Mao, Y. et al., *BMC Biotechnol.* 16:1 (2016), an AAV2 N587A capsid, an AAV2 E548A capsid, an AAV2 N708A capsid, an AAV V708K capsid, a goat AAV capsid, an AAV1/AAV2 chimeric capsid, a bovine AAV capsid, a mouse AAV capsid, or an AAV capsid described in US Pat. 8,283,151 or WO 2003/042397. In some of the above embodiments described herein, the rAAV particles comprise at least one AAV1 ITR, AAV2 ITR, AAV3 ITR, AAV4 ITR, AAV5 ITR, AAV6 ITR, AAV7 ITR, AAV8 ITR, AAVrh8 ITR, AAV9 ITR, AAV10 ITR, AAVrh10 ITR, AAV11 ITR, AAV12 ITR, AAV DJ ITR, goat AAV ITR, bovine AAV ITR, or mouse AAV ITR. In some aspects, the rAAV particles comprise ITRs from one AAV serotype and AAV capsid from another serotype. For example, the rAAV particles may comprise the nucleic acid to be delivered (e.g., encoding any of the DNA-targeting systems, fusion proteins, gRNA, compositions or components thereof) flanked by at least one AAV2 ITR encapsidated into an AAV9 capsid. Such combinations may be referred to as pseudotyped rAAV particles. Exemplary AAV vectors include those described, for example, in WO 2020/113034, US 20220001028, US 20220001028, US 20210317474, and US 20160097061.

**[0420]** In some aspects, the viral particle is a recombinant AAV particle comprising a nucleic acid to be delivered flanked by one or two ITRs. The nucleic acid is encapsidated in the AAV particle. The AAV particle also comprises capsid proteins. In some aspects, the nucleic acid comprises the protein coding sequence or RNA-expressing sequences to be delivered (e.g., any of the DNA-targeting systems, fusion proteins, gRNA, compositions or components thereof)



operatively linked components in the direction of transcription, control sequences including transcription initiation and termination sequences, thereby forming an expression cassette. The expression cassette is flanked on the 5' and 3' end by at least one functional AAV ITR sequences. By "functional AAV ITR sequences" it is meant that the ITR sequences function as intended for the rescue, replication and packaging of the AAV virion. See Davidson et al., PNAS, 2000, 97(7):3428-32; Passini et al., J. Virol., 2003, 77(12):7034-40; and Pechan et al., Gene Ther., 2009, 16:10-16, all of which are incorporated herein in their entirety by reference. For practicing some aspects of the invention, the recombinant vectors comprise at least all of the sequences of AAV essential for encapsidation and the physical structures for infection by the rAAV. AAV ITRs for use in the vectors of the invention need not have a wild-type nucleotide sequence (e.g., as described in Kotin, Hum. Gene Ther., 1994, 5:793-801), and may be altered by the insertion, deletion or substitution of nucleotides or the AAV ITRs may be derived from any of several AAV serotypes. More than 40 serotypes of AAV are currently known, and new serotypes and variants of existing serotypes continue to be identified. See Gao et al., PNAS, 2002, 99(18): 11854-6; Gao et al., PNAS, 2003, 100(10):6081-6; and Bossis et al., J. Virol., 2003, 77(12):6799-810. Use of any AAV serotype is considered within the scope of the present invention. In some aspects, a rAAV vector is a vector derived from an AAV serotype, including without limitation, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh.10, AAV11, AAV12, a tyrosine capsid mutant, a heparin binding capsid mutant, an AAV2R471A capsid, an AAVAAV2/2-7m8 capsid, an AAV DJ capsid, an AAV2 N587A capsid, an AAV2 E548A capsid, an AAV2 N708A capsid, an AAV V708K capsid, a goat AAV capsid, an AAV1/AAV2 chimeric capsid, a bovine AAV capsid, or a mouse AAV capsid, or the like. In some aspects, the nucleic acid in the AAV comprises an ITR of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh.10, AAV11, AAV12 or the like. In further embodiments, the rAAV particle comprises capsid proteins of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh.10, AAV11, AAV12 or the like. In further embodiments, the rAAV particle comprises capsid proteins of an AAV serotype from Clades A-F (Gao, et al. J. Virol. 2004, 78(12):6381).

**[0421]** Different AAV serotypes are used to optimize transduction of particular target cells or to target specific cell types within a particular target tissue (e.g., a diseased tissue). A rAAV particle can comprise viral proteins and viral nucleic acids of the same serotype or a mixed serotype. For example, a rAAV particle can comprise AAV9 capsid proteins and at least one AAV2 ITR or it can comprise AAV2 capsid proteins and at least one AAV9 ITR. In yet another example, a rAAV particle can comprise capsid proteins from both AAV9 and AAV2, and

further comprise at least one AAV2 ITR. Any combination of AAV serotypes for production of a rAAV particle is provided herein as if each combination had been expressly stated herein.

**[0422]** In some aspects, the AAV comprises at least one AAV1 ITR and capsid protein from any of AAV-DJ, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV2 ITR and capsid protein from any of AAV-DJ, AAV1, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV3 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV4 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV5 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV6, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV6 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV7 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV8, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV8 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV9, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV9 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh.8, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAVrh8 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV8, AAV9, AAVrh10, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAVrh10 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV11, and/or AAV12. In some aspects, the AAV comprises at least one AAV11 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV9, AAVrh10, and/or AAV12. In some aspects, the AAV comprises at least one AAV12 ITR and capsid protein from any of AAV-DJ, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV rh8, AAV9, AAVrh10, and/or AAV11. In some aspects, the AAV comprises at least one AAV-DJ ITR and capsid protein from any of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV rh8, AAV9, AAVrh10, and/or AAV11.

**[0423]** In some aspects, the viral particles comprise a recombinant self-complementing genome. AAV viral particles with self-complementing genomes and methods of use of self-complementing AAV genomes are described in US Patent Nos. 6,596,535; 7,125,717; 7,765,583; 7,785,888; 7,790,154; 7,846,729; 8,093,054; and 8,361,457; and Wang Z., et al., (2003) *Gene Ther* 10:2105-2111, each of which are incorporated herein by reference in its entirety. A rAAV comprising a self-complementing genome will quickly form a double stranded DNA molecule by virtue of its partially complementing sequences (e.g., complementing coding and non-coding strands). In some aspects, an AAV viral particle comprises an AAV genome, wherein the rAAV genome comprises a first heterologous polynucleotide sequence (e.g., a coding strand) and a second heterologous polynucleotide sequence (e.g., the noncoding or antisense strand) wherein the first heterologous polynucleotide sequence can form intrastrand base pairs with the second polynucleotide sequence along most or all of its length. In some aspects, the first heterologous polynucleotide sequence and a second heterologous polynucleotide sequence are linked by a sequence that facilitates intrastrand base-pairing; e.g., a hairpin DNA structure. Hairpin structures include, for example in siRNA molecules. In some aspects, the first heterologous polynucleotide sequence and a second heterologous polynucleotide sequence are linked by a mutated ITR (e.g., the right ITR). The mutated ITR comprises a deletion of the D region comprising the terminal resolution sequence. As a result, on replicating an AAV viral genome, the rep proteins will not cleave the viral genome at the mutated ITR and as such, a recombinant viral genome comprising the following in 5' to 3' order will be packaged in a viral capsid: an AAV ITR, the first heterologous polynucleotide sequence including regulatory sequences, the mutated AAV ITR, the second heterologous polynucleotide in reverse orientation to the first heterologous polynucleotide and a third AAV ITR.

**[0424]** Methods for production of rAAV vectors, including transfection, stable cell line production, and infectious hybrid virus production systems which include adenovirus-AAV hybrids, herpesvirus-AAV hybrids (Conway, JE et al., (1997) *J. Virology* 71(11):8780-8789) and baculovirus-AAV hybrids can be employed. Typically, rAAV production cultures for the production of rAAV virus particles all require; 1) suitable host cells, including, for example, human-derived cell lines such as HeLa, A549, or 293 cells, or insect-derived cell lines such as SF-9, in the case of baculovirus production systems; 2) suitable helper virus function, provided by wild-type or mutant adenovirus (such as temperature sensitive adenovirus), herpes virus, baculovirus, or a plasmid construct providing helper functions; 3) AAV rep and cap genes and gene products; 4) a nucleic acid to be delivered (such as any of the DNA-targeting systems,

fusion proteins, compositions or components thereof) flanked by at least one AAV ITR sequences; and 5) suitable media and media components to support rAAV production. In some aspects, the AAV rep and cap gene products may be from any AAV serotype. In general, but not obligatory, the AAV rep gene product is of the same serotype as the ITRs of the rAAV vector genome as long as the rep gene products may function to replicated and package the rAAV genome. Suitable media may be used for the production of rAAV vectors. These media include, without limitation, media produced by Hyclone Laboratories and JRH including Modified Eagle Medium (MEM), Dulbecco's Modified Eagle Medium (DMEM), custom formulations such as those described in U.S. Patent No. 6,566,118, and Sf-900 II SFM media as described in U.S. Patent No. 6,723,551. In some aspects, the AAV helper functions are provided by adenovirus or HSV. In some aspects, the AAV helper functions are provided by baculovirus and the host cell is an insect cell (e.g., *Spodoptera frugiperda* (Sf9) cells).

**[0425]** Suitable rAAV production culture media of the present invention may be supplemented with serum or serum-derived recombinant proteins at a level of 0.5%-20% (v/v or w/v). Alternatively, rAAV vectors may be produced in serum-free conditions which may also be referred to as media with no animal-derived products. Commercial or custom media designed to support production of rAAV vectors may also be supplemented with one or more cell culture components, including without limitation glucose, vitamins, amino acids, and or growth factors, in order to increase the titer of rAAV in production cultures.

**[0426]** rAAV production cultures can be grown under a variety of conditions (over a wide temperature range, for varying lengths of time, and the like) suitable to the particular host cell being utilized. rAAV production cultures include attachment-dependent cultures which can be cultured in suitable attachment-dependent vessels such as, for example, roller bottles, hollow fiber filters, microcarriers, and packed-bed or fluidized-bed bioreactors. rAAV vector production cultures may also include suspension-adapted host cells such as HeLa, 293, and SF-9 cells which can be cultured in a variety of ways including, for example, spinner flasks, stirred tank bioreactors, and disposable systems such as the Wave bag system.

**[0427]** rAAV vector particles of the invention may be harvested from rAAV production cultures by lysis of the host cells of the production culture or by harvest of the spent media from the production culture, provided the cells are cultured under conditions to cause release of rAAV particles into the media from intact cells, as described in U.S. Patent No. 6,566,118). Suitable methods of lysing cells include for example multiple freeze/thaw cycles, sonication, microfluidization, and treatment with chemicals, such as detergents and/or proteases.

**[0428]** In some aspects, recombinant viral particles for delivery of the nucleic acids,

compositions or components thereof are highly purified, suitably buffered, and concentrated. In some aspects, the viral particles are concentrated to at least about  $1 \times 10^7$  vg/mL to about  $9 \times 10^{13}$  vg/mL or any concentration therebetween.

**[0429]** In some aspects, adeno-associated virus (AAV)-based vectors are generally used vector system for neurologic gene therapy, with an excellent safety record established in multiple clinical trials (Kaplitt et al., (2007) *Lancet* 369:2097-2105; Eberling et al., (2008) *Neurology* 70:1980-1983; Fiandaca et al., (2009) *Neuroimage* 47 Suppl. 2:T27-35). In some cases, effective treatment of neurologic disorders has been hindered by problems associated with the delivery of AAV vectors to affected cell populations. This delivery issue has been especially problematic for disorders involving the cerebral cortex. Simple injections do not distribute AAV vectors effectively, relying on diffusion, which is effective only within a 1- to 3-mm radius. An alternative method, convection-enhanced delivery (CED) (Nguyen et al., (2003) *J. Neurosurg.* 98:584-590), has been used clinically in gene therapy (AAV2-hAADC) for Parkinson's disease (Fiandaca et al., (2008) *Exp. Neurol.* 209:51-57). The underlying principle of CED involves pumping infusate into brain parenchyma under sufficient pressure to overcome the hydrostatic pressure of interstitial fluid, thereby forcing the infused particles into close contact with the dense perivascular space of the brain. Pulsation of these vessels acts as a pump, distributing the particles over large distances throughout the parenchyma (Hadaczek et al., (2006) *Hum. Gene Ther.* 17:291-302). To increase the safety and efficacy of CED a reflux-resistant cannula (Krauze et al., (2009) *Methods Enzymol.* 465:349-362) can be employed along with monitored delivery with real-time MRI. Monitored delivery allows for the quantification and control of aberrant events, such as cannula reflux and leakage of infusate into ventricles (Eberling et al., (2008) *Neurology* 70:1980-1983; Fiandaca et al., (2009) *Neuroimage* 47 Suppl. 2:T27-35; Saito et al., (2011) *Journal of Neurosurgery Pediatrics* 7:522-526).

**[0430]** In some aspects, the nucleic acid to be delivered is operably linked to a promoter. In some aspects, the promoter expresses the nucleic acid to be delivered in a cell of the CNS. In some aspects, the promoter expresses the nucleic acid to be delivered in a brain cell. In some aspects, the promoter expresses the nucleic acid to be delivered in a neuron and/or a glial cell. In some aspects, the neuron is a medium spiny neuron of the caudate nucleus, a medium spiny neuron of the putamen, a neuron of the cortex layer IV and/or a neuron of the cortex layer V. In some aspects, the glial cell is an astrocyte. In some aspects, the promoter is a CBA promoter, a minimum CBA promoter, a CMV promoter or a GUSB promoter. In some aspects, the promoter is inducible. In further embodiments, the rAAV vector comprises one or more of an enhancer, a splice donor/splice acceptor pair, a matrix attachment site, or a polyadenylation

signal.

**[0431]** In some aspects, the methods for delivering a recombinant adeno-associated viral (rAAV) particle to the central nervous system of a subject involve administering the rAAV particle to the striatum, wherein the rAAV particle comprises a rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum of the subject. In some aspects, methods for delivering a rAAV particle to the central nervous system of a subject involve administering the rAAV particle to the striatum, wherein the rAAV particle comprises an rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum of the subject and wherein the rAAV particle comprises an AAV serotype 1 (AAV1) capsid. In some aspects, methods for delivering a rAAV particle to the central nervous system of a subject comprise administering the rAAV particle to the striatum, wherein the rAAV particle comprises an rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum of the subject and wherein the rAAV particle comprises an AAV serotype 2 (AAV2) capsid. In some aspects, methods for treating a central nervous system-related disease in a subject involve administering a rAAV particle to the striatum, wherein the rAAV particle comprises a rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum of the subject. In some aspects, the subject is a human.

**[0432]** In some aspects, a rAAV particle is administered to one or more regions of the central nervous system (CNS). In some aspects, the rAAV particle is administered to the striatum. The striatum is known as a region of the brain that receives inputs from the cerebral cortex (the term “cortex” may be used interchangeably herein) and sends outputs to the basal ganglia (the striatum is also referred to as the striate nucleus and the neostriatum). In some aspects, the striatum controls both motor movements and emotional control/motivation and has been implicated in many neurological diseases, such as Huntington’s disease. Several cell types of interest are located in the striatum, including without limitation spiny projection neurons (also known as medium spiny neurons), GABAergic interneurons, and cholinergic interneurons. Medium spiny neurons make up most of the striatal neurons. These neurons are GABAergic and express dopamine receptors. Each hemisphere of the brain contains a striatum.

**[0433]** In some aspects, important substructures of the striatum include the caudate nucleus and the putamen. In some aspects, the rAAV particle is administered to the caudate nucleus (the term “caudate” may be used interchangeably herein). The caudate nucleus is known as a structure of the dorsal striatum. The caudate nucleus has been implicated in control of functions such as directed movements, spatial working memory, memory, goal-directed actions, emotion,

sleep, language, and learning. Each hemisphere of the brain contains a caudate nucleus.

**[0434]** In some aspects, the rAAV particle is administered to the putamen. Along with the caudate nucleus, the putamen is known as a structure of the dorsal striatum. The putamen comprises part of the lenticular nucleus and connects the cerebral cortex with the substantia nigra and the globus pallidus. Highly integrated with many other structures of the brain, the putamen has been implicated in control of functions such as learning, motor learning, motor performance, motor tasks, and limb movements. Each hemisphere of the brain contains a putamen.

**[0435]** In some aspects, rAAV particles may be administered to one or more sites of the striatum. In some aspects, the rAAV particle is administered to the putamen and the caudate nucleus of the striatum. In some aspects, the rAAV particle is administered to the putamen and the caudate nucleus of each hemisphere of the striatum. In some aspects, the rAAV particle is administered to at least one site in the caudate nucleus and two sites in the putamen.

**[0436]** In some aspects, the rAAV particle is administered to one hemisphere of the brain. For example, in some aspects, the rAAV particle is administered to both hemispheres of the brain. In some aspects, the rAAV particle is administered to the putamen and the caudate nucleus of each hemisphere of the striatum. In some aspects, the composition containing rAAV particles is administered to the striatum of each hemisphere. In some aspects, the composition containing rAAV particles is administered to striatum of the left hemisphere or the striatum of the right hemisphere and/or the putamen of the left hemisphere or the putamen of the right hemisphere. In some aspects, the composition containing rAAV particles is administered to any combination of the caudate nucleus of the left hemisphere, the caudate nucleus of the right hemisphere, the putamen of the left hemisphere and the putamen of the right hemisphere.

**[0437]** In some aspects, the methods involving administration to CNS an effective amount of recombinant viral particles to the striatum can be employed for delivery, wherein the rAAV particle comprises a rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum. In some aspects, the viral titer of the rAAV particles is at least about any of  $5 \times 10^{12}$ ,  $6 \times 10^{12}$ ,  $7 \times 10^{12}$ ,  $8 \times 10^{12}$ ,  $9 \times 10^{12}$ ,  $10 \times 10^{12}$ ,  $11 \times 10^{12}$ ,  $15 \times 10^{12}$ ,  $20 \times 10^{12}$ ,  $25 \times 10^{12}$ ,  $30 \times 10^{12}$ , or  $50 \times 10^{12}$  genome copies/mL. In some aspects, the viral titer of the rAAV particles is about any of  $5 \times 10^{12}$  to  $6 \times 10^{12}$ ,  $6 \times 10^{12}$  to  $7 \times 10^{12}$ ,  $7 \times 10^{12}$  to  $8 \times 10^{12}$ ,  $8 \times 10^{12}$  to  $9 \times 10^{12}$ ,  $9 \times 10^{12}$  to  $10 \times 10^{12}$ ,  $10 \times 10^{12}$  to  $11 \times 10^{12}$ ,  $11 \times 10^{12}$  to  $15 \times 10^{12}$ ,  $15 \times 10^{12}$  to  $20 \times 10^{12}$ ,  $20 \times 10^{12}$  to  $25 \times 10^{12}$ ,  $25 \times 10^{12}$  to  $30 \times 10^{12}$ ,  $30 \times 10^{12}$  to  $50 \times 10^{12}$ , or  $50 \times 10^{12}$  to  $100 \times 10^{12}$  genome copies/mL. In some aspects, the viral titer of the rAAV particles is about any of  $5 \times 10^{12}$  to  $10 \times 10^{12}$ ,  $10 \times 10^{12}$  to  $25 \times 10^{12}$ , or  $25 \times 10^{12}$  to  $50 \times 10^{12}$  genome

copies/mL. In some aspects, the viral titer of the rAAV particles is at least about any of  $5 \times 10^9$ ,  $6 \times 10^9$ ,  $7 \times 10^9$ ,  $8 \times 10^9$ ,  $9 \times 10^9$ ,  $10 \times 10^9$ ,  $11 \times 10^9$ ,  $15 \times 10^9$ ,  $20 \times 10^9$ ,  $25 \times 10^9$ ,  $30 \times 10^9$ , or  $50 \times 10^9$  transducing units/mL. In some aspects, the viral titer of the rAAV particles is about any of  $5 \times 10^9$  to  $6 \times 10^9$ ,  $6 \times 10^9$  to  $7 \times 10^9$ ,  $7 \times 10^9$  to  $8 \times 10^9$ ,  $8 \times 10^9$  to  $9 \times 10^9$ ,  $9 \times 10^9$  to  $10 \times 10^9$ ,  $10 \times 10^9$  to  $11 \times 10^9$ ,  $11 \times 10^9$  to  $15 \times 10^9$ ,  $15 \times 10^9$  to  $20 \times 10^9$ ,  $20 \times 10^9$  to  $25 \times 10^9$ ,  $25 \times 10^9$  to  $30 \times 10^9$ ,  $30 \times 10^9$  to  $50 \times 10^9$  or  $50 \times 10^9$  to  $100 \times 10^9$  transducing units/mL. In some aspects, the viral titer of the rAAV particles is about any of  $5 \times 10^9$  to  $10 \times 10^9$ ,  $10 \times 10^9$  to  $15 \times 10^9$ ,  $15 \times 10^9$  to  $25 \times 10^9$ , or  $25 \times 10^9$  to  $50 \times 10^9$  transducing units/mL. In some aspects, the viral titer of the rAAV particles is at least any of about  $5 \times 10^{10}$ ,  $6 \times 10^{10}$ ,  $7 \times 10^{10}$ ,  $8 \times 10^{10}$ ,  $9 \times 10^{10}$ ,  $10 \times 10^{10}$ ,  $11 \times 10^{10}$ ,  $15 \times 10^{10}$ ,  $20 \times 10^{10}$ ,  $25 \times 10^{10}$ ,  $30 \times 10^{10}$ ,  $40 \times 10^{10}$ , or  $50 \times 10^{10}$  infectious units/mL. In some aspects, the viral titer of the rAAV particles is at least any of about  $5 \times 10^{10}$  to  $6 \times 10^{10}$ ,  $6 \times 10^{10}$  to  $7 \times 10^{10}$ ,  $7 \times 10^{10}$  to  $8 \times 10^{10}$ ,  $8 \times 10^{10}$  to  $9 \times 10^{10}$ ,  $9 \times 10^{10}$  to  $10 \times 10^{10}$ ,  $10 \times 10^{10}$  to  $11 \times 10^{10}$ ,  $11 \times 10^{10}$  to  $15 \times 10^{10}$ ,  $15 \times 10^{10}$  to  $20 \times 10^{10}$ ,  $20 \times 10^{10}$  to  $25 \times 10^{10}$ ,  $25 \times 10^{10}$  to  $30 \times 10^{10}$ ,  $30 \times 10^{10}$  to  $40 \times 10^{10}$ ,  $40 \times 10^{10}$  to  $50 \times 10^{10}$ , or  $50 \times 10^{10}$  to  $100 \times 10^{10}$  infectious units/mL. In some aspects, the viral titer of the rAAV particles is at least any of about  $5 \times 10^{10}$  to  $10 \times 10^{10}$ ,  $10 \times 10^{10}$  to  $15 \times 10^{10}$ ,  $15 \times 10^{10}$  to  $25 \times 10^{10}$ , or  $25 \times 10^{10}$  to  $50 \times 10^{10}$  infectious units/mL.

**[0438]** In some aspects, an effective amount of recombinant viral particles is administered to the striatum, wherein the rAAV particle comprises a rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum. In some aspects, the dose of viral particles administered to the individual is at least about any of  $1 \times 10^8$  to about  $1 \times 10^{13}$  genome copies/kg of body weight. In some aspects, the dose of viral particles administered to the individual is about  $1 \times 10^8$  to  $1 \times 10^{13}$  genome copies/kg of body weight.

**[0439]** In some aspects, an effective amount of recombinant viral particles is administered to the striatum, wherein the rAAV particle comprises a rAAV vector encoding a nucleic acid to be delivered that is expressed in at least the cerebral cortex and striatum. In some aspects, the total amount of viral particles administered to the individual is at least about  $1 \times 10^9$  to about  $1 \times 10^{14}$  genome copies. In some aspects, the total amount of viral particles administered to the individual is about  $1 \times 10^9$  to about  $1 \times 10^{14}$  genome copies.

## 2. *Non-viral vectors*

**[0440]** In some embodiments, the vector is a non-viral vector. In some aspects, exemplary non-viral vectors include polymers, lipids, peptides, inorganic materials, and hybrid systems. In some aspects, the non-viral vector is a lipid nanoparticle (LNP), a liposome, an exosome, or a



cell penetrating peptide. In some aspects, the non-viral vector is a lipid nanoparticle (LNP). In some aspects, the LNP can be used for delivery to the liver. Exemplary non-viral vectors include those described in WO 2020/051561, US 20210301274, Zu et al., *The AAPS Journal* volume 23, Article number: 78 (2021), and Sung et al., *Biomaterials Research* volume 23, Article number: 8 (2019), Nyamay' Antu et al., *Cell & Gene Therapy Insights* 2019; 5(S1):51-57, and Yin et al., *Nature Reviews Genetics* 15:541–555 (2014).

**[0441]** In some embodiments, the vector is a non-viral vector selected from: a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide

**[0442]** In some embodiments, a vector described herein is or comprises a lipid nanoparticle (LNP). In some embodiments, any of the epigenetic-modifying DNA-targeting systems, gRNAs, Cas-gRNA combinations, polynucleotides, fusion proteins, or components thereof described herein, are incorporated in lipid nanoparticles (LNPs), such as for delivery. In some embodiments, the lipid nanoparticle is a vector for delivery. In some embodiments, the nanoparticle may comprise at least one lipid. The lipid may be selected from, but is not limited to, DLin-DMA, DLin-K-DMA, 98N12- 5, C12-200, DLin-MC3-DMA, DLin-KC2-DMA, DODMA, PLGA, PEG, PEG-DMG and PEGylated lipids. In another aspect, the lipid may be a cationic lipid such as, but not limited to, DLin-DMA, DLin-D-DMA, DLin-MC 3 -DMA, DLin-KC2-DMA and DODMA.

**[0443]** Lipid nanoparticles can be used for the delivery of encapsulated or associated (e.g., complexed) therapeutic agents, including nucleic acids and proteins, such as those encoding and/or comprising CRISPR/Cas systems. See, e.g., US Patent No. 10,723,692, US Patent No. 10,941,395, and WO 2015/035136.

**[0444]** In some embodiments, the provided methods involve use of a lipid nanoparticle (LNP) comprising mRNA, such as mRNA encoding a protein component of any of the provided DNA-targeting systems, for example any of the fusion proteins provided herein. In some embodiments, the mRNA can be produced using methods known in the art such as in vitro transcription. In some embodiments of the method, the mRNA comprises a 5' cap. In some embodiments, the 5' cap is an altered nucleotide on the 5' end of primary transcripts such as messenger RNA. In some aspects, the 5' caps of the mRNA improves one or more of RNA stability and processing, mRNA metabolism, the processing and maturation of an RNA transcript in the nucleus, transport of mRNA from the nucleus to the cytoplasm, mRNA stability, and efficient translation of mRNA to protein. In some embodiments, a 5' cap can be a naturally-occurring 5' cap or one that differs from a naturally-occurring cap of an mRNA. A 5' cap may be any 5' cap known to a skilled artisan. In certain embodiments, the 5' cap is selected from the

group consisting of an Anti-Reverse Cap Analog (ARCA) cap, a 7-methyl-guanosine (7mG) cap, a CleanCap® analog, a vaccinia cap, and analogs thereof. For instance, the 5' cap may include, without limitation, an anti-reverse cap analogs (ARCA) (US7074596), 7-methyl-guanosine, CleanCap® analogs, such as Cap 1 analogs (Trilink; San Diego, CA), or enzymatically capped using, for example, a vaccinia capping enzyme or the like. In some embodiments, the mRNA may be polyadenylated. The mRNA may contain various 5' and 3' untranslated sequence elements to enhance expression of the encoded protein and/or stability of the mRNA itself. Such elements can include, for example, posttranslational regulatory elements such as a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE). In some embodiments, the mRNA comprises at least one nucleoside modification. The mRNA may contain modifications of naturally-occurring nucleosides to nucleoside analogs. Any nucleoside analogs known in the art are envisioned. Such nucleoside analogs can include, for example, those described in US 8,278,036. In certain embodiments of the method, the nucleoside modification is selected from the group consisting of a modification from uridine to pseudouridine and uridine to N1- methyl pseudouridine. In particular embodiments of the method the nucleoside modification is from uridine to pseudouridine.

**[0445]** In some embodiments, LNPs useful for in the present methods comprise a cationic lipid selected from DLin-DMA (1,2-dilinoleyloxy-3 -dimethylaminopropane), DLin-MC3 -DMA (dilinoleylmethyl-4-dimethylaminobutyrate), DLin-KC2-DMA (2,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane), DODMA (1,2- dioleyloxy-N,N-dimethyl-3-aminopropane), SS-OP (Bis[2-(4-{2-[4-(cis-9 octadecenoyloxy)phenylacetoxylethyl]piperidinylethyl} disulfide), and derivatives thereof. DLin-MC3-DMA and derivatives thereof are described, for example, in WO 2010/144740. DODMA and derivatives thereof are described, for example, in US 7,745,651 and Mok et al. (1999), *Biochimica et Biophysica Acta*, 1419(2): 137-150. DLin-DMA and derivatives thereof are described, for example, in US 7,799,565. DLin-KC2-DMA and derivatives thereof are described, for example, in US 9,139,554. SS-OP (NOF America Corporation, White Plains, NY) is described, for example, at [https://www.nofamerica.com/store/index.php?dispatch=products.view&product\\_id=962](https://www.nofamerica.com/store/index.php?dispatch=products.view&product_id=962). Additional and non-limiting examples of cationic lipids include methylpyridiyl-dialkyl acid (MPDACA), palmitoyl-oleoyl-nor-arginine (PONA), guanidino-dialkyl acid (GUADACA), 1,2-di-O-octadecenyl-3-trimethylammonium propane (DOTMA), 1,2- dioleyl-3-trimethylammonium-propane (DOTAP), Bis{2-[N-methyl-N-(a-D-tocopherolhemisuccinatepropyl)amino]ethyl} disulfide (SS-33/3AP05), Bis{2-[4-(a-D-

tocopherolhemisuccinateethyl)piperidyl] ethyl} disulfide (SS33/4PE15), Bis{2-[4-(cis-9-octadecenoateethyl)-1-piperidinyl] ethyl} disulfide (SS18/4PE16), and Bis{2-[4-(cis,cis-9,12-octadecadienoateethyl)-1-piperidinyl] ethyl} disulfide (SS18/4PE13). In further embodiments, the lipid nanoparticles also comprise one or more non-cationic lipids and a lipid conjugate.

**[0446]** In some embodiments, the molar concentration of the cationic lipid is from about 20% to about 80%, from about 30% to about 70%, from about 40% to about 60%, from about 45% to about 55%, or about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, or about 80% of the total lipid molar concentration, wherein the total lipid molar concentration is the sum of the cationic lipid, the non-cationic lipid, and the lipid conjugate molar concentrations. In certain embodiments, the lipid nanoparticles comprise a molar ratio of cationic lipid to any of the polynucleotides of from about 1 to about 20, from about 2 to about 16, from about 4 to about 12, from about 6 to about 10, or about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, or about 20.

**[0447]** In some embodiments, the lipid nanoparticles can comprise at least one non-cationic lipid. In particular embodiments, the molar concentration of the non-cationic lipids is from about 20% to about 80%, from about 30% to about 70%, from about 40% to about 70%, from about 40% to about 60%, from about 46% to about 50%, or about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 48.5%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, or about 80% of the total lipid molar concentration. Non-cationic lipids include, in some embodiments, phospholipids and steroids.

**[0448]** In some embodiments, phospholipids useful for the lipid nanoparticles described herein include, but are not limited to, 1,2-Distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-Didecanoyl-sn-glycero-3-phosphocholine (DDPC), 1,2-Dierucoyl-sn-glycero-3-phosphate(Sodium Salt) (DEPA-NA), 1,2-Dierucoyl-sn-glycero-3-phosphocholine (DEPC), 1,2-Dierucoyl-sn-glycero-3-phosphoethanolamine (DEPE), 1,2-Dierucoyl-sn-glycero-3[Phospho-rac-(1-glycerol)](Sodium Salt) (DEPG-NA), 1,2-Dilinoleoyl-sn-glycero-3-phosphocholine (DLOPC), 1,2-Dilauroyl-sn-glycero-3-phosphate(Sodium Salt) (DLPA-NA), 1,2-Dilauroyl-sn-glycero-3-phosphocholine (DLPC), 1,2-Dilauroyl-sn-glycero-3-phosphoethanolamine (DLPE), 1,2-Dilauroyl-sn-glycero-3[Phospho-rac-(1-glycerol)](Sodium Salt) (DLPG-NA), 1,2-Dilauroyl-sn-glycero-3[Phospho-rac-(1-glycerol)](Ammonium Salt) (DLPG-NH<sub>4</sub>), 1,2-Dilauroyl-sn-glycero-3-phosphoserine(Sodium Salt) (DLPS-NA), 1,2-Dimyristoyl-sn-glycero-3-phosphate(Sodium Salt) (DMPA-NA), 1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC),

1,2-Dimyristoyl- sn-glycero-3-phosphoethanolamine (DMPE), 1,2-Dimyristoyl-sn-glycero-3[Phospho-rac-(1- glycerol)(Sodium Salt) (DMPG-NA), 1,2-Dimyristoyl-sn-glycero-3[Phospho-rac-(1- glycerol)(Ammonium Salt) (DMPG-NH<sub>4</sub>), 1,2-Dimyristoyl-sn-glycero-3[Phospho-rac-(1- glycerol)(Sodium/ Ammonium Salt) (DMPG-NH<sub>4</sub>/NA), 1,2-Dimyristoyl-sn-glycero-3-phosphoserine(Sodium Salt) (DMPS-NA), 1,2-Dioleoyl-sn-glycero-3-phosphate(Sodium Salt) (DOPA-NA), 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-Dioleoyl-sn- glycero-3-phosphoethanolamine (DOPE), 1,2-Dioleoyl-sn-glycero-3[Phospho-rac-(1- glycerol)(Sodium Salt) (DOPG-NA), 1,2-Dioleoyl-sn-glycero-3-phosphoserine(Sodium Salt) (DOPS-NA), 1,2-Dipalmitoyl-sn-glycero-3-phosphate(Sodium Salt) (DPPA-NA), 1,2- Dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE), 1,2-Dipalmitoyl-sn-glycero- 3[Phospho-rac-(1-glycerol)(Sodium Salt) (DPPG-NA), 1,2-Dipalmitoyl-sn-glycero- 3[Phospho-rac-(1-glycerol)(Ammonium Salt) (DPPG-NH<sub>4</sub>), 1,2-Dipalmitoyl-sn-glycero-3- phosphoserine(Sodium Salt) (DPPS-NA), 1,2-Distearoyl-sn-glycero-3-phosphate(Sodium Salt) (DSPA-NA), 1,2-Distearoyl-sn-glycero-3-phosphoethanolamine (DSPE), 1,2- Distearoyl-sn-glycero-3[Phospho-rac-(1-glycerol)(Sodium Salt) (DSPG-NA), 1,2-Distearoyl- sn-glycero-3[Phospho-rac-(1-glycerol)(Ammonium Salt) (DSPG-NH<sub>4</sub>), 1,2-Distearoyl-sn- glycero-3-phosphoserine(Sodium Salt) (DSPS-NA), Egg-PC (EPC), Hydrogenated Egg PC (HEPC), Hydrogenated Soy PC (HSPC), 1-Myristoyl-sn-glycero-3-phosphocholine (LY S OPCM YRIS TIC ), 1-Palmitoyl-sn-glycero-3-phosphocholine (LYSOPCPALMITIC), 1- Stearoyl-sn-glycero-3-phosphocholine (LYSOPC STEARIC), 1- Myristoyl-2-palmitoyl-sn- glycero3-phosphocholine (MPPC), 1-Myristoyl-2-stearoyl-sn-glycero-3-phosphocholine (MSPC), 1-Palmitoyl-2-myristoyl-sn-glycero-3-phosphocholine (PMPC), 1- Palmitoyl-2- oleoyl-sn-glycero-3-phosphocholine (POPC), 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-Palmitoyl-2-olcoyl-sn-glycero-3[Phospho-rac-(1- glycerol)] (Sodium Salt) (POPG-NA), 1-Palmitoyl-2-stearoyl-sn-glycero-3-phosphocholine (PS PC), 1- Stearoyl-2-myristoyl-sn-glycero-3-phosphocholine (SMPC), 1-Stearoyl-2-oleoyl- sn-glycero-3-phosphocholine (SOPC), and 1-Stearoyl-2-palmitoyl-sn-glycero-3- phosphocholine (SPPC). In particular embodiments, the phospholipid is DSPC. In particular embodiments, the phospholipid is DOPE. In particular embodiments, the phospholipid is DOPC.

**[0449]** In some embodiments, the non-cationic lipids comprised by the lipid nanoparticles include one or more steroids. Steroids useful for the lipid nanoparticles described herein include, but are not limited to, cholestanes such as cholesterol, cholanes such as cholic acid, pregnanes such as progesterone, androstanes such as testosterone, and estranes such as estradiol. Further steroids include, but are not limited to, cholesterol (ovine), cholesterol sulfate, desmosterol-d6,

cholesterol-d7, lathosterol-d7, desmosterol, stigmasterol, lanosterol, dehydrocholesterol, dihydrolanosterol, zymosterol, lathosterol, zymosterol-d5, 14-demethyl-lanosterol, 14-demethyl-lanosterol-d6, 8(9)-dehydrocholesterol, 8(14)-dehydrocholesterol, diosgenin, DHEA sulfate, DHEA, lanosterol-d6, dihydrolanosterol-d7, campesterol-d6, sitosterol, lanosterol-95, Dihydro FF-MAS-d6, zymostenol-d7, zymostenol, sitostanol, campestanol, campesterol, 7-dehydrodesmosterol, pregnenolone, sitosterol-d7, Dihydro T-MAS, Delta 5-avenasterol, Brassicasterol, Dihydro FF-MAS, 24-methylene cholesterol, cholic acid derivatives, cholesteryl esters, and glycosylated sterols. In particular embodiments, the lipid nanoparticles comprise cholesterol.

**[0450]** In some embodiments, the lipid nanoparticles comprise a lipid conjugate. Such lipid conjugates include, but are not limited to, ceramide PEG derivatives such as C8 PEG2000 ceramide, C16 PEG2000 ceramide, C8 PEG5000 ceramide, C16 PEG5000 ceramide, C8 PEG750 ceramide, and C16 PEG750 ceramide, phosphoethanolamine PEG derivatives such as 16:0 PEG5000PE, 14:0 PEG5000 PE, 18:0 PEG5000 PE, 18:1 PEG5000 PE, 16:0 PEG3000 PE, 14:0 PEG3000 PE, 18:0 PEG3000 PE, 18:1 PEG3000 PE, 16:0 PEG2000 PE, 14:0 PEG2000 PE, 18:0 PEG2000 PE, 18:1 PEG2000 PE 16:0 PEG1000 PE, 14:0 PEG1000 PE, 18:0 PEG1000 PE, 18:1 PEG 1000 PE, 16:0 PEG750 PE, 14:0 PEG750 PE, 18:0 PEG750 PE, 18:1 PEG750 PE, 16:0 PEG550 PE, 14:0 PEG550 PE, 18:0 PEG550 PE, 18:1 PEG550 PE, 16:0 PEG350 PE, 14:0 PEG350 PE, 18:0 PEG350 PE, and 18:1 PEG350, sterol PEG derivatives such as Chol-PEG600, and glycerol PEG derivatives such as DMG-PEG5000, DSG-PEG5000, DPG-PEG5000, DMG-PEG3000, DSG-PEG3000, DPG-PEG3000, DMG-PEG2000, DSG-PEG2000, DPG-PEG2000, DMG-PEG1000, DSG-PEG1000, DPG-PEG1000, DMG-PEG750, DSG-PEG750, DPG-PEG750, DMG-PEG550, DSG-PEG550, DPG-PEG550, DMG-PEG350, DSG-PEG350, and DPG-PEG350. In some embodiments, the lipid conjugate is a DMG-PEG. In some particular embodiments, the lipid conjugate is DMG-PEG2000. In some particular embodiments, the lipid conjugate is DMG-PEG5000.

**[0451]** It is within the level of a skilled artisan to select the cationic lipids, non-cationic lipids and/or lipid conjugates which comprise the lipid nanoparticle, as well as the relative molar ratio of such lipids to each other, such as based upon the characteristics of the selected lipid(s), the nature of the delivery to the intended target cells, and the characteristics of the nucleic acids and/or proteins to be delivered. Additional considerations include, for example, the saturation of the alkyl chain, as well as the size, charge, pH, pKa, fusogenicity and toxicity of the selected lipid(s). Thus, the molar ratios of each individual component may be adjusted accordingly.

**[0452]** The lipid nanoparticles for use in the method can be prepared by various techniques

which are known to a skilled artisan. Nucleic acid-lipid particles and methods of preparation are disclosed in, for example, U.S. Patent Publication Nos. 20040142025 and 20070042031.

**[0453]** In some embodiments, the lipid nanoparticles will have a size within the range of about 25 to about 500 nm. In some embodiments, the lipid nanoparticles have a size from about 50 nm to about 300 nm, or from about 60 nm to about 120 nm. The size of the lipid nanoparticles may be determined by quasi-electric light scattering (QELS) as described in Bloomfield, *Ann. Rev. Biophys. Bioeng.*, 10:421A150 (1981). A variety of methods are known in the art for producing a population of lipid nanoparticles of particular size ranges, for example, sonication or homogenization. One such method is described in U.S. Pat. No. 4,737,323.

**[0454]** In some embodiments, the lipid nanoparticles comprise a cell targeting molecule such as, for example, a targeting ligand (e.g., antibodies, scFv proteins, DART molecules, peptides, aptamers, and the like) anchored on the surface of the lipid nanoparticle that selectively binds the lipid nanoparticles to the targeted cell, such as any cell described herein.

**[0455]** In some embodiments, the vector exhibits tropism for one or more cell types. For example, the vector may exhibit liver cell and/or hepatocyte tropism, neural cell (e.g. neuron or glia) tropism, immune cell tropism, or tropism for any suitable cell type.

**[0456]** In some aspects, provided herein are pluralities of vectors that comprise any of the vectors described herein, and one or more additional vectors. In some embodiments, the one or more additional vectors comprise one or more additional polynucleotides encoding any additional transcriptional activation domain, multipartite effector such as multipartite activator, DNA-targeting domain, gRNA, fusion protein, DNA-targeting system, or a portion, component, or combination thereof. In some aspects, provided are pluralities of vectors, that include: a first vector comprising any of the polynucleotides described herein; a second vector comprising any of the polynucleotides described herein; and optionally one or more additional vectors comprising any of the polynucleotides described herein.

**[0457]** In some aspects, vectors provided herein may be referred to as delivery vehicles. In some aspects, any of the DNA-targeting systems, components thereof, or polynucleotides disclosed herein can be packaged into or on the surface of delivery vehicles for delivery to cells. Delivery vehicles contemplated include, but are not limited to, nanospheres, liposomes, quantum dots, nanoparticles, polyethylene glycol particles, hydrogels, and micelles. As described in the art, a variety of targeting moieties can be used to enhance the preferential interaction of such vehicles with desired cell types or locations.

**[0458]** Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a cell.

Suitable methods include, include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery, and the like. In some embodiments, the composition may be delivered by mRNA delivery and ribonucleoprotein (RNP) complex delivery. Direct delivery of the RNP complex, including the DNA-targeting domain complexed with the sgRNA, can eliminate the need for intracellular transcription and translation and can offer a robust platform for host cells with low transcriptional and translational activity. The RNP complexes can be introduced into the host cell by any of the methods known in the art.

**[0459]** Nucleic acids or RNPs of the disclosure can be incorporated into a host using virus-like particles (VLP). VLPs contain normal viral vector components, such as envelope and capsids, but lack the viral genome. For instance, nucleic acids expressing the Cas and sgRNA can be fused to the viral vector components such as gag and introduced into producer cells. The resulting virus-like particles containing the sgRNA-expressing vectors can infect the host cell for efficient editing.

**[0460]** Introduction of the complexes, polypeptides, and nucleic acids of the disclosure can occur by protein transduction domains (PTDs). PTDs, including the human immunodeficiency virus-1 TAT, herpes simplex virus-1 VP22, *Drosophila Antennapedia Antp*, and the polyarginines, are peptide sequences that can cross the cell membrane, enter a host cell, and deliver the complexes, polypeptides, and nucleic acids into the cell.

**[0461]** Introduction of the complexes, polypeptides, and nucleic acids of the disclosure into cells can occur by viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, nucleofection, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro-injection, nanoparticle-mediated nucleic acid delivery, and the like, for example as described in WO 2017/193107, WO 2016/123578, WO 2014/152432, WO 2014/093661, WO 2014/093655, or WO 2021/226555.

**[0462]** Various methods for the introduction of polynucleotides are well known and may be used with the provided methods and compositions. Exemplary methods include those for transfer of polynucleotides encoding the DNA targeting systems provided herein, including via viral, e.g., retroviral or lentiviral, transduction, transposons, and electroporation.

### C. Pharmaceutical Compositions and Formulations

[0463] Also provided are compositions, such as pharmaceutical compositions and formulations for administration, that include any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing. In some aspects, the pharmaceutical composition comprises one or more pharmaceutically acceptable carriers.

[0464] In some aspects, the pharmaceutical composition contains one or more DNA-targeting systems provided herein or a component thereof. In some aspects, the pharmaceutical composition comprises one or more vectors, e.g., viral vectors that contain polynucleotides that encode one or more components of the DNA-targeting systems provided herein. Such compositions can be used in accord with the provided methods, and/or with the provided articles of manufacture or compositions, such as in the prevention or treatment of diseases, conditions, and disorders, or in detection, diagnostic, and prognostic methods.

[0465] The term “pharmaceutical formulation” refers to a preparation which is in such form as to permit the biological activity of an active ingredient contained therein to be effective, and which contains no additional components which are unacceptably toxic to a subject to which the formulation would be administered.

[0466] A “pharmaceutically acceptable carrier” refers to an ingredient in a pharmaceutical formulation, other than an active ingredient, which is nontoxic to a subject. A pharmaceutically acceptable carrier includes, but is not limited to, a buffer, excipient, stabilizer, or preservative.

[0467] In some aspects, the choice of carrier is determined in part by the particular cell or agent and/or by the method of administration. Accordingly, there are a variety of suitable formulations. For example, the pharmaceutical composition can contain preservatives. Suitable preservatives may include, for example, methylparaben, propylparaben, sodium benzoate, and benzalkonium chloride. In some aspects, a mixture of two or more preservatives is used. The preservative or mixtures thereof are typically present in an amount of about 0.0001% to about 2% by weight of the total composition. Carriers are described, e.g., by Remington’s *Pharmaceutical Sciences* 16th edition, Osol, A. Ed. (1980). Pharmaceutically acceptable carriers are generally nontoxic to recipients at the dosages and concentrations employed, and include, but are not limited to: buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride; benzethonium chloride;



phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g. Zn-protein complexes); and/or non-ionic surfactants such as polyethylene glycol (PEG).

**[0468]** The pharmaceutical composition in some embodiments contains components in amounts effective to treat or prevent the disease or condition, such as a therapeutically effective or prophylactically effective amount. Therapeutic or prophylactic efficacy in some embodiments is monitored by periodic assessment of treated subjects. For repeated administrations over several days or longer, depending on the condition, the treatment is repeated until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful and can be determined. The desired dosage can be delivered by a single bolus administration of the composition, by multiple bolus administrations of the composition, or by continuous infusion administration of the composition.

**[0469]** The composition can be administered by any suitable means, for example, by bolus infusion, by injection, e.g., intravenous or subcutaneous injections, intraocular injection, periocular injection, subretinal injection, intravitreal injection, trans-septal injection, subcleral injection, intrachoroidal injection, intracameral injection, subconjunctival injection, subconjunctival injection, sub-Tenon's injection, retrobulbar injection, peribulbar injection, or posterior juxtasceral delivery. In some embodiments, they are administered by parenteral, intrapulmonary, and intranasal, and, if desired for local treatment, intralesional administration. Parenteral infusions include intramuscular, intravenous, intraarterial, intraperitoneal, or subcutaneous administration. In some embodiments, a given dose is administered by a single bolus administration of the composition. In some embodiments, it is administered by multiple bolus administrations of the composition, for example, over a period of no more than 3 days, or by continuous infusion administration of the composition.

**[0470]** For the prevention or treatment of disease, the appropriate dosage may depend on the type of disease to be treated, the type of agent or agents, the type of cells or recombinant receptors, the severity and course of the disease, whether the agent or cells are administered for preventive or therapeutic purposes, previous therapy, the subject's clinical history and response

to the agent or the cells, and the discretion of the attending physician. The compositions are in some embodiments suitably administered to the subject at one time or over a series of treatments.

[0471] Formulations include those for oral, intravenous, intraperitoneal, subcutaneous, pulmonary, transdermal, intramuscular, intranasal, buccal, sublingual, or suppository administration. In some embodiments, the agent or cell populations are administered parenterally. The term “parenteral,” as used herein, includes intravenous, intramuscular, subcutaneous, rectal, vaginal, and intraperitoneal administration. In some embodiments, the agent or cell populations are administered to a subject using peripheral systemic delivery by intravenous, intraperitoneal, or subcutaneous injection.

[0472] Compositions in some embodiments are provided as sterile liquid preparations, e.g., isotonic aqueous solutions, suspensions, emulsions, dispersions, or viscous compositions, which may in some aspects be buffered to a selected pH. Liquid preparations are normally easier to prepare than gels, other viscous compositions, and solid compositions. Additionally, liquid compositions are somewhat more convenient to administer, especially by injection. Viscous compositions, on the other hand, can be formulated within the appropriate viscosity range to provide longer contact periods with specific tissues. Liquid or viscous compositions can comprise carriers, which can be a solvent or dispersing medium containing, for example, water, saline, phosphate buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol) and suitable mixtures thereof.

[0473] Sterile injectable solutions can be prepared by incorporating the agent or cells in a solvent, such as in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose, dextrose, or the like.

[0474] The formulations to be used for in vivo administration are generally sterile. Sterility may be readily accomplished, e.g., by filtration through sterile filtration membranes.

#### **IV. METHODS OF MODULATING AND METHODS OF TREATMENT**

[0475] Provided herein are methods of treatment, e.g., including administering any of the compositions, such as pharmaceutical compositions described herein. In some aspects, also provided are methods of administering any of the compositions described herein to a subject, such as a subject that has a disease or disorder. The compositions, such as pharmaceutical compositions, described herein are useful in a variety of therapeutic, diagnostic and prophylactic indications. For example, the compositions are useful in treating a variety of diseases and disorders in a subject. Such methods and uses include therapeutic methods and uses, for

example, involving administration of the compositions, to a subject having a disease, condition, or disorder, such as a tumor or cancer. In some embodiments, the compositions are administered in an effective amount to effect treatment of the disease or disorder. Uses include uses of the compositions in such methods and treatments, and in the preparation of a medicament in order to carry out such therapeutic methods. In some embodiments, the methods are carried out by administering the compositions, to the subject having or suspected of having the disease or condition. In some embodiments, the methods thereby treat the disease or condition or disorder in the subject. Also provided are therapeutic methods for administering the cells and compositions to subjects, e.g., patients.

**[0476]** Provided are methods for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a cell, that involve: introducing any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, into the cell.

**[0477]** In some embodiments, the cell is from a subject that has or is suspected of having Rett syndrome.

**[0478]** Also provided herein are methods for modulating the expression of MeCP2 in a subject, the method comprising: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations of gRNAs described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the plurality of vectors described herein, or a portion or a component of any of the foregoing, to the subject.

**[0479]** In some embodiments, the subject has or is suspected of having Rett syndrome.

**[0480]** Also provided herein are methods of treating Rett syndrome, the method comprising: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

**[0481]** In some embodiments, the cell is a heart cell, a skeletal muscle cell, a nervous system

cell, or an induced pluripotent stem cell. In some embodiments, the introducing, contacting or administering is carried out in vivo or ex vivo. In some embodiments, following the introducing, contacting or administering, the expression of MeCP2 is increased in the cell or the subject. In some embodiments, the expression of MeCP2 is increased at least about 1.2-fold, 1.25-fold, 1.3-fold, 1.4-fold, 1.5-fold, 1.6-fold, 1.7-fold, 1.75-fold, 1.8-fold, 1.9-fold, 2-fold, 2.5-fold, 3-fold, 4-fold, or 5-fold. In some embodiments, the expression is increased by less than about 10-fold, 9-fold, 8-fold, 7-fold or 6-fold. In some embodiments, the subject is a human.

**[0482]** In some embodiments, the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some embodiments, the cell is from a subject that has or is suspected of having Rett syndrome.

**[0483]** Provided are methods for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a subject, that involve: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to the subject.

**[0484]** In some embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some embodiments, the subject has or is suspected of having Rett syndrome.

**[0485]** Provided are methods of treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, that involve: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0486]** Provided are methods of treating Rett syndrome, that involve: administering any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any

of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

**[0487]** Provided are pharmaceutical compositions that include any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing.

**[0488]** Provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0489]** Provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in treating Rett syndrome.

**[0490]** Provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0491]** Provided are pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for use in the manufacture of a medicament for treating Rett syndrome.

**[0492]** In some embodiments, the pharmaceutical composition is to be administered to a subject. In some embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some embodiments, the subject has or is suspected of having Rett syndrome.

**[0493]** Provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0494]** Provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, for treating Rett syndrome.

**[0495]** Provided are uses of pharmaceutical compositions, such as any of the pharmaceutical compositions described herein, in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

**[0496]** Provided are uses of pharmaceutical compositions, such as any of the pharmaceutical

compositions described herein, in the manufacture of a medicament for treating Rett syndrome.

**[0497]** In some embodiments, the pharmaceutical composition is to be administered to a subject. In some embodiments, the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome. In some embodiments, the subject has or is suspected of having Rett syndrome.

**[0498]** Provided are cells comprising any of the DNA-targeting systems described herein, any of the gRNAs described herein, any of the combinations described herein, any of the fusion proteins described herein, any of the polynucleotides described herein, any of the pluralities of polynucleotides described herein, any of the vectors described herein, any of the pluralities of vectors described herein, or a portion or a component of any of the foregoing.

**[0499]** In some embodiments, a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome. In some embodiments, the mutant MeCP2 allele comprises a mutation corresponding to R255X. In some embodiments, a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome. In some embodiments, a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject. In some embodiments, the cell is a nervous system cell, or an induced pluripotent stem cell.

**[0500]** In some embodiments, the introducing, contacting or administering is carried out in vivo or ex vivo.

**[0501]** In some embodiments, following the introducing, contacting or administering, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject. In some embodiments, the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold. In some embodiments, the expression is increased by less than about 200-fold, 150-fold, or 100-fold. In some embodiments, the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

**[0502]** In some embodiments, the subject is a human.

**[0503]** In some aspects, methods of treating of treating a disease or disorder, such as diseases or disorders associated with dysregulation or reduced activity, function or expression of MeCP2, such as Rett syndrome, in an individual or a subject, involve administering to the individual or the subject AAV particles. The AAV particles may be administered to a particular tissue of interest, or it may be administered systemically. In some aspects, an effective amount of the AAV particles may be administered parenterally. Parenteral routes of administration may

include without limitation intravenous, intraosseous, intra-arterial, intracerebral, intramuscular, intrathecal, subcutaneous, intracerebroventricular, and others. In some aspects, an effective amount of AAV particles may be administered through one route of administration. In some aspects, an effective amount of AAV particles may be administered through a combination of more than one route of administration. In some aspects, the individual is a mammal. In some aspects, the individual is a human.

**[0504]** An effective amount of AAV particles comprising an oversized AAV genome is administered, depending on the objectives of treatment. For example, where a low percentage of transduction can achieve the desired therapeutic effect, then the objective of treatment is generally to meet or exceed this level of transduction. In some instances, this level of transduction can be achieved by transduction of only about 1 to 5% of the target cells of the desired tissue type, In some aspects at least about 20% of the cells of the desired tissue type, In some aspects at least about 50%, In some aspects at least about 80%, In some aspects at least about 95%, In some aspects at least about 99% of the cells of the desired tissue type. As a guide, the number of particles administered per injection is generally between about  $1 \times 10^6$  and about  $1 \times 10^{14}$  particles, between about  $1 \times 10^7$  and  $1 \times 10^{13}$  particles, between about  $1 \times 10^9$  and  $1 \times 10^{12}$  particles or about  $1 \times 10^9$  particles, about  $1 \times 10^{10}$  particles, or about  $1 \times 10^{11}$  particles. The rAAV composition may be administered by one or more administrations, either during the same procedure or spaced apart by days, weeks, months, or years. One or more of any of the routes of administration described herein may be used. In some aspects, multiple vectors may be used to treat the human.

**[0505]** Methods to identify cells transduced by AAV viral particles can be employed; for example, immunohistochemistry or the use of a marker such as enhanced green fluorescent protein can be used to detect transduction of viral particles; for example viral particles comprising a rAAV capsid with one or more substitutions of amino acids.

**[0506]** In some aspects the AAV viral particles comprising an oversized AAV genome with are administered to more than one location simultaneously or sequentially. In some aspects, multiple injections of rAAV viral particles are no more than one hour, two hours, three hours, four hours, five hours, six hours, nine hours, twelve hours or 24 hours apart.

## V. KITS AND ARTICLES OF MANUFACTURE

**[0507]** Also provided are articles of manufacture, systems, apparatuses, and kits useful in performing the provided embodiments. In some embodiments, the provided articles of manufacture or kits contain one or more components of the one or more components of the

DNA-targeting system provided herein. In some embodiments, the articles of manufacture or kits include polypeptides, nucleic acids, vectors and/or polynucleotides useful in performing the provided methods.

**[0508]** In some embodiments, the articles of manufacture or kits include one or more containers, typically a plurality of containers, packaging material, and a label or package insert on or associated with the container or containers and/or packaging, generally including instructions for use, *e.g.*, instructions for introducing or administering.

**[0509]** Also provided are articles of manufacture, systems, apparatuses, and kits useful in administering the provided compositions, *e.g.*, pharmaceutical compositions, *e.g.*, for use in therapy or treatment. In some embodiments, the articles of manufacture or kits provided herein contain vectors and/or plurality of vectors, such as any vectors and/or plurality of vectors described herein. In some aspects, the articles of manufacture or kits provided herein can be used for administration of the vectors and/or plurality of vectors, and can include instructions for use.

**[0510]** The articles of manufacture and/or kits containing cells or cell compositions for therapy, may include a container and a label or package insert on or associated with the container. Suitable containers include, for example, bottles, vials, syringes, IV solution bags, etc. The containers may be formed from a variety of materials such as glass or plastic. The container in some embodiments holds a composition which is by itself or combined with another composition effective for treating, preventing and/or diagnosing the condition. In some embodiments, the container has a sterile access port. Exemplary containers include an intravenous solution bags, vials, including those with stoppers pierceable by a needle for injection, or bottles or vials for orally administered agents. The label or package insert may indicate that the composition is used for treating a disease or condition. The article of manufacture may further include a package insert indicating that the compositions can be used to treat a particular condition. Alternatively, or additionally, the article of manufacture may further include another or the same container comprising a pharmaceutically-acceptable buffer. It may further include other materials such as other buffers, diluents, filters, needles, and/or syringes.

## **VI. DEFINITIONS**

**[0511]** Unless defined otherwise, all terms of art, notations and other technical and scientific terms or terminology used herein are intended to have the same meaning as is commonly understood by one of ordinary skill in the art to which the claimed subject matter pertains. In



some cases, terms with commonly understood meanings are defined herein for clarity and/or for ready reference, and the inclusion of such definitions herein should not necessarily be construed to represent a substantial difference over what is generally understood in the art.

**[0512]** As used herein, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. For example, “a” or “an” means “at least one” or “one or more.” It is understood that aspects and variations described herein include “consisting” and/or “consisting essentially of” aspects and variations.

**[0513]** Throughout this disclosure, various aspects of the claimed subject matter are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the claimed subject matter. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges as well as individual numerical values within that range. For example, where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the claimed subject matter. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the claimed subject matter, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the claimed subject matter. This applies regardless of the breadth of the range.

**[0514]** The term “about” as used herein refers to the usual error range for the respective value readily known. Reference to “about” a value or parameter herein includes (and describes) embodiments that are directed to that value or parameter per se. For example, description referring to “about X” includes description of “X”. In some embodiments, “about” may refer to  $\pm 25\%$ ,  $\pm 20\%$ ,  $\pm 15\%$ ,  $\pm 10\%$ ,  $\pm 5\%$ , or  $\pm 1\%$ .

**[0515]** In some aspects, corresponding positions of the one or more modifications, such as one or more substitutions, can be determined in reference to positions of a reference amino acid sequence or a reference nucleotide sequence. As used herein, recitation that nucleotides or amino acid positions “correspond to” nucleotides or amino acid positions in a disclosed sequence, such as set forth in the Sequence listing, refers to nucleotides or amino acid positions identified upon alignment with the disclosed sequence to maximize identity using a standard alignment algorithm, such as the GAP algorithm or other available algorithms. By aligning the sequences, corresponding residues can be identified, for example, using conserved and identical amino acid residues as guides. In general, to identify corresponding positions, the sequences of

amino acids are aligned so that the highest order match is obtained (see, e.g. : Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; Carrillo et al. (1988) SIAM J Applied Math 48: 1073). Alignment for determining corresponding positions can be obtained in various ways, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Appropriate parameters for aligning sequences can be determined, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For example, corresponding residues can be determined by alignment of a reference sequence that is a wild-type Cas protein by available alignment methods. By aligning the sequences, one skilled in the art can identify corresponding residues, for example, using conserved and/or identical amino acid residues as guides.

**[0516]** The term “vector,” as used herein, refers to a nucleic acid molecule capable of propagating another nucleic acid to which it is linked. The term includes the vector as a self-replicating nucleic acid structure as well as the vector incorporated into the genome of a host cell into which it has been introduced. Certain vectors are capable of directing the expression of nucleic acids to which they are operatively linked. Such vectors are referred to herein as “expression vectors.” Among the vectors are viral vectors, such as adenoviral vectors.

**[0517]** As used herein, “percent (%) amino acid sequence identity” and “percent identity” when used with respect to an amino acid sequence (reference polypeptide sequence) is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the reference polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various known ways, in some embodiments, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Appropriate parameters for aligning sequences can be determined, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

**[0518]** In some embodiments, “operably linked” may include the association of components, such as a DNA sequence, e.g. a heterologous nucleic acid) and a regulatory sequence(s), in such

a way as to permit gene expression when the appropriate molecules (e.g. transcriptional activator proteins) are bound to the regulatory sequence. Hence, it means that the components described are in a relationship permitting them to function in their intended manner.

**[0519]** An amino acid substitution may include replacement of one amino acid in a polypeptide with another amino acid. The substitution may be a conservative amino acid substitution or a non-conservative amino acid substitution. Amino acid substitutions may be introduced into a binding molecule, e.g., antibody, of interest and the products screened for a desired activity, e.g., retained/improved antigen binding, decreased immunogenicity, or improved ADCC or CDC.

**[0520]** Amino acids generally can be grouped according to the following common side-chain properties:

- (1) hydrophobic: Norleucine, Met, Ala, Val, Leu, Ile;
- (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln;
- (3) acidic: Asp, Glu;
- (4) basic: His, Lys, Arg;
- (5) residues that influence chain orientation: Gly, Pro;
- (6) aromatic: Trp, Tyr, Phe.

**[0521]** In some embodiments, conservative substitutions can involve the exchange of a member of one of these classes for another member of the same class. In some embodiments, non-conservative amino acid substitutions can involve exchanging a member of one of these classes for another class.

**[0522]** As used herein, a composition refers to any mixture of two or more products, substances, or compounds, including cells. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

**[0523]** As used herein, a “subject” is a mammal, such as a human or other animal, and typically is human.

## VII. EXEMPLARY EMBODIMENTS

**[0524]** Among the provided embodiments are:

1. A DNA-targeting system comprising a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.
2. The DNA-targeting system of embodiment 1, wherein binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site.
3. The DNA-targeting system of embodiment 1 or 2, wherein the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA

(gRNA) combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

4. The DNA-targeting system of any of embodiments 1-3, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA.

5. The DNA-targeting system of embodiment 3 or 4, wherein the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein.

6. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a variant Cas protein that lacks nuclease activity or that is a deactivated Cas (dCas) protein; and

(b) at least one gRNA, comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus or is complementary to the target site.

7. The DNA-targeting system of any of embodiments 3-6, wherein the at least one gRNA is capable of complexing with the Cas protein or variant thereof.

8. The DNA-targeting system of any of embodiments 3-5 and 7, wherein the at least one gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

9. The DNA-targeting system of any of embodiments 3-8, wherein the Cas protein or a variant thereof is a Cas9 protein or a variant thereof.

10. The DNA-targeting system of any of embodiments 4-9, wherein the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein.

11. The DNA-targeting system of embodiment 9 or 10, wherein the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof.

12. The DNA-targeting system of any of embodiments 9-11, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96.

13. The DNA-targeting system of any of embodiments 9-12, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

14. The DNA-targeting system of embodiment 9 or 10, wherein the Cas9 protein or a variant thereof is a *Staphylococcus aureus* Cas9 (SaCas9) protein or a variant thereof.

15. The DNA-targeting system of any of embodiments 9, 10, and 14, wherein the variant Cas9 is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation

selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99.

16. The DNA-targeting system of any of embodiments 9, 10, 14, and 15, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

17. The DNA-targeting system of any of embodiments 1-16, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

18. The DNA-targeting system of any of embodiments 1-17, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

19. The DNA-targeting system of any of embodiments 1-18, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

20. The DNA-targeting system of any of embodiments 3-19, wherein the at least one gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

21. The DNA-targeting system of embodiment 20, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30.

22. The DNA-targeting system of any of embodiments 3-21, wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:69, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:69.

23. The DNA-targeting system of any of embodiments 1-18, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

24. The DNA-targeting system of any of embodiments 3-18 and 23, wherein the at least one gRNA comprises a gRNA that comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt.

25. The DNA-targeting system of embodiment 24, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30.

26. The DNA-targeting system of any of embodiments 3-18 and 23-25, wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:87, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:87.

27. The DNA-targeting system of any of embodiments 6-26, wherein the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length.

28. The DNA-targeting system of any of embodiments 6-27, wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

29. The DNA-targeting system of any of embodiments 3-28, wherein the gRNA comprises

modified nucleotides for increased stability.

30. The DNA-targeting system of any of embodiments 1-29, wherein the DNA-targeting system further comprises at least one effector domain.

31. The DNA-targeting system of embodiment 30, wherein the DNA-targeting domain or a component thereof is fused to the at least one effector domain.

32. The DNA-targeting system of embodiment 31, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA, and the component thereof fused to the at least one effector domain is the Cas protein or a variant thereof.

33. The DNA-targeting system of any of embodiments 30-32, wherein the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, heterochromatin formation, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation.

34. The DNA-targeting system of any of embodiments 30-33, wherein the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation.

35. The DNA-targeting system of any of embodiments 30-34, wherein the effector domain induces transcription de-repression.

36. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

- (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and
- (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

37. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

- (a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and
- (b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

38. The DNA-targeting system of any of embodiments 30-37, wherein the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof.

39. The DNA-targeting system of any of embodiments 30-38, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a

portion or a variant thereof.

40. The DNA-targeting system of embodiment 39, wherein the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

41. The DNA-targeting system of any of embodiments 30-40, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof.

42. The DNA-targeting system of any of embodiments 30-41, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

43. The DNA-targeting system of any of embodiments 39-42, wherein the DNA-targeting system comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

44. The DNA-targeting system of any of embodiments 1-43, wherein the DNA-targeting domain is a first DNA-targeting domain, and the DNA-targeting system further comprises one or more second DNA-targeting domain.

45. The DNA-targeting system of embodiment 44, wherein:

the first DNA-targeting domain binds a first target site in the MeCP2 locus; and

the second DNA-targeting domain binds a second target site in the MeCP2 locus.

46. A DNA-targeting system that binds to one or more target sites in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, the DNA-targeting system comprising:

a first DNA-targeting domain that binds a first target site in a MeCP2 locus; and

a second DNA-targeting domain that binds a second target site in a MeCP2 locus.

47. The DNA-targeting system of any of embodiments 44-47, wherein the first target site and the second target site independently are located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

48. The DNA-targeting system of any of embodiments 44-47, wherein:

the first DNA-targeting domain comprises a first Cas-gRNA combination comprising (a) a first Cas protein or a variant thereof and (b) a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site; and

the second DNA-targeting domain comprises a second Cas-gRNA combination comprising (a) a second Cas protein or a variant thereof and (b) a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site.

49. The DNA-targeting system of embodiment 48, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein.

50. The DNA-targeting system of embodiment 49, wherein the first variant Cas protein and/or

the second variant Cas protein is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96; or comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

51. The DNA-targeting system of embodiment 49, wherein the first variant Cas protein and/or the second variant Cas protein is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; or comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

52. The DNA-targeting system of any of embodiments 48-51, wherein the first Cas protein and the second Cas protein are the same.

53. The DNA-targeting system of any of embodiments 48-51, wherein the first Cas protein and the second Cas protein are different.

54. The DNA-targeting system of any of embodiments 48-53, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is fused to at least one effector domain.

55. The DNA-targeting system of embodiment 54, wherein the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, heterochromatin formation, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation.

56. The DNA-targeting system of embodiment 54 or 55, wherein the effector domain induces transcription de-repression.

57. The DNA-targeting system of any of embodiments 44-56, wherein the first DNA-targeting domain and the second DNA-targeting domain are encoded in a first polynucleotide.

58. The DNA-targeting system of any of embodiments 44-57, wherein the first Cas protein and the second Cas protein are encoded in a first polynucleotide.

59. The DNA-targeting system of any of embodiments 44-52 and 54-58, wherein the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence.

60. The DNA-targeting system of any of embodiments 44-59, wherein the first gRNA and the second gRNA are encoded in a first polynucleotide.

61. The DNA-targeting system of any of embodiments 44-52 and 54-60, wherein the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence, and the Cas protein, the first gRNA, and the second gRNA are encoded in a first polynucleotide.

62. The DNA-targeting system of any of embodiments 44-56, wherein the first DNA-targeting



domain is encoded in a first polynucleotide and the second DNA-targeting domain is encoded in a second polynucleotide.

63. The DNA-targeting system of any of embodiments 44-56 and 62, wherein the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide.

64. The DNA-targeting system of any of embodiments 44-56, 62, and 63, wherein the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide.

65. The DNA-targeting system of any of embodiments 44-56, 62, and 63, wherein the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

66. A guide RNA (gRNA) that binds a target site located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

67. The gRNA of embodiment 66, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

68. The gRNA of embodiment 66 or 67, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

69. The gRNA of any of embodiments 66-68, wherein the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

70. The gRNA of any of embodiments 66-68, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30.

71. The gRNA of any of embodiments 66-69, wherein the gRNA comprises the sequence set forth in SEQ ID NO:69, optionally wherein the gRNA sequence is set forth in SEQ ID NO:69.

72. The gRNA of embodiment 66 or 67, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

73. The gRNA of any of embodiments 66, 67, and 72, wherein the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt.

74. The gRNA of any of embodiments 66, 67, 72, and 73, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30.

75. The gRNA of any of embodiments 66, 67, and 72-74, wherein the gRNA comprises the sequence set forth in SEQ ID NO:87, optionally wherein the gRNA sequence is set forth in SEQ ID NO:87.

76. The gRNA of any of embodiments 66-75, wherein the gRNA spacer sequence is between 14

nt and 24 nt, or between 16 nt and 22 nt in length.

77. The gRNA of any of embodiments 66-76, wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

78. The gRNA of any of embodiments 66-77, wherein the gRNA comprises modified nucleotides for increased stability.

79. The gRNA of any of embodiments 66-78, wherein the gRNA is capable of complexing with the Cas protein or variant thereof.

80. The gRNA of any of embodiments 66-79, wherein the gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

81. A combination, comprising a first gRNA comprising the gRNA of any of embodiments 66-80, and one or more second gRNAs that binds to a second target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

82. The combination of embodiment 81, wherein the second gRNA comprises the gRNA of any of embodiments 66-80.

83. A combination, comprising:

a first gRNA that binds a first target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158; and

a second gRNA that binds a second target site in a regulatory DNA element of a MeCP2 locus, wherein the second target site is located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

84. A fusion protein comprising (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein:

the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and

the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, heterochromatin formation, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation.

85. The fusion protein of embodiment 84, wherein binding of the DNA-targeting domain or a component thereof to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

86. The fusion protein of embodiment 84 or 85, wherein the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA)

combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

87. The fusion protein of any of embodiments 84-86, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising a Cas protein or a variant thereof and at least one gRNA, and the component of the DNA-targeting domain is a Cas protein or a variant thereof.

88. A fusion protein comprising (1) a Cas protein or a variant thereof and (2) at least one effector domain, wherein the effector domain induces, catalyzes or leads to transcription activation, transcription co-activation, transcription elongation, transcription de-repression, transcription repression, transcription factor release, polymerization, histone modification, histone acetylation, histone deacetylation, nucleosome remodeling, chromatin remodeling, heterochromatin formation, reversal of heterochromatin formation, nuclease, signal transduction, proteolysis, ubiquitination, deubiquitination, phosphorylation, dephosphorylation, splicing, nucleic acid association, DNA methylation, DNA demethylation, histone methylation, histone demethylation, or DNA base oxidation.

89. The fusion protein of any of embodiments 86-88, wherein the variant Cas protein lacks nuclease activity or is a deactivated Cas (dCas) protein.

90. The fusion protein of any of embodiments 86-89, wherein the Cas protein or a variant thereof is a Cas9 protein or a variant thereof.

91. The fusion protein of any of embodiments 86-90, wherein the variant Cas protein is a variant Cas9 protein that lacks nuclease activity or that is a deactivated Cas9 (dCas9) protein.

92. The fusion protein of embodiment 90 or 91, wherein the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof.

93. The fusion protein of any of embodiments 90-92, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96.

94. The fusion protein of any of embodiments 90-93, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

95. The fusion protein of embodiment 90 or 91, wherein the Cas9 protein or a variant thereof is a *Streptococcus pyogenes* Cas9 (SaCas9) protein or a variant thereof.

96. The fusion protein of any of embodiments 90, 91, and 95, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99.

97. The fusion protein of any of embodiments 90, 91, 95, and 96, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

98. The fusion protein of any of embodiments 84-97, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.
99. The fusion protein of any of embodiments 84-98, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.
100. The fusion protein of any of embodiments 84-99, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.
101. The fusion protein of any of embodiments 84-99, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.
102. The fusion protein of any of embodiments 84-101, wherein the effector domain induces, catalyzes or leads to transcription de-repression, DNA demethylation or DNA base oxidation.
103. The fusion protein of any of embodiments 84-102, wherein the effector domain induces transcription de-repression.
104. The fusion protein of any of embodiments 84-103, wherein the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof.
105. The fusion protein of any of embodiments 84-104, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof.
106. The fusion protein of embodiment 105, wherein the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.
107. The fusion protein of any of embodiments 84-106, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof, optionally wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus of *the Cas protein or a variant thereof*.
108. The fusion protein of any of embodiments 84-107, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof, optionally the Cas protein or variant thereof, to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).
109. The fusion protein of any of embodiments 84-108, wherein the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.
110. A combination comprising the fusion protein of any of embodiments 85-109 and at least one

gRNA, optionally wherein the at least one gRNA is a gRNA of any of embodiments 66-80.

111. A polynucleotide encoding the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, or the fusion protein of any of embodiments 84-109, or a portion or a component of any of the foregoing.

112. A polynucleotide encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of the DNA-targeting system of any of embodiments 56-90 or the combination of any of embodiments 80-83 and 110.

113. A polynucleotide encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of the DNA-targeting system of any of embodiments 56-90 or the combination of any of embodiments 80-83 and 110.

114. A plurality of polynucleotides, comprising the polynucleotide of any of embodiments 111-113, and one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, or the fusion protein of any of embodiments 84-109, or a portion or a component of any of the foregoing.

115. A plurality of polynucleotides, comprising:

a first polynucleotide comprising the polynucleotide of embodiment 112; and

a second polynucleotide comprising the polynucleotide of embodiment 113.

116. A vector comprising the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, or a first polynucleotide or a second polynucleotide of the plurality of polynucleotides of embodiment 114 or 115, or a portion or a component of any of the foregoing.

117. The vector of embodiment 116, wherein the vector is a viral vector, optionally wherein the viral vector is an AAV vector.

118. The vector of embodiment 117, wherein the viral vector, optionally the AAV vector, exhibits central nervous system (CNS) tropism.

119. The vector of embodiment 117 or 118, wherein the viral vector is an AAV vector and the AAV vector is selected from among AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 vector, optionally an AAV5 vector or an AAV9 vector.

120. The vector of embodiment 116, wherein the vector is a non-viral vector selected from: a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide

121. A plurality of vectors, comprising the vector of any of embodiments 116-120, and one or more additional vectors comprising one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, or the fusion protein of any of embodiments 84-109, or a portion or a component of any of the foregoing.

122. A plurality of vectors, comprising:

a first vector comprising the polynucleotide of embodiment 112; and  
a second vector comprising the polynucleotide of embodiment 113.

123. A cell comprising the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, the fusion protein of any of embodiments 84-109, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 116-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing.

124. The cell of embodiment 123, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

125. The cell of embodiment 123 or 124, wherein the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

126. The cell of any of embodiments 123-125, wherein the cell is from a subject that has or is suspected of having Rett syndrome.

127. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a cell, the method comprising:

introducing the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, the fusion protein of any of embodiments 84-109, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 116-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing, into the cell.

128. The method of embodiment 127, wherein the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

129. The method of embodiment 127 or 128, wherein the cell is from a subject that has or is suspected of having Rett syndrome.

130. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a subject, the method comprising:

administering the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, the fusion protein of any of embodiments 84-109, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 116-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing, to the subject.

131. The method of any of embodiments 128-130, wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

132. The method of any of embodiments 128-131, wherein the subject has or is suspected of

having Rett syndrome.

133. A method of treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, the method comprising:

administering the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, the fusion protein of any of embodiments 84-109, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 116-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

134. A method of treating Rett syndrome, the method comprising:

administering the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83 and 110, the fusion protein of any of embodiments 84-109, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 116-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

135. The method of any of embodiments 128-134, wherein a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, optionally wherein the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome.

136. The method of any of embodiments 128-135, wherein a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

137. The method of any of embodiments 127-136, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

138. The method of any of embodiments 127-137, wherein the introducing, contacting or administering is carried out in vivo or ex vivo.

139. The method of any of embodiments 135-138, wherein following the introducing, contacting or administering, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject.

140. The method of embodiment 139, wherein the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold.

141. The method of embodiment 139 or 140, wherein the expression is increased by less than about 200-fold, 150-fold, or 100-fold.

142. The method of any of embodiments 139-141, wherein the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the

expression of the wild-type MeCP2 of a cell from a normal subject.

143. The method of any of embodiments 128-143, wherein the subject is a human.

144. A pharmaceutical composition comprising the DNA-targeting system of any of embodiments 1-65, the gRNA of any of embodiments 66-80, the combination of any of embodiments 80-83, the fusion protein of any of embodiments 84-109 or 110, the polynucleotide of any of embodiments 111-113, the plurality of polynucleotides of embodiment 114 or 115, the vector of any of embodiments 115-120, the plurality of vectors of embodiment 121 or 122, or a portion or a component of any of the foregoing.

145. The pharmaceutical composition of embodiment 144, for use in treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

146. The pharmaceutical composition of embodiment 144, for use in treating Rett syndrome.

147. The pharmaceutical composition of embodiment 183, for use in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

148. The pharmaceutical composition of embodiment 183, for use in the manufacture of a medicament for treating Rett syndrome.

149. The pharmaceutical composition for use of any of embodiments 145-148, wherein the pharmaceutical composition is to be administered to a subject, optionally wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally Rett syndrome.

150. Use of the pharmaceutical composition of embodiment 144, for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

151. Use of the pharmaceutical composition of embodiment 144, for treating Rett syndrome.

152. Use of the pharmaceutical composition of embodiment 144 in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

153. Use of the pharmaceutical composition of embodiment 144 in the manufacture of a medicament for treating Rett syndrome.

154. The use of embodiment 148 or 149, wherein the pharmaceutical composition is to be administered to a subject, optionally wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally Rett syndrome.

155. The pharmaceutical composition for use or the use of embodiment 149 or 154, wherein a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, optionally wherein the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome.

156. The pharmaceutical composition for use or the use of any of embodiments 149, 154, and



155, wherein a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

157. The pharmaceutical composition for use or the use of any of embodiments 155-156, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

158. The pharmaceutical composition for use or the use of any of embodiments 149 and 154-157, wherein the administration is carried out in vivo or ex vivo.

159. The pharmaceutical composition for use or the use of any of embodiments 149 and 154-158, wherein following the administration, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject.

160. The pharmaceutical composition for use or the use of embodiment 160, wherein the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 7.5-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold.

161. The pharmaceutical composition for use or the use of embodiment 159 or 160, wherein the expression is increased by less than about 200-fold, 150-fold, or 100-fold.

162. The pharmaceutical composition for use or the use of any of embodiments 159-161, wherein the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

163. The pharmaceutical composition for use or the use of any of embodiments 149 and 154-162, wherein the subject is a human.

201. A DNA-targeting system comprising a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

202. A DNA-targeting system comprising:

(a) a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and

(b) at least one effector domain that increases transcription of the MeCP2 locus.

203. The DNA-targeting system of embodiment 201 or 202, wherein binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

204. The DNA-targeting system of any of embodiments 201-203, wherein the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

205. The DNA-targeting system of any of embodiments 201-204, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA.

206. The DNA-targeting system of embodiment 204 or 205, wherein the variant Cas protein is a deactivated Cas (dCas) protein.

207. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a deactivated Cas (dCas) protein; and

(b) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus or is complementary to the target site.

208. The DNA-targeting system of any of embodiments 204-207, wherein the at least one gRNA is capable of complexing with the Cas protein or variant thereof.

209. The DNA-targeting system of any of embodiments 204-206 and 208, wherein the at least one gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

210. The DNA-targeting system of any of embodiments 204-209, wherein the Cas protein or a variant thereof is a Cas9 protein or a variant thereof.

211. The DNA-targeting system of any of embodiments 205-210, wherein the variant Cas protein is a deactivated Cas9 (dCas9) protein.

212. The DNA-targeting system of embodiment 210 or 211, wherein the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof.

213. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a *Streptococcus pyogenes* dCas9 (dSpCas9) protein;

(b) at least one effector domain that increases transcription of a methyl-CpG-binding protein 2 (MeCP2) locus; and

(c) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a MeCP2 locus or is complementary to the target site.

214. The DNA-targeting system of any of embodiments 210-212, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96.

215. The DNA-targeting system of any of embodiments 210-212 and 214, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

216. The DNA-targeting system of embodiment 210 or 211, wherein the Cas9 protein or a variant thereof is a *Staphylococcus aureus* Cas9 (SaCas9) protein or a variant thereof.

217. The DNA-targeting system of any of embodiments 210, 211, and 216, wherein the variant Cas9 is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99.

218. The DNA-targeting system of any of embodiments 210, 211, 216, and 217, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

219. The DNA-targeting system of any of embodiments 210-212, wherein the variant Cas protein is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein.

220. The DNA-targeting system of embodiment 219, wherein when the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein.

221. The DNA-targeting system of embodiment 219 or 220, wherein the N-terminal Intein comprises an N-terminal Npu Intcin, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

222. The DNA-targeting system of any of embodiments 219-221, wherein the N-terminal fragment of the variant Cas protein comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

223. The DNA-targeting system of any of embodiments 219-222, wherein the first polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

224. The DNA-targeting system of any of embodiments 219-223, wherein the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

225. The DNA-targeting system of any of embodiments 219-224, wherein the C-terminal fragment of the variant Cas protein comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

226. The DNA-targeting system of any of embodiments 219-225, wherein the second polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

227. The DNA-targeting system of any of embodiments 201-226, wherein the target site comprises the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

228. The DNA-targeting system of any of embodiments 201-227, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

229. The DNA-targeting system of any of embodiments 201-228, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

230. The DNA-targeting system of any of embodiments 201-229, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

231. The DNA-targeting system of any of embodiments 204-230, wherein the at least one gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

232. The DNA-targeting system of embodiment 231, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30.

233. The DNA-targeting system of any of embodiments 204-232, wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:69, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:69.

234. The DNA-targeting system of any of embodiments 201-229, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

235. The DNA-targeting system of any of embodiments 204-229 and 234, wherein the at least one gRNA comprises a gRNA that comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt.

236. The DNA-targeting system of embodiment 235, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30.

237. The DNA-targeting system of any of embodiments 204-229 and 234-236, wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:87, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:87.

238. The DNA-targeting system of any of embodiments 207-237, wherein the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length.

239. The DNA-targeting system of any of embodiments 207-238, wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

240. The DNA-targeting system of any of embodiments 204-239, wherein the gRNA comprises modified nucleotides for increased stability.

241. The DNA-targeting system of any of embodiments 201-240, wherein the DNA-targeting system further comprises at least one effector domain.

242. The DNA-targeting system of embodiment 241, wherein the DNA-targeting domain or a component thereof is fused to the at least one effector domain.

243. The DNA-targeting system of embodiment 242, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA, and the component thereof fused to the at least one effector domain is the Cas protein or a variant thereof.

244. The DNA-targeting system of any of embodiments 241-243, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

245. The DNA-targeting system of any of embodiments 241-244, wherein the effector domain induces transcription de-repression, DNA demethylation or DNA base oxidation.

246. The DNA-targeting system of any of embodiments 241-245, wherein the effector domain induces transcription de-repression.

247. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

248. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

249. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a

*Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

250. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

251. The DNA-targeting system of embodiment 249 or 250, further comprising a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of the dSpCas9 fused to a C-terminal Intein.

252. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an C-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

253. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to a C-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

254. The DNA-targeting system of embodiment 252 or 253, further comprising a first polypeptide of a split variant Cas9 protein an N-terminal fragment of the dSpCas9 fused to an N-terminal Intein.

255. The DNA-targeting system of any of embodiments 249-254, wherein when the first polypeptide and the second polypeptide of the split variant Cas9 are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

256. The DNA-targeting system of embodiment 255, wherein the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has

at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

257. The DNA-targeting system of embodiment 255 or 256, wherein the N-terminal fragment of the variant Cas9 comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

258. The DNA-targeting system of any of embodiments 255-257, wherein the first polypeptide of the split variant Cas9 comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

259. The DNA-targeting system of any of embodiments 255-258, wherein the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

260. The DNA-targeting system of any of embodiments 255-259, wherein the C-terminal fragment of the variant Cas9 comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

261. The DNA-targeting system of any of embodiments 255-260, wherein the second polypeptide of the split variant Cas9 comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

262. The DNA-targeting system of any of embodiments 241-248, wherein the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof.

263. The DNA-targeting system of any of embodiments 241-248 and 262, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof.

264. The DNA-targeting system of embodiment 263, wherein the effector domain comprises the

sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

265. The DNA-targeting system of any of embodiments 241-248 and 262-264, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof.

266. The DNA-targeting system of any of embodiments 241-248 and 262-265, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

267. The DNA-targeting system of any of embodiments 263-266, wherein the DNA-targeting system comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

268. A combination, comprising:

a first DNA-targeting domain comprising the DNA targeting domain of any of embodiments 201-267, and

one or more second DNA-targeting domains, optionally wherein the one or more second DNA-targeting domains comprises the DNA targeting domain of any of embodiments 201-267.

269. The combination of embodiment 268, wherein:

the first DNA-targeting domain binds a first target site in the MeCP2 locus; and

the second DNA-targeting domain binds a second target site in the MeCP2 locus.

270. A combination comprising:

a first DNA-targeting domain that binds a first target site in a MeCP2 locus; and

a second DNA-targeting domain that binds a second target site in a MeCP2 locus.

271. The combination of any of embodiments 268-270, wherein the first target site and the second target site independently are located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

272. The combination of any of embodiments 268-271, wherein:

the first DNA-targeting domain comprises a first Cas-gRNA combination comprising (a) a first Cas protein or a variant thereof and (b) a first gRNA that is capable of hybridizing to the target site or is complementary to the first target site; and

the second DNA-targeting domain comprises a second Cas-gRNA combination comprising (a) a second Cas protein or a variant thereof and (b) a second gRNA that is capable of hybridizing to the target site or is complementary to the second target site.

273. The combination of embodiment 272, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a deactivated Cas9 (dCas9) protein.

274. The combination of embodiment 273, wherein the first variant Cas protein and/or the second variant Cas protein is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of



SEQ ID NO:96; or comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

275. The combination of embodiment 273, wherein the first variant Cas protein and/or the second variant Cas protein is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; or comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

276. The combination of any of embodiments 272-275, wherein the first variant Cas protein and/or the second variant Cas protein is a split variant Cas9 protein, wherein the split variant Cas9 protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas9 and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas9 and a C-terminal Intein.

277. The combination of any of embodiments 272-276, wherein the first Cas protein and the second Cas protein are the same.

278. The combination of any of embodiments 272-276, wherein the first Cas protein and the second Cas protein are different.

279. The combination of any of embodiments 272-278, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is fused to at least one effector domain.

280. The combination of embodiment 279, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

281. The combination of embodiment 279 or 280, wherein the effector domain induces transcription de-repression.

282. The combination of any of embodiments 268-281, wherein the first DNA-targeting domain and the second DNA-targeting domain are encoded in a first polynucleotide.

283. The combination of any of embodiments 268-282, wherein the first Cas protein and the second Cas protein are encoded in a first polynucleotide.

284. The combination of any of embodiments 268-277 and 279-283, wherein the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence.

285. The combination of any of embodiments 268-284, wherein the first gRNA and the second gRNA are encoded in a first polynucleotide.

286. The combination of any of embodiments 268-277 and 279-285, wherein the first Cas protein and the second Cas protein are encoded by the same nucleotide sequence, and the Cas protein, the first gRNA, and the second gRNA are encoded in a first polynucleotide.

287. The combination of any of embodiments 268-281, wherein the first DNA-targeting domain is encoded in a first polynucleotide and the second DNA-targeting domain is encoded in a second

polynucleotide.

288. The combination of any of embodiments 268-281 and 287, wherein the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide.

289. The combination of any of embodiments 268-281, 287, and 288, wherein the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide.

290. The combination of any of embodiments 268-281, 287, and 288, wherein the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

291. A guide RNA (gRNA) that binds a target site located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

292. A guide RNA (gRNA) that binds a target site comprising the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

293. The gRNA of embodiment 291 or 292, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

294. The gRNA of any of embodiments 291-293, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

295. The gRNA of any of embodiments 291-294, wherein the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

296. The gRNA of any of embodiments 291-294, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30.

297. The gRNA of any of embodiments 291-295, wherein the gRNA comprises the sequence set forth in SEQ ID NO:69, optionally wherein the gRNA sequence is set forth in SEQ ID NO:69.

298. The gRNA of any of embodiments 291-293, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

299. The gRNA of any of embodiments 291-293 and 298, wherein the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt.

300. The gRNA of any of embodiments 291-293, 298, and 299, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30.

301. The gRNA of any of embodiments 291-293 and 298-3100, wherein the gRNA comprises the sequence set forth in SEQ ID NO:87, optionally wherein the gRNA is set forth in SEQ ID NO:87.

302. The gRNA of any of embodiments 291-301, wherein the gRNA spacer sequence is between

14 nt and 24 nt, or between 16 nt and 22 nt in length.

303. The gRNA of any of embodiments 291-302, wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

304. The gRNA of any of embodiments 291-303, wherein the gRNA comprises modified nucleotides for increased stability.

305. The gRNA of any of embodiments 291-304, wherein the gRNA is capable of complexing with the Cas protein or variant thereof.

306. The gRNA of any of embodiments 291-305, wherein the gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

307. A combination, comprising a first gRNA comprising the gRNA of any of embodiments 291-306, and one or more second gRNAs that binds to a second target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

308. The combination of embodiment 307, wherein the second gRNA comprises the gRNA of any of embodiments 266-280.

309. A combination, comprising:

a first gRNA that binds a first target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158; and

a second gRNA that binds a second target site in a regulatory DNA element of a MeCP2 locus, wherein the second target site is located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

310. A fusion protein comprising (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein:

the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and

the effector domain increases transcription of the MeCP2 locus.

311. A fusion protein comprising (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein:

the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and

the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

312. The fusion protein of embodiment 310 or 311, wherein the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing

endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

313. The fusion protein of any of embodiments 310-312, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising a Cas protein or a variant thereof and at least one gRNA, and the component of the DNA-targeting domain is a Cas protein or a variant thereof.

314. The fusion protein of embodiment 313, wherein the gRNA binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

315. A fusion protein comprising (1) a Cas protein or a variant thereof and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

316. A fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

317. A fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

318. The fusion protein of embodiment 317, wherein the first polypeptide of the split variant Cas protein, and a second polypeptide of the split variant Cas protein comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

319. A fusion protein comprising (1) a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

320. A fusion protein comprising (1) a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

321. The fusion protein of embodiment 320, wherein the second polypeptide of the split variant Cas protein, and a first polypeptide of the split variant Cas protein comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excise and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

322. The fusion protein of any of embodiments 312-321, wherein the Cas protein or a variant

thereof is capable of complexing with at least one gRNA, optionally wherein the gRNA binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

323. The fusion protein of any of embodiments 310-322, wherein binding of the DNA-targeting domain or a component thereof targeted to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

324. The fusion protein of any of embodiments 312-323, wherein the variant Cas protein is a deactivated Cas (dCas) protein.

325. The fusion protein of any of embodiments 312-324, wherein the Cas protein or a variant thereof is a Cas9 protein or a variant thereof.

326. The fusion protein of any of embodiments 312-325, wherein the variant Cas protein is a deactivated Cas9 (dCas9) protein.

327. The fusion protein of embodiment 325 or 326, wherein the Cas9 protein or variant thereof is a *Streptococcus pyogenes* Cas9 (SpCas9) protein or a variant thereof.

328. The fusion protein of any of embodiments 325-327, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96.

329. The fusion protein of any of embodiments 325-328, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

330. The fusion protein of embodiment 325 or 326, wherein the Cas9 protein or a variant thereof is a *Streptococcus pyogenes* Cas9 (SaCas9) protein or a variant thereof.

331. The fusion protein of any of embodiments 325, 326, and 330, wherein the variant Cas9 is a *Streptococcus pyogenes* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99.

332. The fusion protein of any of embodiments 325, 326, 330, and 331, wherein the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

333. The fusion protein of any of embodiments 312-326, wherein the variant Cas protein is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein.

334. The fusion protein of embodiment 333, wherein when the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein.

335. The fusion protein of embodiment 333 or 334, wherein the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has

at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

336. The fusion protein of any of embodiments 333-335, wherein the N-terminal fragment of the variant Cas protein comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

337. The fusion protein of any of embodiments 333-336, wherein the first polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:121, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

338. The fusion protein of any of embodiments 333-337, wherein the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

339. The fusion protein of any of embodiments 333-338, wherein the C-terminal fragment of the variant Cas protein comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

340. The fusion protein of any of embodiments 333-339, wherein the second polypeptide of the split variant Cas protein comprises the sequence set forth in SEQ ID NO:131, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

341. The fusion protein of any of embodiments 310-340, wherein the target site comprises the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

342. The fusion protein of any of embodiments 310-341, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

343. The fusion protein of any of embodiments 310-342, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a

complementary sequence of any of the foregoing.

344. The fusion protein of any of embodiments 310-343, wherein the target site comprises the sequence set forth in SEQ ID NO:9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

345. The fusion protein of any of embodiments 310-343, wherein the target site comprises the sequence set forth in SEQ ID NO:27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

346. The fusion protein of any of embodiments 310-345, wherein the effector domain induces transcription de-repression, DNA demethylation or DNA base oxidation.

347. The fusion protein of any of embodiments 310-346, wherein the effector domain induces transcription de-repression.

348. The fusion protein of any of embodiments 310-347, wherein the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof.

349. The fusion protein of any of embodiments 310-348, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof.

350. The fusion protein of embodiment 349, wherein the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

351. The fusion protein of any of embodiments 310-350, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof, optionally wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus of *the Cas protein or a variant thereof*.

352. The fusion protein of any of embodiments 310-351, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof, optionally the Cas protein or variant thereof, to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

353. The fusion protein of any of embodiments 310-352, wherein the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

354. A combination comprising the fusion protein of any of embodiments 310-353 and at least one gRNA, optionally wherein the at least one gRNA is a gRNA of any of embodiments 266-280.

355. A polynucleotide encoding the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 91-306, the combination of any of embodiments 268-290, 307-309, and 354, or the fusion protein of any of embodiments 310-353, or a portion or a component of any of the

foregoing.

356. A polynucleotide encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of the DNA-targeting system of any of embodiments 201-267 or the combination of any of embodiments 268-290, 307-309, and 354.

357. A polynucleotide encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of the DNA-targeting system of any of embodiments 201-267 or the combination of any of embodiments 268-290, 307-309, and 354.

358. A plurality of polynucleotides, comprising the polynucleotide of any of embodiments 155-157, and one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, or the fusion protein of any of embodiments 310-353, or a portion or a component of any of the foregoing.

359. A plurality of polynucleotides, comprising:

a first polynucleotide comprising the polynucleotide of embodiment 356; and

a second polynucleotide comprising the polynucleotide of embodiment 357.

360. A vector comprising the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, or a first polynucleotide or a second polynucleotide of the plurality of polynucleotides of embodiment 358 or 359, or a portion or a component of any of the foregoing.

361. The vector of embodiment 360, wherein the vector is a viral vector, optionally wherein the viral vector is an AAV vector.

362. The vector of embodiment 361, wherein the viral vector, optionally the AAV vector, exhibits tropism for a cell of the central nervous system (CNS), a heart cell, optionally a cardiomyocyte, a skeletal muscle cell, a fibroblast, an induced pluripotent stem cell, or a cell derived from any of the foregoing.

363. The vector of embodiment 361 or 362, wherein the viral vector is an AAV vector and the AAV vector is selected from among AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 vector, optionally an AAV5 vector or an AAV9 vector.

364. The vector of any of embodiments 361-363, wherein the viral vector is an AAV9 vector.

365. The vector of embodiment 360, wherein the vector is a non-viral vector selected from: a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide.

366. A plurality of vectors, comprising the vector of any of embodiments 360-365, and one or more additional vectors comprising one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, or the fusion protein of any of embodiments 310-353, or a portion or a component of any of the foregoing.



367. A plurality of vectors, comprising:

a first vector comprising the polynucleotide of embodiment 356; and

a second vector comprising the polynucleotide of embodiment 357.

368. A cell comprising the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing.

369. The cell of embodiment 368, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

370. The cell of embodiment 368 or 369, wherein the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

371. The cell of any of embodiments 368-370, wherein the cell is from a subject that has or is suspected of having Rett syndrome.

372. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a cell, the method comprising:

introducing the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing, into the cell.

373. The method of embodiment 372, wherein the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

374. The method of embodiment 372 or 373, wherein the cell is from a subject that has or is suspected of having Rett syndrome.

375. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a subject, the method comprising:

administering the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing, to the subject.

376. The method of any of embodiments 373-375, wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

377. The method of any of embodiments 373-376, wherein the subject has or is suspected of having Rett syndrome.

378. A method of treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, the method comprising:

administering the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

379. A method of treating Rett syndrome, the method comprising:

administering the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

380. The method of any of embodiments 373-379, wherein a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, optionally wherein the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome.

381. The method of any of embodiments 373-380, wherein a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

382. The method of any of embodiments 372-381, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

383. The method of any of embodiments 372-382, wherein the introducing, contacting or administering is carried out in vivo or ex vivo.

384. The method of any of embodiments 380-383, wherein following the introducing, contacting or administering, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject.

385. The method of embodiment 384, wherein the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold.

386. The method of embodiment 384 or 385, wherein the expression is increased by less than about 200-fold, 150-fold, or 100-fold.

387. The method of any of embodiments 384-386, wherein the expression of the wild-type

MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

388. The method of any of embodiments 373-387, wherein the subject is a human.

389. A pharmaceutical composition comprising the DNA-targeting system of any of embodiments 201-267, the gRNA of any of embodiments 291-306, the combination of any of embodiments 268-290, 307-309, and 354, the fusion protein of any of embodiments 310-353, the polynucleotide of any of embodiments 355-357, the plurality of polynucleotides of embodiment 358 or 359, the vector of any of embodiments 360-365, the plurality of vectors of embodiment 366 or 367, or a portion or a component of any of the foregoing.

390. The pharmaceutical composition of embodiment 389, for use in treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

391. The pharmaceutical composition of embodiment 389 or 390, for use in treating Rett syndrome.

392. The pharmaceutical composition of embodiment 389, for use in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

393. The pharmaceutical composition of embodiment 389 or 392, for use in the manufacture of a medicament for treating Rett syndrome.

394. The pharmaceutical composition for use of any of embodiments 391-393, wherein the pharmaceutical composition is to be administered to a subject, optionally wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally Rett syndrome.

395. Use of the pharmaceutical composition of embodiment 389, for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

396. Use of the pharmaceutical composition of embodiment 389 or 395, for treating Rett syndrome.

397. Use of the pharmaceutical composition of embodiment 389 in the manufacture of a medicament for treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

398. Use of the pharmaceutical composition of embodiment 389 or 397 in the manufacture of a medicament for treating Rett syndrome.

399. The use of any of embodiments 395-398, wherein the pharmaceutical composition is to be administered to a subject, optionally wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally Rett syndrome.

400. The pharmaceutical composition for use or the use of any of embodiments 390-399, wherein a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, optionally wherein

the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome.

401. The pharmaceutical composition for use or the use of any of embodiments 390-400, wherein a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

402. The pharmaceutical composition for use or the use of any of embodiments 390-401, wherein the cell is a nervous system cell, or an induced pluripotent stem cell.

403. The pharmaceutical composition for use or the use of any of embodiments 390-402, wherein the administration is carried out in vivo or ex vivo.

404. The pharmaceutical composition for use or the use of any of embodiments 390-403, wherein following the administration, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject.

405. The pharmaceutical composition for use or the use of embodiment 404, wherein the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 7.5-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold.

406. The pharmaceutical composition for use or the use of embodiment 404 or 405, wherein the expression is increased by less than about 200-fold, 150-fold, or 100-fold.

407. The pharmaceutical composition for use or the use of any of embodiments 404-406, wherein the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

408. The pharmaceutical composition for use or the use of any of embodiments 390-407, wherein the subject is a human.

## VIII. EXAMPLES

[0525] The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

### **Example 1: CRISPR/Cas-effector fusion protein-mediated transcriptional activation of MeCP2 in induced pluripotent stem cells (iPSCs) generated from Rett syndrome patient**

[0526] Transcriptional re-activation (de-repression) of the methyl-CpG-binding protein 2 (MeCP2) allele from the inactive X chromosome in Rett syndrome patient-derived cells by MeCP2-targeting CRISPR/Cas-effector fusion protein was assessed.

[0527] Guide RNAs (gRNAs) targeting the promoter and first exon of human methyl-CpG-binding protein 2 (MeCP2) gene were generated, and transduced together with nucleic acid sequences encoding a deactivated Cas9 (dCas9)-TET catalytic domain fusion protein into

induced pluripotent stem cells (iPSCs) generated from a patient with Rett syndrome. Expression of mutant and wild-type alleles of MeCP2 were assessed by RT-qPCR and flow cytometry.

#### A. iPSCs generated from Rett Syndrome patients

**[0528]** Experiments were performed in iPSCs generated from a Rett syndrome patient, harboring one nonsense mutation allele of MeCP2 (R255X) on the X-chromosome (R255X-iPSCs). In this cell line, the wild-type (WT) allele was present on the inactive X chromosome (Xi), and the R255X mutant allele was present on the active X chromosome (Xa), as shown in **FIG. 1A**.

#### B. dSpCas9-TET1 and gRNA constructs

**[0529]** Plasmids encoding an exemplary deactivated Cas9 (dCas9)-TET catalytic domain fusion protein, dSpCas9-TET1 (amino acid sequence set forth in SEQ ID NO:91) were prepared. dSpCas9-TET1 included a fusion of a modified Cas9 engineered to lack endonuclease activity (dCas9) from *S. pyogenes* (dSpCas9) and the catalytic domain of Ten-eleven translocation methylcytosine dioxygenase 1 (TET1).

**[0530]** Plasmids encoding gRNAs targeted to one of multiple sequences in the human MeCP2 gene promoter and first exon, were also prepared. The gRNAs included a DNA-targeting spacer sequence and a constant scaffold sequence. gRNAs were designed based on the SpCas9 protospacer-adjacent motif (PAM) sequence, 5'-NGG-3'. The MeCP2-targeting gRNAs are indicated in **Table E1**.

**Table E1.** MeCP2-targeting gRNAs

SEQ ID NO:	gRNA name	gRNA spacer sequence
1	gRNA 1	TAAGGATTAATGGACCCTTG
2	gRNA 2	CGCCTCTTTTCCCCAAACGA
3	gRNA 3	CCATCACAGCCAATGACGGG
4	gRNA 4	TCGGTGCATCTGTGGACAGA
5	gRNA 5	GGGGCGCGACGTCGGCCGTG
6	gRNA 6	TCAGCGGCGATGCCGTCAAT
7	gRNA 7	CGCGGAGGGACTGGTTTAGT
8	gRNA 8	AAGAGGGCGGGGCGCGACGT
9	gRNA 9	GCTGCGAGCCCCGCCGTCAT
10	gRNA 10	ACTTGCCCCAGCATCCGCAA
11	gRNA 11	CCGTTACTCGGCCCCCCCAC
12	gRNA 12	GCGCCCCCTCTCCCGTACT
13	gRNA 13	GGAGGGGGAGAGCGCGATCC
14	gRNA 14	TCGCCGGGGCTTCGCCTGTC
15	gRNA 15	AGGCGAAGCCCCGGCGACAG

16	gRNA 16	CCGGGATGCGCGTCGAGGGC
17	gRNA 17	GGACGGTCACCCGCGAGCAG
18	gRNA 18	CGGCCCGTCACCCCTGCTCG
19	gRNA 19	GCCGAGGGGAGAGTCGCCAC
20	gRNA 20	CGGACGACACGGCTGGCGGA
21	gRNA 21	GCCGTGTCGTCCGACCCCGC
22	gRNA 22	CTGACCCCCGCCCCCGGCA
23	gRNA 23	CGGGCGGGGACCCTTGCCGG
24	gRNA 24	GACTGTGAGTGGGACCGCCG
25	gRNA 25	GGTAAAAGCCGTCCGGAAAA
26	gRNA 26	GCGCGCGCGCTCCCTCCTCT
27	gRNA 27	AATGACGGGCGGGCTCGCAG
28	gRNA 28	TTTCCTGTCCATTTCGGCCA
29	gRNA 29	GCTGCTTTCGGCCGTCGTTT

### C. Upregulation of WT allele of MeCP2 on the inactive X chromosome in R255X-iPSCs

[0531] Individual gRNAs targeted to the promoter and first exon of MeCP2, gRNAs 1-10, as described in **Table E1** above, were co-expressed with dSpCas9-TET1 in R255X-iPSCs, transduced using two separate lentiviral vectors.

[0532] Plasmids were prepared using the QIAGEN Plasmid Plus Midi Kit (#12945). Lentivirus was generated in HEK293FT cells using Lipofectamine 3000 (ThermoFisher #L3000015) and concentrated to 50x using Lenti-X Concentrator (Takara #631232). Lentivirus was added to R255X iPSCs at a 1x final concentration. 48 hours after the addition of lentivirus, the cells were selected with 0.5  $\mu$ g/ml puromycin to enrich for cells expressing dSpCas9-TET1 and gRNA. Cells were harvested at day 10 post-transduction for RT-qPCR and flow cytometry.

[0533] Levels of mRNA expression of the mutant and wild-type alleles were assessed by RT-qPCR, as follows. Total RNA was extracted using a Total RNA Purification Kit (Norgen Biotek #17200) and reverse transcribed into cDNA using the SuperScript VILO cDNA Synthesis Kit (Invitrogen #11754050). qRT-PCR was performed on a Quant Studio 3 using SYBR green reagents (Quantbio #95054). Data were normalized to a GAPDH loading control gene and presented as fold change in mRNA expression relative to the average of the control conditions (Ctrl Tet1 and Ctrl VP64).

[0534] As shown in **FIG. 1B**, in the R255X-iPSCs, gRNA 9 and dSpCas9-TET1 led to a greater than 20-fold increase in mRNA expression of the Xi WT allele of MeCP2. Other single gRNAs did not facilitate a comparable increase in expression of the WT allele. Expression of the mutant allele was not substantially affected by any of the individual gRNAs, including gRNA 9, as shown in **FIG. 1C**.

[0535] The results support the utility of an exemplary MeCP2-targeting gRNA, together with a dCas9-TET catalytic domain fusion protein, in reactivating the expression of a WT MeCP2 allele from an inactive X chromosome.

#### **D. Screening of gRNAs targeting MeCP2**

[0536] Additional gRNAs targeting the promoter and first exon of MeCP2 were designed and tested for upregulation of the Xi WT MeCP2 allele in R255X-iPSCs.

[0537] Twenty-nine (29) total gRNAs, as described in **Table E1** above, were screened for re-activation of the Xi WT MeCP2 allele on the inactive X chromosome by co-expression with dSpCas9-TET1.

[0538] In addition to gRNA 9, gRNA 27, which targets an overlapping sequence compared to the gRNA 9 (**FIG. 2**), was found to increase expression of the Xi WT MeCP2 allele. These results show that the region of MeCP2 promoter targeted by gRNA 9 and gRNA 27 represents a specific regulatory region for reactivation of MeCP2 expression from the Xi.

[0539] The results together support the utility of the MeCP2-targeting gRNAs and dCas9-effector domain fusion proteins in reactivating a MeCP2 allele on an inactive X chromosome, and in therapeutic applications for the treatment of Rett syndrome.

#### **Example 2: Increased expression of MeCP2 over time**

[0540] Expression of MeCP2 after transduction with MeCP2-targeting CRISPR/Cas-effector fusion protein was assessed over time.

[0541] iPSCs generated from a patient with Rett syndrome expressing a gRNA targeting MeCP2 and a dCas9-TET catalytic domain fusion protein were assessed over time for expression of WT and mutant MeCP2 mRNA.

[0542] R255X-iPSCs were transduced using lentivirus with plasmids encoding dSpCas9-TET1 and an MeCP2 promoter-targeting gRNA 9, generally as described above in Example 1. Cells were harvested at days 5, 9, 13, 17, 21 and 25 post-transduction, and assessed for MeCP2 expression by RT-qPCR. qRT-PCR data were normalized to a GAPDH loading control gene and presented as fold change in mRNA expression relative to Day 5 MECP2 levels with a non-targeting gRNA.

[0543] As shown in **FIG. 3A**, expression of the Xi WT MeCP2 allele progressively increased from 5 to 21 days post-transduction, and did not further increase from 21 to 25 days post-transduction. In contrast, expression of the Xa mutant MeCP2 allele remained similar throughout the time course (**FIG. 3B**), indicating that MeCP2 activation with dSpCas9-TET1

was specific to the Xi WT allele. The results showed that the expression of the WT MeCP2 allele increased over time for an extended period of time, such as at least 21 days, after introduction of a gRNA targeting MeCP2 and a dCas9-TET catalytic domain fusion protein.

### **Example 3: Single vector delivery for MeCP2 activation**

[0544] The effect of MeCP2-targeting CRISPR/Cas-effector fusion protein, delivered using a dual vector system encoding the Cas-effector fusion protein and gRNA in separate nucleic acids, compared to a single vector system encoding all components on MeCP2 expression, was assessed.

[0545] iPSCs generated from a patient with Rett syndrome were transduced with a dCas9-TET catalytic domain fusion protein and gRNA targeting MeCP2, using a dual vector system or a single vector system and assessed by flow cytometry for expression of MeCP2.

[0546] A single lentiviral vector was designed to allow co-delivery of dSpCas9-TET1 and gRNA 9 from the same plasmid.

[0547] R255X-iPSCs were transduced with dSpCas9-TET1 and gRNA 9, generally as described in Example 1B above, using the one vector system, or using two separate vectors. Cells were assessed for MeCP2 expression by flow cytometry.

[0548] As shown in **FIG. 4A**, 3.8% of cells expressed MeCP2 when transduced with the two vector system. In comparison, 30.5% of cells expressed MeCP2 when transduced with the single vector system, as shown in **FIG. 4B**.

[0549] The results showed that the single vector system substantially increases re-activation of MeCP2 in R255X-iPSCs, showing approximately a 7.5-fold difference in MeCP2<sup>+</sup> cells compared to a two vector system, as assessed by flow cytometry.

### **Example 4: MeCP2 activation using multiple gRNAs**

[0550] The effect of transducing one or two different gRNAs targeting MeCP2 and a Cas-effector fusion protein on MeCP2 expression was assessed. iPSCs generated from a patient with Rett syndrome were transduced with a dCas9-TET catalytic domain fusion protein and one or two different gRNAs targeting an overlapping region in MeCP2 assessed by flow cytometry for expression of MeCP2.

[0551] R255X-iPSCs were transduced with a lentiviral vector encoding dSpCas9-TET1 and gRNA 9, a lentiviral vector encoding dSpCas9-TET1 and gRNA 27 (see Example 1D), or both of the aforementioned vectors together, and assessed for MeCP2 expression by flow cytometry at 20 days after lentiviral transduction.



[0552] As shown in **FIG. 4C**, gRNA 9 and gRNA 27 each individually led to substantial expression of MeCP2 as assessed by flow cytometry. The percentage of MeCP2 expressing cells when both gRNA 9 and gRNA 27 were transduced was similar to the percentage of MeCP2 cells transduced with either of the individual gRNAs. The results indicated that a substantially higher MeCP2 expression was not observed when two different gRNAs targeting the same region were transduced together. The results further support that the region of MeCP2 promoter targeted by gRNA 9 and gRNA 27 represents a specific regulatory region for reactivation of MeCP2 expression, and the substantial MeCP2 reactivation by each of gRNA 9 and gRNA 27.

**Example 5: MeCP2 activation in neurons differentiated from iPSCs**

[0553] iPSCs generated from a patient with Rett syndrome transduced with a gRNA targeting MeCP2 and nucleic acid sequences encoding a dCas9-TET catalytic domain fusion protein were differentiated into neurons, and assessed for neuronal differentiation and MeCP2 expression.

[0554] R255X-iPSCs from a Rett syndrome patient were transduced with dSpCas9-TET1 and gRNA 9 using the one-vector system, generally as described in Example 3 above.

[0555] Plasmids were prepared using the QIAGEN Plasmid Plus Midi Kit (#12945). Lentivirus was generated in HEK293FT cells using Lipofectamine 3000 (ThermoFisher #L3000015) and concentrated to 50x using Lenti-X Concentrator (Takara #631232). Lentivirus encoding dSpCas9-TET1 and gRNA 9 was added to R255X iPSCs at a 1x final concentration. 48 hours after the addition of lentivirus, the cells were selected with 0.5 µg/ml puromycin to enrich for cells expressing dSpCas9-TET1 and gRNA.

[0556] At day 10 post transduction, the cells were transduced with a second lentivirus encoding Ngn2 and switched from iPSC media (mTesR, StemCell Tech #85850 ) to N3 neuronal induction media (DMEM/F12, 1x N2 supplement + 1x B27 supplement). At day 7 post transduction of Ngn2, the cells were fixed with 4% paraformaldehyde and stained with TUBB3 (Biolegend #801201) and MeCP2 (Cell Signaling #3456) antibodies for immunofluorescence.

[0557] As shown in **FIG. 5**, immunofluorescence labeling with antibodies for the neuronal protein TUBB3 and MeCP2 showed that MeCP2 protein expression was observed in TUBB3<sup>+</sup> neurons differentiated from iPSCs.

[0558] The results showed that MeCP2 activation in iPSCs was maintained in differentiated neurons, further supporting the utility of gRNA targeting MeCP2 and dCas9-effector fusion proteins in therapeutic applications for treating Rett syndrome.

**Example 6: Targeted demethylation of MeCP2 promoter in R255X-iPSCs**

[0559] Methylation status of the promoter region of MeCP2 was assessed by bisulfite sequencing, in Rett syndrome patient-derived cells transduced with MeCP2 promoter-targeting gRNA and nucleic acid sequences encoding a Cas-effector fusion protein.

[0560] R255X-iPSCs were transduced using lentivirus with plasmids encoding dSpCas9-TET1 and either the MeCP2 promoter-targeting gRNA 9, or a control non-targeting gRNA, generally as described above in Examples 1 and 3. 48 hours after the addition of lentivirus, the cells were selected with 0.5 µg/ml puromycin to enrich for cells expressing dSpCas9-TET1 and gRNA, and cultured until day 12 post-transduction.

[0561] Cells were harvested on day 12 post-transduction for bisulfite sequencing to assess methylation of the MeCP2 promoter. Harvested cells transfected with the MeCP2 promoter-targeting gRNA 9 were fixed and stained using the Transcription Factor Staining Kit (ThermoFisher #00-5523-00) with a primary conjugated anti-MeCP2 antibody (Cell Signaling #34113), and sorted by fluorescence-activated cell sorting (FACS) on a Sony Sorter MA900 into MeCP2<sup>+</sup> and MeCP2<sup>-</sup> populations.

[0562] Bisulfite sequencing was performed for cells transduced with the non-targeting gRNA, and for the sorted populations of cells transduced with the MeCP2 promoter-targeting gRNA 9. Genomic DNA (gDNA) was extracted using QIAGEN DNeasy Blood and Tissue kit (#69504). gDNA was bisulfite treated using the ZYMO EZ DNA Methylation-Gold Kit (#D5005). The CpG island region at the MeCP2 promoter was PCR amplified. Sequencing libraries were prepared using standard Illumina adapters and barcodes and sequenced on an Illumina Miseq.

[0563] As shown in **FIG. 6**, cells transduced with dSpCas9-TET1 and the MeCP2 promoter-targeting gRNA 9 exhibited reduced overall methylation of the MeCP2 promoter in comparison to control cells transduced with a non-targeting gRNA. In addition, the MeCP2<sup>+</sup> sorted population had reduced overall methylation in comparison to the MeCP2<sup>-</sup> population. The results show that dSpCas9-TET1 with gRNA 9 leads to demethylation of the MeCP2 promoter, and that demethylation is associated with increased MeCP2 expression.

[0564] Taken together, the results indicate that dSpCas9-TET1 with a MeCP2 promoter-targeting gRNA induces demethylation of the MeCP2 promoter to re-activate (e.g., de-repress) allele-specific transcription of the Xi WT MeCP2 allele in Rett syndrome patient-derived cells. The results support the utility of gRNAs targeting MeCP2 and dCas9-effector fusion proteins in therapeutic applications for treating Rett syndrome.

**Example 7: Improved activation of MeCP2 using a modified linker**

[0565] A modified linker and NLS sequence between the TET1 catalytic domain and the dSpCas9 domain of the dSpCas9-TET1 fusion protein was tested for the effect in MeCP2 re-activation.

[0566] R255X-iPSCs were transduced using lentivirus with the MeCP2 promoter-targeting gRNA 9, and either the original dSpCas9-TET1 used in Examples 1-6 above (set forth in SEQ ID NO:91), which includes a 16 amino acid linker sequence (set forth in SEQ ID NO:119) and NLS or a modified dSpCas9-TET1 (set forth in SEQ ID NO:114) with a modified longer 80 amino acid linker (set forth in SEQ ID NO:117) and NLS, as illustrated in **FIG. 7A**.

[0567] 48 hours after transduction, the cells were selected with 0.5 µg/ml puromycin to enrich for cells expressing dSpCas9-TET1 and gRNA. Cells were harvested on day 11 (D11) and day 17 (D17) post-transduction to assess MeCP2 expression.

[0568] MeCP2 expression was assessed by flow cytometry. For flow cytometry, cells were fixed and stained using the Transcription Factor Staining Kit (ThermoFisher #00-5523-00) with a primary conjugated MECP2 antibody (Cell Signaling #34113), and analyzed on a Sony Sorter MA900 to determine the percentage of cells expressing MeCP2.

[0569] As shown in **FIG. 7B**, cells transduced with dSpCas9-TET1 with the modified longer linker exhibited increased expression of MeCP2 (as assessed by % MeCP2 positive cells) in comparison to cells transduced with dSpCas9-TET1 with the original linker. The results support the improved effector activity of a Cas-effector fusion protein using a modified longer linker linking the two domains.

**Example 8: An engineered self-assembling split dCas9-TET1 for MeCP2 activation**

[0570] A self-assembling split dCas9-TET1 fusion protein was engineered and tested in MeCP2 re-activation.

[0571] A two-vector system was engineered for expression of a split dCas9-TET1 fusion protein, using trans-splicing inteins. Inteins are internal protein elements that self-excite from their host protein and catalyze ligation of flanking sequences with a peptide bond. In this two-vector system, the first vector encoded a polypeptide comprising the TET1 catalytic domain and an N-terminal fragment of dSpCas9, followed by an N-terminal Npu Intein (TET1-dSpCas9-573N; set forth in SEQ ID NO:121). The second vector encoded a polypeptide comprising a C-terminal Npu Intein, followed by a C-terminal fragment of dSpCas9 (dSpCas9-573C; set forth in SEQ ID NO:131). The N- and C-terminal fragments of the encoded fusion protein were split at position 573Glu of the dSpCas9 molecule, with reference to positions of SEQ ID NO:96. The N-

terminal Npu Intein (SEQ ID NO:129) and C-terminal Npu Intein (set forth in SEQ ID NO:133) were engineered to self-excise and ligate the N- and C-terminal fragments, thereby forming the full-length dSpCas9-TET1 fusion protein when expressed in a cell, as illustrated in **FIG. 8**.

**[0572]** The split dSpCas9-TET1 fusion protein was assessed for the ability to activate MeCP2 in R255X-iPSCs. Plasmids were prepared for each of the split dSpCas9-TET1 components, with each plasmid including a gRNA expression cassette for the MeCP2 promoter-targeting gRNA 9. Plasmids were prepared using a QIAGEN Plasmid Plus Midi Kit (#12945). Lentivirus was generated in HEK293FT cells using Lipofectamine 3000 (ThermoFisher #L3000015) and concentrated to 50x using Lenti-X Concentrator (Takara #631232).

**[0573]** Lentivirus encoding dSpCas9-573C/gRNA 9 was incubated first with the R255X iPSCs at a 1x final concentration. 48 hours after the addition of lentivirus, the cells were selected with 2 µg/ml blasticidin to enrich for transduced cells. Following selection for 10 days, cells were then transduced with lentivirus encoding TET1-dCas9-573N/gRNA 9 and selected with 0.5 µg/ml puromycin. Negative control cells were included that were only transduced with the dSpCas9-573C/gRNA 9 lentivirus, and positive control cells were included that were transduced with lentivirus encoding a non-split dSpCas9-TET1 fusion protein (set forth in SEQ ID NO:91) and gRNA 9.

**[0574]** Cells were harvested to assess MeCP2 expression on day 12 post-transduction of TET1-dSpCas9-573N. MeCP2 expression was assessed by flow cytometry. For flow cytometry, cells were fixed and stained using the Transcription Factor Staining Kit (ThermoFisher #00-5523-00) with a primary conjugated anti-MeCP2 antibody (Cell Signaling #34113), and analyzed on a Sony Sorter MA900 to determine the percentage of cells expressing MeCP2.

**[0575]** As shown in **FIG. 9**, the expression of both components of the split dSpCas9-TET1 fusion protein with gRNA 9 led to activation of MeCP2, to an extent comparable to the non-split dSpCas9-TET1 fusion protein.

**[0576]** The results support the utility of using the dSpCas9-TET1 split fusion protein for targeted demethylation and activation of MeCP2 in therapeutic applications for treating Rett syndrome. A two-vector system encoding a split fusion protein is advantageous in some therapeutic applications. For example, a split fusion protein may assist or improve packaging of larger components (e.g., larger Cas9-effector fusion proteins) in therapeutic delivery vectors with limited capacity, such as an adeno-associated virus (AAV) vector.

#### **Example 9: dSpCas9-TET1 mediated activation of MeCP2 in mouse fibroblasts**

**[0577]** Guide RNAs (gRNAs) targeting the mouse methyl-CpG-binding protein 2 (MeCP2)

gene was designed and screened for transcriptional re-activation of MeCP2 allele in the inactive X (Xi) chromosome, using a mouse fibroblast reporter cell line.

**[0578]** 7 gRNAs were designed to target regulatory region of mouse MeCP2. The mouse MeCP2-targeting gRNAs are indicated in **Table E2**.

**Table E2.** Mouse MeCP2-targeting gRNAs

SEQ ID NO:	gRNA name	gRNA spacer sequence
223	gRNA m1	GCCGCGCCGAGCGGAGGAGG
224	gRNA m2	CTGGCGTTGTTCCAAGCCAA
225	gRNA m3	AGTGGGACCGCCAAGGCCGC
226	gRNA m4	TGCTGACTGGTATCAGGGTA
227	gRNA m5	GACCGCCAAGGCCGCGGGCG
228	gRNA m6	TTGGAAAAAAGAGGCGGCTA
229	gRNA m7	GGGGCAAAAAGTCACGGAAT
230	Non-targeting gRNA	GATCGGTTATGTTTAGGGTT

**[0579]** Plasmids and lentiviral vectors were generated, generally as described in Example 1. The gRNAs were then screened for the ability to increase expression of an inactivated allele of MeCP2 when co-expressed with dSpCas9-TET1 in a mouse fibroblast cell line with a transgenic inactive X (Xi) allele of MeCP2 with a luciferase reporter, generally as described in Sripathy et al., PNAS 114(7):1619-1624, 2017. Lentivirus encoding dSpCas9-TET1 and each gRNA or a non-targeting gRNA was incubated with MeCP2-Luciferase mouse fibroblasts at a 1x final concentration. 48 hours after the addition of lentivirus, cells were selected with 1 µg/ml puromycin to enrich for cells expressing dSpCas9-TET1 and gRNA. Cells were harvested on day 16 and day 29 post-transduction to assess activation of the Xi MeCP2 reporter allele by qRT-PCR.

**[0580]** To measure the expression of MeCP2, Xi MeCP2 reporter-allele-specific primers were designed for qRT-PCR. qRT-PCR was performed, generally as described in Example 1. Data were normalized to a Gapdh loading control gene, and the fold change in mRNA expression relative to a non-targeting gRNA control was determined.

**[0581]** As shown in **FIG. 10**, several of the mouse MeCP2-targeting gRNAs resulted in an increase in MeCP2 mRNA expression, including gRNA m1 which led to ~3-fold and ~10-fold increase in the expression of the Xi MeCP2 reporter allele at Day 15 and Day 29 post-transduction, respectively. The results support the utility of dCas9-TET1 with MeCP2-targeting gRNAs in reactivating the expression of MeCP2 from an inactive X chromosome in various different species, such as in a mouse, and in diverse cell types, such as in a fibroblast.

**Example 10: ZFP-mediated transcriptional activation of MeCP2**

[0582] Fusion proteins containing DNA-targeting domains based on zinc finger proteins (ZFP) that target the MeCP2 locus were designed, generated, and assessed for their effect in re-activation of MeCP2 in cells.

[0583] ZFP-based DNA-targeting domains targeting regulatory elements of MeCP2, including promoter-targeting and enhancer-targeting ZFP DNA-targeting domains, were designed, based on available methods for designing ZFP targeting specific target sequences. Exemplary ZFP DNA-targeting domains target sequences within the genomic coordinates human genome assembly GRCh38 (hg38) 154,097,151-154,098,158. The exemplary genomic regions specified above contained multiple sequentially tiled target sites, designing ZFPs targeting one of the tiled target sites in the region. Fusion proteins were designed, each comprising one of the designed ZFP DNA-targeting domains fused to a TET catalytic domain, such as a TET1 catalytic domain (set forth in SEQ ID NO:93). Exemplary fusion proteins included MeCP2-targeting ZFP-TET1.

[0584] Viral vectors, including lentiviral vectors, were designed and cloned, each comprising nucleic acid sequences encoding a MeCP2-targeting ZFP-TET1. Vectors further encoded a selectable marker (*e.g.* puromycin resistance cassette).

[0585] R255X-iPSCs, generally as described above in Example 1, are transduced using lentivirus with plasmids encoding one of the MeCP2-targeting ZFP-TET1, and enriched for transduced cells (*e.g.* using puromycin selection). Negative control cells are transduced with a non-targeting ZFP-TET1 fusion protein, or the ZFP MeCP2-targeting DNA-targeting domains without TET1. Cells are harvested and assessed for MeCP2 expression by RT-qPCR, generally as described in Example 1. qRT-PCR data are normalized to a GAPDH loading control gene and presented as fold change in mRNA expression relative to negative control cells.

[0586] Cells transduced with MeCP2-targeting ZFP-TET1 show increased mRNA expression of the Xi WT allele of MeCP2 compared to the negative control and the expression of the mutant allele is not substantially affected.

[0587] The results support the utility of an exemplary MeCP2-targeting ZFP-TET1 fusion protein in reactivating the expression of a WT MeCP2 allele from an inactive X chromosome.

**Example 11: TALE-mediated transcriptional activation of MeCP2**

[0588] Fusion proteins containing DNA-targeting domains based on transcription activator-like effector (TALE) binding domains that target the MeCP2 locus are designed, generated, and assessed for their effect in re-activation of MeCP2 in cells.

**[0589]** TALE-based DNA-targeting domains targeting regulatory elements of MeCP2, including promoter-targeting and enhancer-targeting TALE DNA-targeting domains, are designed, based on available methods for designing TALE targeting specific target sequences. Exemplary TALE DNA-targeting domains target sequences within the genomic coordinates human genome assembly GRCh38 (hg38) 154,097,151-154,098,158. The exemplary genomic regions specified above contained multiple sequentially tiled target sites, designing TALEs targeting one of the tiled target sites in the region. Fusion proteins are designed, each comprising one of the designed TALE DNA-targeting domains fused to a TET catalytic domain, such as a TET1 catalytic domain (set forth in SEQ ID NO:93). Exemplary fusion proteins included MeCP2-targeting TALE-TET1.

**[0590]** Viral vectors, including lentiviral vectors, are designed and cloned, each comprising nucleic acid sequences encoding a MeCP2-targeting TALE-TET1. Vectors further encoded a selectable marker (*e.g.* puromycin resistance cassette).

**[0591]** R255X-iPSCs, generally as described above in Example 1, are transduced using lentivirus with plasmids encoding one of the MeCP2-targeting TALE-TET1, and enriched for transduced cells (*e.g.* using puromycin selection). Negative control cells are transduced with a non-targeting TALE-TET1 fusion protein, or the TALE MeCP2-targeting DNA-targeting domains without TET1. Cells are harvested and assessed for MeCP2 expression by RT-qPCR, generally as described in Example 1. qRT-PCR data are normalized to a GAPDH loading control gene and presented as fold change in mRNA expression relative to negative control cells.

**[0592]** Cells transduced with MeCP2-targeting TALE-TET1 show increased mRNA expression of the Xi WT allele of MeCP2 compared to the negative control and the expression of the mutant allele is not substantially affected.

**[0593]** The results support the utility of an exemplary MeCP2-targeting TALE-TET1 fusion protein in reactivating the expression of a WT MeCP2 allele from an inactive X chromosome.

**[0594]** The present invention is not intended to be limited in scope to the particular disclosed embodiments, which are provided, for example, to illustrate various aspects of the invention. Various modifications to the compositions and methods described will become apparent from the description and teachings herein. Such variations may be practiced without departing from the true scope and spirit of the disclosure and are intended to fall within the scope of the present disclosure.

Sequences

#	SEQUENCE	ANNOTATION
1	TAAGGATTAATGGACCCTTG	gRNA 1 spacer DNA
2	CGCCTCTTTTCCCAAACGA	gRNA 2 spacer DNA
3	CCATCACAGCCAATGACGGG	gRNA 3 spacer DNA
4	TCGGTGCATCTGTGGACAGA	gRNA 4 spacer DNA
5	GGGGCGCGACGTCGGCCGTG	gRNA 5 spacer DNA
6	TCAGCGGCGATGCCGTCAAT	gRNA 6 spacer DNA
7	CGCGGAGGGACTGGTTTAGT	gRNA 7 spacer DNA
8	AAGAGGGCGGGGCGCGACGT	gRNA 8 spacer DNA
9	GCTGCGAGCCCGCCCGTCAT	gRNA 9 spacer DNA
10	ACTTGCCCCAGCATCCGCAA	gRNA 10 spacer DNA
11	CCGTTACTCGGCCCCCCCAC	gRNA 11 spacer DNA
12	GCGCCCCCTCTCCCGTTACT	gRNA 12 spacer DNA
13	GGAGGGGGAGAGCGCGATCC	gRNA 13 spacer DNA
14	TCGCCGGGGCTTCGCCTGTC	gRNA 14 spacer DNA
15	AGGCGAAGCCCCGGCGACAG	gRNA 15 spacer DNA
16	CCGGGATGCGCGTCGAGGGC	gRNA 16 spacer DNA
17	GGACGGTCACCCGCGAGCAG	gRNA 17 spacer DNA
18	CGGCCCGTCACCCCTGCTCG	gRNA 18 spacer DNA
19	GCCGAGGGGAGAGTCGCCAC	gRNA 19 spacer DNA
20	CGGACGACACGGCTGGCGGA	gRNA 20 spacer DNA
21	GCCGTGTCTGTCGACCCCGC	gRNA 21 spacer DNA
22	CTGACCCCGCCCCCGGCA	gRNA 22 spacer DNA
23	CGGGCGGGGACCCTTGCCGG	gRNA 23 spacer DNA
24	GACTGTGAGTGGGACCGCCG	gRNA 24 spacer DNA
25	GGTAAAAGCCGTCCGGAAAA	gRNA 25 spacer DNA
26	GCGCGCGCGCTCCCTCCTCT	gRNA 26 spacer DNA
27	AATGACGGGCGGGCTCGCAG	gRNA 27 spacer DNA



28	TTTCCTGTCCATTTTCGGCCA	gRNA 28 spacer DNA
29	GCTGCTTTTCGGCCGTCGTTT	gRNA 29 spacer DNA
30	GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCCGAGUCGGUGC	SpCas9 gRNA scaffold sequence
31	UAAGGAUUAUUGGACCCUUG	gRNA 1 spacer RNA
32	CGCCUCUUUUCCCCAAACGA	gRNA 2 spacer RNA
33	CCAUCACAGCCAAUGACGGG	gRNA 3 spacer RNA
34	UCGGUGCAUCUGUGGACAGA	gRNA 4 spacer RNA
35	GGGGCGCGACGUCGGCCGUG	gRNA 5 spacer RNA
36	UCAGCGGCGAUGCCGUCAAU	gRNA 6 spacer RNA
37	CGCGGAGGGACUGGUUUAGU	gRNA 7 spacer RNA
38	AAGAGGGCGGGGCGCGACGU	gRNA 8 spacer RNA
39	GCUGCGAGCCC GCCCGUCAU	gRNA 9 spacer RNA
40	ACUUGCCCCAGCAUCCGCAA	gRNA 10 spacer RNA
41	CCGUUACUCGGCCCCCCCAC	gRNA 11 spacer RNA
42	GCGCCCCUCUCCCGUUACU	gRNA 12 spacer RNA
43	GGAGGGGGAGAGCGCGAUCC	gRNA 13 spacer RNA
44	UCGCCGGGGCUUCGCCUGUC	gRNA 14 spacer RNA
45	AGGCGAAGCCCCGGCGACAG	gRNA 15 spacer RNA
46	CCGGGAUGCGCGUCGAGGGC	gRNA 16 spacer RNA
47	GGACGGUCACCCGCGAGCAG	gRNA 17 spacer RNA
48	CGGCCCGUCACCCUGCUCG	gRNA 18 spacer RNA
49	GCCGAGGGGAGAGUCGCCAC	gRNA 19 spacer RNA
50	CGGACGACACGGCUGGCGGA	gRNA 20 spacer RNA
51	GCCGUGUCGUCCGACCCCGC	gRNA 21 spacer RNA
52	CUGACCCCGCCCCCGGCA	gRNA 22 spacer RNA
53	CGGGCGGGGACCCUUGCCGG	gRNA 23 spacer RNA
54	GACUGUGAGUGGGACCCCG	gRNA 24 spacer RNA
55	GGUAAAAGCCGUCCGGAAAA	gRNA 25 spacer RNA
56	GCGCGCGCGCUCCUCCUCU	gRNA 26 spacer RNA

57	AAUGACGGGCGGGCUCGCAG	gRNA 27 spacer RNA
58	UUUCCUGUCCAUUUUCGGCCA	gRNA 28 spacer RNA
59	GCUGCUUUCGGCCGUCGUUU	gRNA 29 spacer RNA
60	<p>ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGCACGTTAAGCGACCTGCCGCCACAAAGAAGGCTGGACAGGCTAAGAAGAAGAACTGGACGTTCTGCCACCTGcagctgtcttgatcgagttatatacaaaaagacaaa ggcccatattatatacacaccttggggcaggaccaagtgttgcctgtcagggaaatcat ggagaataggtatggtcaaaaaggaaacgcaataaggatagaaatagtgtacaccg gtaaagaagggaaaagctctcatgggtgtccaattgctaagtgggttttaagaagaagc agtgtgaagaaaaagttctttgtttgggtcgggcagcgtacaggccaccactgtccaac tgctgtgatgggtggtgctcatcatgggtgtgggatggcatccctctccaatggccgacc ggctatacacagagctcacagagaatctaaagtcatacaatgggcaccctaccgacaga agatgcaccctcaatgaaaaatcgtacctgtacatgtcaaggaaattgatccagagacttg tggagcttcattctcttttggctgttcatggagtatgtactttaatggctgtaaagttg gtagaagcccaagccccagaagattagaattgatccaagctctccctacatgaaaa aaccttgaagataacttacagagtttggctacacgattagctccaatttataagcagta tgcctcagtagcttaccaaaatcagggtggaatatgaaaatggtgcccgagaatgtcggc ttggcagcaaggaaggtcgacccttctctggggctcactgcttgccctggacttctgtgct catccccacagggacattcacacatgaataatggaagcactgtgggttgtacctaac tcgagaagataaccgctctttgggtgttattcctcaagatgagcagctccatgtgctac ctctttataagctttcagacacagatgagtttggctccaaggaaggaatggaagccaag atcaaatctggggccatcgaggtcctggcacccegcgcaaaaaaagaacgtgttccac tcagcctgttccccgttctggaaagaagggctgcgatgatgacagaggttcttgcaac ataagataaagggcagtggaagaaagaaacctattccccgaatcaagcggagaataactca acaacaacaacaacagtaagccttcgctcactgccaaccttagggagtAACACTGAGAC cgtgcaacctgaagtaaaaagtgaaaccgaacccttttatcttaaaaagttcagaca acactaaaacttattcgtctgatgccatccgctcctcaccaccagtgaagagggcatctcca ggcttctcctgggtccccgaagactgcttcagccacaccagctccactgaagaatgacgc aacagcctcatgcggttttccagaagaagcagcactcccactgtacgatgcttggg gaagactcagtggtgccaatgctgcagctgctgatggccctggcatttccacagcttggc gaagtggtcctctccccaccctgtctgctcctgtgatggagccctcaatattctga gccttccactgggtgtgactgagccgctaacgcctcatcagccaaaccaccagccctcct tccctacctctcctcaagaccttgccctcttctccaatggaagaagatgagcagcattct gaagcagatgagcctccatcagacgaaccctatctgatgacccctgtcaccctgctga ggagaaattgccccacattgatgagtttgggtcagacagtgagcacaatcttttggatg caaatattgggtgggtggccatcgcacctgctcagggctcgggttttgattgagtgccc cggcgagagctgcaagctaccaactcctgttgagcaccoccaaccgtaataatcaaccocg cctctcccttgtcttttaccagcacaacaaacctaaataagccccaacatggttttgaac taaacaaagattaaagtttgaggctaaagaagctaaagaataagaaaatgaaggcctcagag caaaaagaccaggcagctaatgaaggtccagaacagtcctctgaagtaaatgaattgaa ccaaattccttctcataaagcattaacattaaccctatgacaatgttgtcacogtgcct cttatgctctcacacacggttgcggggcCCTATAACCATTGGGTCAACCCAAAGAAGAAG CGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACGAAGAAGTACTCCATTGGGCTCGC CATCGGCACAAACAGCGTCCGGCTGGGCGTCATTACGGACGAGTACAAGGTGCCGAGCA AAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTTGGC GCCCTCCTGTTGACTCCGGGGAAACCGCCGAAGCCACGCGGCTCAAAGAACAGCACG GCGCAGATATACCCGAGAAAGAATCGGATCTGCTACCTgcaggGAGATCTTTAGTAATG AGATGGCTAAGGTGGATGACTCTTTCTTCATAGGCTGGAGGAGTCCTTTTTGGTGGAG GAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTA CCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATA AGGCTGACTTGCAGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCCGGGACAC TTCCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCA ACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTG ACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAATCCCGGCGGCTCGAAAACCTCATC GCACAGCTCCCTGGGGAGAAGAAGACGGCCTGTTTGGTAATCTTATCGCCCTGTCACT CGGGCTGACCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAAC TGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAG TACGCAGACCTTTTTTTGGCGGCAAGAACCTGTGACAGCCATTCTGCTGAGTGATAT</p>	FLAG-NLS-TET1-NLS-dSpCas9-NLS-P2A-PuroR vector sequence

TCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCT  
ATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTGACAGCAACTGCCT  
GAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA  
CGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGAAAAAATGG  
ACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC  
ACTTTCGACAATGGAAGCATCCCCACCAGATTACCTGGGCGAACTGCACGCTATCCT  
CAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAAGATAACAGGGAAAAAGATTGAGAAAA  
TCCTCACATTTCCGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCAGATTC  
GCGTGGATGACTCGCAAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAGGAAGTCGT  
GGATAAGGGGGCCTCTGCCAGTCTTTCATCGAAAGGATGACTAACTTTGATAAAAAATC  
TGCCTAACGAAAAGGTGCTTCCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTAT  
AACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAAGCCAGCATTCTGTGTC  
TGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCCGAAAAGTTACCG  
TGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTTCGACTCTGTTGAAAATC  
AGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCAGGATCTTCTGAAAAAT  
CATTAAAGACAAGGACTTCTTGACAATGAGGAGAACGAGGACATTCTTGAGGACTTG  
TCCTCACCCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTAC  
GCTCATCTCTTCGACGACAAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATG  
GGGGCGGCTGTCAAGAAAATGATCAATGGgat cCGAGACAAGCAGAGTGGAAAAGACAA  
TCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCAT  
GATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAAGCACAAGTTTCTGGCCAGGGGGA  
CAGTCTTCCAGGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAAATAC  
TGCAGACCGTTAAGGTCGTGGATGAACCTGTCAAAAGTAATGGGAAGGCATAAGCCCGAG  
AATATCGTTATCGAGATGGCCCGAGAGAACCAAACCTACCCAGAAGGGACAGAAGAACAG  
TAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCAAAATCCTTA  
AGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTG  
CAGAACGGCAGGGACATGTACGTGGATCAGGAACCTGGACATCAATCGGCTCTCCGACTA  
CGACGTGGATGCCATCGTGCCCCAGTCTTTTTCTCAAAGATGATTCTATTGATAATAAAG  
TGTTGACAAGATCCGATAAAAAATAGAGGGAAAGAGTGATAACGTCCCTCAGAAGAAGTT  
GTCAAGAAAAATGAAAAATTTATTGGCGGAGCTGCTGAACGCCAAACTGATCACACAACG  
GAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCG  
GCTTCATCAAAAGGCAGCTTGTGAGACACGCCAGATCACCAA<sub>gca</sub>cGTGGCCAAAAT  
CTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTGAGAGGTGAA  
AGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTGAGAAAGGACTTTCAGTTTTATA  
AGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTA  
GGCAGTGCCTTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTA  
TAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCA  
CCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTG  
GCCAATGGAGAGATTGGAAGCGACCCTTATCGAAAACAAACGGAGAAAACAGGAGAAAT  
CGTGTGGGACAAGGGTAGGGATTTTCGCGACAGTCCGGAAGGTCTGTCCATGCCGCAGG  
TGAACATCGTTAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTC  
CCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATA  
CGGCGGATTCGATTCTCTACAGTGCCTTACAGTGTACTGGTTGTGGCCAAAAGTGGAGA  
AAGGGAAGTCTAAAAAACTCAAAGCGTCAAGGAACCTGCTGGGCATCACAATCATGGAG  
CGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGT  
CAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAAACGGCC  
GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAAGGTAACGAGCTGGCACTGCC  
TCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCC  
CGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGA  
TCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGAT  
AAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAA  
CATTATCCACTTGTTACTCTGACCAACTTGGGCGGCGCTGCAGCCTTCAAGTACTTCG  
ACACCACCATAGACAGAAAGCGGTACACCTCTACAAAAGGAGGTCTTGACGCCACTG  
ATTCATCAGTCAATTACGGGGCTCTATGAAAACAAGAATCGACCTCTCTCAGCTCGGTGG  
AGACAAAAGGCCGGCGGCCACGAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGct agCG  
CAACAAACTTCTCTCTGCTGAAACAAGCCGGAGATGTCGAAGAGAATCCTGGACCGATG  
ACCGAGTACAAGCCCACGGTGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACG  
CACCTCGCCCGCGCTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACC  
GCCACATCGAGCGGGTACCAGCTGCAAGAACTCTTCTCACGCGCTCGGCTCGAC  
ATCGGCAAGGTGTGGGTGCGGACGACGCGCCGCGGTGGCGGTCTGGACCACGCCGGA  
GAGCGTCAAGCGGGGGCGGTGTTGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCG



88	UUUCCUGUCCAUUUUCGGCCAGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAA UAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCCGAGUCGGUGC	gRNA 28
89	GCUGCUUUCGGCCGUCGUUUGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAA UAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCCGAGUCGGUGC	gRNA 29
90	AAGCGACCTGCCGCCACAAAGAAGGCTGGACAGGCTAAGAAGAAGAACTGGACGTTCT GCCCACCTGcagctgtcttgatcgagttatacaaaaagacaaaggccatattatacac accttggggcaggaccaagtgttgctgctgtcagggaaatcatggagaataggtatggt caaaaaggaaacgcaataaggatagaaatagtgtgtacaccggtaaagaagggaag ctctcatgggtgtccaattgctaatgtgggttttaagaagaagcagtgatgaagaaaag tctttgtttgggtcggcagcgtacagggcaccactgtccaactgctgtgatggtggtg ctcatcatggtgtgggatggcatccctctccaatggcgcaccggctatacacagagct cacagagaatctaaagtcatacaatgggcaccctaccgacagaagatgcaccctcaatg aaaatcgtacctgtacatgtcaaggaattgatccagagacttgtggagcttcattctct tttggctgttcatggagtatgtactttaatggctgtaagtttggtagaagcccaagccc cagaagattagaattgatccaagctctcccttacctgaaaaaaaccttgaagataact tacagagtttggctacacgattagctccaatttataagcagtgatgctccagtagcttac caaaatcaggtggaatagaaaatglttgcccagagaatgtcggctggcagcaaggaag tcgacccttctctgggtcactgcttgcctggacttctgtgctcatccccacagggaca ttcacaacatgaataatggaagcactgtggtttgtaccttaactcgagaagataaccgc tctttgggtgttattctctcaagatgagcagctccatgtgctacctctttataagcttct agacacagatgagtttggctccaaggaaggaatggaagccaagatcaaatctggggcca tcgaggtcctggcaccctcgcgcaaaaaagaaagctgttccactcagcctgttcccctg tctggaaagaagagggctgcgatgatgacagaggttcttgcacataagataagggcag ggaaaagaacctattccccgaatcaagcgggaagaataactcaacaacaacaacaaca gtaagcctctcgtcactgccaaccttagggagtaaacactgagaccgtgcaacctgaagta aaaagtgaaccgaaccctttttatcttaaaaagttcagacaacactaaaacttattc gctgatgccatccgctcctcaccagtgaaagagggcatctccaggcttctcctggtccc cgaagactgcttcagccacaccagctccactgaagaatgacgcaacagcctcatgccc ttttcagaaagaagcagcactcccactgtacgatgcttcgggaagactcagtggtgc caatgctgcagctgctgatggcctggcatttcacagcttggcgaagtggtcctctcc ccacctgtctgctcctgtgatggagccctcatttaattctgagcctccactggtg actgagccgctaacgctcatcagccaaaccaccagcctcctcctcactctcctca agaccttgctcttctccaatggaagaagatgagcagcattctgaagcagatgagcctc catcagacgaaccctatctgatgacccctgtcactgctgaggagaaattgccccac attgatgagtattggtcagacagtgagcacaatcttttggatgcaaatattggtgggt ggccatcgcacctgctcagcgtcggttttgattgagtgtgcccggcgagagctgcacg ctaccactcctgttgagcaccceaacctaatcatccaaccgcctctcccttgtcttt taccagcacaacctaataaagcccaacatggttttgaactaaacaagataagtt tgaggctaagaagctaagaataagaaaatgaaggcctcagagcaaaaagaccagcag ctaataagaggtccagaacagctcctctgaagtaaatgaattgaaccaaattccttctcat aaagcattaacattaacccatgacaatgttgtcacogtgcoccttatgctctcacaca cgttgcggggcCCTATAACCATTTGGGTCAACCCAAAGAAGAAGCGGAAGGTTCGGTATCC ACGGAGTCCCAGCAGCCGACAAGAAGTACTCCATTGGGCTCGCCATCGGCACAAAACAGC GTCGGCTGGGCCGTCAATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTCT GGCAATACCGATCGCCACAGCATAAAGAAGAAGCACTTATGGCGCCCTCCTGTTCCGACT CCGGGAAACCGCCGAAGCCAGCCGCTCAAAAGAACAGCACGGCGCAGATATAACCCGC AGAAAGAATCGGATCTGCTACCTgcaggAGATCTTTAGTAATGAGATGGCTAAGGTGGA TGACTCTTTCTTCCATAGGCTGGAGGAGTCTTTTGGTGGAGGAGGATAAAAAGCACG AGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCA ACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT GATCTATCTCGCGCTGGCGCATATGATCAAATTTGGGGACACTTCTCATCGAGGGGG ACCTGAACCCGACAAACAGCGATGTCGACAACTCTTTATCCAAGTGGTTGAGACTTAC AATGACCTTTTCGAAGAGAACCAGTCAACGCATCCGGAGTTGACGCCAAAGCAATCCT GAGCGCTAGGCTGTCCAATCCCGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGG AGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAAC TTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTA CGATGATGATCTCGACAATCTGCTGGCCAGATCGGCAGCAGTACGCAGACCTTTTTT TGGCGGCAAAAGAACTGTGAGCGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACG GAGATCAACCAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCA AGACTTGGCTTTGCTGAAGGCCCTTGTGACAGCAACTGCCTGAGAAGTACAAGGAAA TTTTCTTCGATCAGTCTAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAG	NLS-TET1- NLS-dSpCas9- NLS (nt)

	<p>GAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCT  GCTGGTAAAGCTTAACAGAGAAGATCTGTGCGCAAACAGCGCACTTTCGACAATGGAA  GCATCCCCCACCAGATTACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGAT  TTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCCGAT  ACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCAGATTTCGCGTGGATGACTCGCA  AATCAGAAGAGACCATCACTCCCTGGAAC TTCGAGGAAAGTCGTGGATAAGGGGGCCTCT  GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAAATCTGCCTAACGAAAAAGGT  GCTTCCTAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGG  TCAAATACGTCACAGAAGGGATGAGAAAAGCCAGCATTCTGTCTGGAGAGCAGAAGAAA  GCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAAGTTACCGTGAAACAGCTCAAAAGA  AGACTATTTCAAAAAGATTGAATGTTTTGACTCTGTTGAAATCAGCGGAGTGGAGGATC  GTTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGAC  TTCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCTCACCCCTTACGTT  GTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCATCTCTTCGACG  ACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGCGGTGTCAAGA  AAACTGATCAATGgatcCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTCTTAA  GTCCGATGGATTTGCCAACCGGAAC TTCATGCAGTTGATCCATGATGACTCTCTCACCT  TTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCAC  ATCGCTAATCTTGAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGT  CGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGT TATCGAGA  TGGCCCGAGAGAACCAAAC TACCCAGAAGGGACAGAAGAACAGTAGGGAAAAGGATGAAG  AGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTGA  AAACACCCAGCTTCAGAATGAGAAGCTTACCTGTACTACCTGCAGAACCGCAGGGACA  TGTACGTGGATCAGGAAC TGGACATCAATCGGCTCTCCGACTACGACGTGGATGCCATC  GTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGA  TAAAAATAGAGGGAAGAGTGATAACGTCCCTCAGAAGAAGTTGTCAAGAAAATGAAAA  ATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACCGGAAGTTCGATAATCTG  ACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAAGCCGGCTTCATCAAAAAGGCA  GCTTGTGAGACACGCCAGATCACCAAgcaCGTGGCCCAAATTCGATTACGCATGA  ACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTTGAAG  TCTAAGCTGGTCTCAGATTTTCAAGAAAGGACTTTTCAAGTTTATAAAGGTGAGAGATCAA  CAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCA  AAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT  AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTT  TTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTC  GGAAGCGACCCTTATCGAAACAAACCGGAGAAACAGGAGAAATCGTGTGGGACAAGGGT  AGGGATTTGCGACAGTCCGGAAGTCTGTCCATGCCGAGGTGAACATCGTTAAAAA  GACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACGCG  ACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAAATACGGCGGATTCGATTCT  CCTACAGTTCGTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAAGGGAAGTCTAAAAA  ACTCAAAGCGTCAAGGAAC TGTGGGCATCACAAATCATGGAGCGATCAAGCTTCGAAA  AAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATC  ATTAAGCTTCCCAAGTACTCTCTTTGAGCTTGAAAACGGCCGAAACGAATGCTCGC  TAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAAT  TCTTGATCTGGCCAGCCACTATGAAAAGTCAAAGGGTCTCCCGAAGATAATGAGCAG  AAGCAGCTGTTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAG  CGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTT  ACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTT  ACTCIGACCAACTTGGGCGCGCTGCAGCCTTCAAGTACTTCGACACCACCATAGACAG  AAAGCGGTACACCTCTACAAAGGAGGTCTGGACGCCACACTGATTTCATCAGTCAATTA  CGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAAAAGGCCGGCG  GCCACGAAAAGGCCGGCCAGGCAAAAAAGAAAAAG</p>	
<p>91</p>	<p>KRPAATKKAGQAKKKKLDVLP TCSCLDRVIQKDKGPYYTHLGAGP SVAAVRE IMENRYG  QKGN AIRIEIVVYTGKEGKSSHGCP IAKWVLRSSDEEKVLC LVRQRTGHHCP TAVMVV  LIMVWDGIPLPMADRLYTEL TENLKS YNGHPTDRRCTLNENRTCTCQIDPETCGASF S  FGCSWSMYFNGCKFGRSP SPRRFRI DPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAY  QNQVEYENVARECRLGSK EGRPFSGV TACLD FCAHPHRDIHNMNNGSTVVCTLTREDNR  SLGVIPQDEQLHVLPLYKLSDTDEF GSKEGMEAKIKSGAIEVLAPRRKKRTCF TQPVPR  SGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNS TTTNNSKPS SLP TLGSNRET VQPEV  KSETEPHFILKSSDNITKTYSLMP SAPHVKEASPGF SWSPKTASATPAPLKN DATASC G  FSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSE PSTGV</p>	<p>NLS-TET1-  NLS-dSpCas9-  NLS (aa)</p>

	<p>TEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSSEADEPPSDEPLSDDPLSPAEEKLPH          IDEYWSDESEHIFLDANIGGVAIAPAHGSQLIECARRELHATTPVEHPNRRNHPTRLSLVF          YQHKNLNKPKQHGFELNKIKFEAKEAKNKKMKASEQKDAQANEGPEQSSEVNELNQIPSH          KALTLTHDNVTVSPYALTHVAGPYNHVWNPKKKRKVGIHGVPAAADKKYSIGLAIGTNS          VGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR          RKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP          TIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTY          NQLFEENP INASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSGLTPN          FKSNFDLAEDAQLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT          EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ          EEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED          FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKAS          AQSFIERTMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK          AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKD          FLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHHLFDDKVMKQLKRRRYTGWGRLSR          KLINGIRDKQSGKTLDFLKSDFANRNFMQLIHDDSLTFKEDIQKAQVSGQDGLHEH          IANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK          RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAI          VPQSFLLKDDSIDNKVLRSDKNRGRKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNL          TKAERGGSELDDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVI TLK          SKLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVGTAI KKYPKLESEFVYGDYKVDV          RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG          RDFAVSRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS          PTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSFEKNPIDFLEAKGYKEVKKDLI          IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLLYLASHYEKLGKSPEDNEQ          KQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENIHLF          TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDKRPA          ATTKAGQAKKKK</p>	
<p>92</p>	<p>CTGCCACCTgcagctgtcttgatcgagttatacaaaaagacaaaggcccatattatac          acaccttggggcaggaccaagtgttgcctgctgcagggaaatcatggagaataggtatg          gtcaaaaaggaacgcaataaggtatgaaatagtgtacacccggtaaagaagggaaa          agctctcatgggtgtccaattgctaagtggttttaagaagaagcagtgatgaagaaaa          agttctttgtttggtccggcagcgtacaggccaccactgtccaactgctgtgatgggtgg          tgctcatcatgggtgtgggatggcatccctcttccaatggccgaccgctatcacagag          ctcacagagaatctaaagtcatacaatgggcaccctaccgacagaagatgcaccctcaa          tgaaaatcgtacctgtacatgtcaaggaattgatccagagacttgtggagcttcatctct          cttttggctgttcatggagatgtactttaatggctgtaagtttggtagaagcccaagc          cccagaagatttagaattgatccaagctctcccttacctgaaaaaaccttgaagataa          cttcacagagtttggctacacgattagctccaatttataaagcagtatgctccagactt          accaaaaatcaggtggaatatagaaaatgttggccgagaatgtcggcttggcagcaaggaa          ggtcgacccttctctggggtcactgcttgcctggacttctgtgctcatcccacagggga          cattcacaacatgaataatggaagcactgtggtttgtaccttaactcgagaagataacc          gctctttgggtgttattcctcaagatgagcagctccatgtgctacctctttataagctt          tcagacacagatgagtttggctccaaggaaggaatggaagccaagatcaaatctggggc          catcgaggtcctggcaccgccgcaaaaaaagaacgtgtttcactcagcctgttcccc          gttctggaagaagagggtcgcgatgatgacagaggttcttgacataagataagggca          gtggaaaagaaacctattccccgaatcaagcgggaagaataactcaacaacaacaaaca          cagtaagccttcgtcactgccaaccttagggagttaactgagaccgtgcaacctgaag          taaaaagtgaaccgaacccccattttatcttaaaaagttcagacaacactaaaacttat          tcgctgatgccatccgctcctcaccagtgaaagaggcatctccaggettctcctggtc          cccgaagactgcttcagccacaccagctccactgaagaatgacgcaacagcctcatgog          ggttttcagaagaagcagcactccccactgtacgatgcttcgggaagactcagtggt          gccaatgctcagctgctgatgccctggcatttcacagcttggcgaagtgctcctct          ccccaccctgtctgctcctgtgatggagccctcattaatctgagccttccactggtg          tgactgagccgctaacgcctcatcagccaaaccaccagccctccttctcactcctct          caagaccttgccctcttccaatggaagaagatgagcagcattctgaagcagatgagcc          tccatcagacgaaccctatctgatgacccctgtcaactgctgaggagaaattgcccc          acattgatgagtttggtcagacagtgagcacaatctttttggatgcaaatattggtggg          gtggccatcgcacctgctcagggctcggttttgattgagtggtgcccggcgagagctgca          cgctaccactcctggtgagcaccceaacctaatcatccaaccgctctccttgctct          tttaccagcacaacaaacctaaataagcccccaacatggttttgaactaaacaagattaag</p>	<p>TET1 catalytic domain (nt)</p>

	<p>tttgaggctaagaagctaagaataagaaaatgaaggcctcagagcaaaaagaccaggc agctaatagaaggtccagaacagtcctctgaagtaaatgaattgaaccaaattccttctc ataaagcattaacattaacccatgacaatggtgtcacogtgtccccttatgctctcaca cacggttgcggggcCCTATAACCATTGGGTC</p>	
<p>93</p>	<p>LPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGK SSHGCP IAKWVLRSSDEEKVLCVLRQRTGHHCP TAVMVVLI MVWDG I PLPMADRLYTE LTENLKS YNGHPTDRRCTLNENRTCTCQGIDPETCGASF SFGCSWSMYFNGCKFGRSPS PRRFRIDPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECLRGSKE GRPFSGVTA CLDFCAHPRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPPLYKL SDTDEFGSKEGMEAKIKSGAIEVLAPRRKRKRTCTQPVPRS GKRAAMTEVLAHKIRA VEKKPIPRIKRKNSTTTNNSKPSLLPTLGSNTE TVQPEVKSETEPHF ILKSSDNTKTY SLMP SAPHVPKEASPGFSWSPKTASATPAPLKN DATASCGFSERSSTPHCTMP SGRLSG ANAAAADGPGI SQLGEVAPLPTLSAPVMEPLINSEPS TGVTEPLTPHQPNHQPSFLTSP QDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWS DSEHIFLDANIGG VAIAPAHGSLVIECARRELHATTPVEHPNRNHP TRLSLVFYQHKNLNKPQHGFELNKIK FEAKEAKNKKMKA SEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDNVVTVSPYALT HVAGPYNHWW</p>	<p>TET1 catalytic domain (aa)</p>
<p>94</p>	<p>GACAAGAAGTACTCCATTGGGCTCGCCATCGGCACAAAACAGCGTCGGCTGGGCCGTCAT TACGGACGAGTACAAGGTGCCGAGCAAAAAATTC AAAGTTCTGGGCAATACCGATCGCC ACAGCATAAAGAAGAACCTCATTGGCGCCCTCTGTTCGACTCCGGGGAAACCGCCGAA GCCACGCGGCTCAAAGAACAGCACGGCGCAGATATACCCG CAGAAAGAATCGGATCTG CTACCTgcaGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTCTTCCATA GGCTGGAGGAGTCTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTT GGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAG GAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGC GGTGATCTATCTCGCGCTGG CGCATATGATCAAATTTCCGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAAC AGCGATGTCGACAAACTCTTTATCCA ACTGGTTCAGACTTACAATCAGCTTTTCGAAGA GAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCA AATCCCGGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTG TTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTTTAAATCTAACTCGA CCTGGCCCAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACA ATCTGCTGGCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTG TCAGACGCCATTCTGCTGAGTGATATTTCTGCGAGTGAACACGGAGATCACC AAAGCTCC GCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGA AGGCCCTTGT CAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTCTTCGATCAGTCT AAAAATGGCTACGCCGATACATTGACGGCGGAGCAAAGCCAGGAGGAATTTTACAAAT TATTAAGCCCATCTTGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACA GAGAAGATCTGTGCGCAAACAGCGCACTTTCGCAATGGAAGCATCCCCACCACTT CACCTGGGCGAACTGCACGCTATCTCAGCGGCAAGAGGATTTTACCCCTTTTGAA AGATAACAGGGAAAAGATTGAGAAAATCTCACATTTTCGGATACCC TACTATGTAGGCC CCCTCGCCCGGGGAAATTCAGATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATC ACTCCCTGGA ACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGCCAGTCTTCATCGA AAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGGTGCTTCTTAAACACTCTC TGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTACAGAA GGGATGAGAAAGCCAGCATTTCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCT CTTCAAGACGAACCGAAAAGTTACCGTGA AACAGCTCAAAGAAGACTATTTCAA AAAAGA TTGAATGTTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTG GGAACGTATCACGATCTCCTGAAAATCATTAAGACAAGGACTTCTTGACAATGAGGA GAACGAGGACATTTCTTGAGGACATTGCTCTACCCTTACGTTGTTTGAAGATAGGGAGA TGATTGAAGAACGCTTGAAA ACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAG CTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAA ACTGATCAATGGgat ccGAGACAAGCAGAGTGGAAAGACAATCTGGATTTTCTTAAGTCCGATGGATTTGCCA ACCGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAG AAAGCACAAAGTTTCTGGCCAGGGGACAGTCTTACGAGCACATCGCTAATCTTGCAGG TAGCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCTGATGAACTCGTCA AAGTAATGGGAAGGCATAAGCCC GAGAATATCGTTATCGAGATGGCCCGAGAGAACCAA ACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTAT AAAAGA ACTGGGGTCCCAAATCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGA ATGAGAAGCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAA CTGGACATCAATCGGCTCTCCGACTACGACTGAGTGGATGCCATCGTGGCCAGTCTTTCT CAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAAAATAGAGGGAAGA</p>	<p>dSpCas9 (nt)</p>



	<p>GTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTG                  CTGAACGCCAAACTGATCACACAACGGAAGTTTCGATAATCTGACTAAGGCTGAACGAGG                  TGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTGAGACACGCC                  AGATCACCAA<sup>g</sup>acGTGGCCCAAATTCTCGATTACGCATGAACACCAAGTACGATGAA                  AATGACAAACTGATTTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGA                  TTTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGC                  ATGATGCCTACCTGAATGCAGTGGTAGGCACTGCCTTATCAAAAAATATCCCAAGCTT                  GAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAA                  GTCTGAGCAGGAAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGA                  ATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTTCGGAAGCGACCCTTATC                  GAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTTCGCGACAGT                  CCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCG                  GAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGC                  AAAAAAGATTGGGACCCCAAGAAATACGGCGGATTTCGATTCTCCTACAGTCGCTTACAG                  TGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAGCGTCAAGG                  AACTGTGGTGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTT                  CTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTA                  CTCTCTCTTTGAGCTTGAAAACGGCCGGAACGAATGCTCGCTAGTGCGGGCGAGCTGC                  AGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGC                  CACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCTGTGGA                  ACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAG                  TGATCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTTGCTTACAATAAGCACAGGGAT                  AAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGG                  CGCGCCTGCACCTTCAAGTACTTCGACACCACCATAGACAGAAAAGCGGTACACCTCTA                  CAAAGGAGGTCTTGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACA                  AGAATCGACCTCTCTCAGCTCGGTGGAGAC</p>	
<p>95</p>	<p>DKKYSITGLATIGTNSVGVAVITDEYKVP SKKFKVILGNTDRHS IKKNI.TGAI.I.FDSGFTAF                  ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP IF                  GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN                  SDVDKLF IQLVQTYNQLFEEENP INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGL                  FGNLIALSLGLTPNFKSNFDLAEDAQLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNL                  SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS                  KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQI                  IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI                  TPWNFEVVVDKGSASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE                  GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL                  LGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKYAHLFDDKVMKQ                  LKRRYTGWGRINGIRDKQSGKTLIDLFLKSDGFANRNFMQLIHDDSLTFKEDIQ                  KAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIV IEMARENQ                  TTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQE                  LDINRLSDYDVAIVPQSFLKDDSIDNKVLRSDKNRGSNDNVP SEEVVKKMKNYWRQL                  LNAKLIITQRKFDNLTKAERGGSELKAGFIKRQLVETRQITKHVAQILD SRMNTKYDE                  NDKLIREVKVIITLKSCLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVG TALIKKYPKL                  ESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI                  ETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIAR                  KKDWDPKKYGGFDSPTVAYSVLVVAKEVEKSKKLSVKELLGITIMERS SFEKPNIDF                  LEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALP SKYVNFYLYLAS                  HYEKLGSPEDNEQQLFVEQHKKHYLDEIEQISEF'SKRVI LADANLDKVL SAYNKHRD                  KPITREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYET                  RIDLSQLGGD</p>	<p>dSpCas9 (aa)</p>
<p>96</p>	<p>MDKKYSIGLDIGTNSVGVAVITDEYKVP SKKFKVILGNTDRHS IKKNIIGALLFDSGETA                  EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP I                  FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD                  NSDVDKLF IQLVQTYNQLFEEENP INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGL                  LFGNLI ALSLGLTPNFKSNFDLAEDAQLQLSKD TYDDDLNLLAQIGDQYADLFLAAKN                  LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ                  SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQ                  IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET                  ITPWNFEVVVDKGSASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE                  EGMKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS                  LGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKYAHLFDDKVMK</p>	<p>WT Sp Cas9 (aa)</p>

	<p>QLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDFANRNFMLI HDDSLTFKEDI                  QKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENI VIEMAREN                  QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQ                  ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLRSDKNRGSNDNVPSEEVVKMKMKNYWRQ                  LLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYD                  ENDKLIREVKVITLTKSLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVTALIKKYPK                  LESEFVYGDYKVIYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPL                  IETNGETGEIVWDKGRDFATVRKVL SMPQVNI VVKTEVQTGGFSKESILPKRNSDKLIA                  RKKDWDPKKYGGFDSPTVAYSVLVVAKEKGSKKLKSVKELLGITIMERS SFEKNPID                  FLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA                  SHYEKLGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKH                  DKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYE                  TRIDLSQLGGD</p>	
<p>97</p>	<p>AAGCGAACTACATCCTGGGCCTGGCCATCGGCATCACCAGCGTGGGCTACGGCATCAT                  CGACTACGAGACACGGGACGTGATCGATGCCGGCGTGC GGCTGTTC AAAGAGGCCAACG                  TGAAAACACAGGGCAGGCGGAGCAAGAGGGCGCAGAACAGGCTGAAGCGGGCAGG                  CGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCA                  CAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGC                  TGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCAC                  AACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAG                  CCGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGA                  AGAAAGACGGCGAAGTGC GGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAA                  GAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCC TACCACCAGCTGGACCAGAGCTT CAT                  CGACACCTACATCGACCTGCTGGAACCCGGCGACCTACTATGAGGGACCTGGCGAGG                  GCAGCCCCCTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGC                  ACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAA                  CGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAAT                  ATTACGAGAAGTTCAGATCATCGAGAACGTGT TCAAGCAGAAGAAGAAGCCACCCCTG                  AAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGAC                  CAGCACCGGCAAGCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGACATTA                  CCGCCCCGAAAGAGATTATTGAGAACCCTGAGCTGCTGGATCAGATTGCCAAGATCTG                  ACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCT                  GACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACC                  TGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAG                  ATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCC CAGCAGAA                  AGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCT                  TCATCCAGAGATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGAC                  ATCAITATCGAGCTGGCCCGCAGAGAAGAACTCCAAGGACGCCAGAAAATGATCAACGA                  GATGCAGAAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCG                  GCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGC                  AAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAA                  CTATGAGGTGGACCACATCATCCCAGAAAGCGTGTCTTCGACAACAGCTTCAACAACA                  AAGTGCTCGTGAAGCAGGAAGAAgcCAGCAAGAAGGGCAACCGGACCCCATTCAGTAC                  CTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCTGAACTCT                  GGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAAGCGG                  ACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCGA                  TACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGA                  CGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGGGGAAGTGGAAAGT                  TTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCC                  AACGCCGATTTTCATCTTCAAAGAGTGAAGAAACTGGACAAGGCCAAAAAAGTGATGGA                  AAACCATGATGTTTCGAGGAAAAGCAGGCCCAGAGCATGCCCGAGATCGAAACCGAGCAGG                  AGTACAAGAGATCTTCATCACCCCCCAGATCAAGCACATTAAGGACTCAAGGAC                  TACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCCT                  GTACTCCACCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCC                  TGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTG                  CTGATGTACCACCAGCACCCTCAGACCTACCAGAAACTGAAGCTGATTATGGAACAGTA                  CGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACCTACCTGACCA                  AGTACTCCAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAA                  CTGAACGCCCATCTGGACATCACCCGACTACCCCAACAGCAGAAACAAGTCTGTGAA                  GCTGICCTTAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCG                  TGACCGTGAAGAATCTGGATGTGATCAAAAAGAAAACCTACTACGAAGTGAATAGCAAG</p>	<p>dSaCas9 (nt)</p>

	TGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTC CTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCCTGA ACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTAC CTGGAAAACATGAACGACAAGAGGCCCCCCCAGGATCATTAAAGACAATCGCCTCCAAGAC CCAGAGCATTAAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTA AGAAGCACCTCAGATCATCAAAAAGGGC	
98	KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRSINRFKTSYVK EAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYEYKFOI IENVFQKQKKPTL KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEI IENAELLDQIAKIL TIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQ IAIFNRLKLVPPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQS IKVINAI IKKYGLPND II IELAREKNSKDAQKMINEMQKRNRQTNERIEE I RTTGKENAKYLIEKIKLHDMQEG KCLYSLEAIPLEDLLNNPFNYEVDHI I PRSVSFDNSFNKVLVKQEEASKKGNRTPFQY LSSSDSKISYETFKKHILNLAKGKGRI SKTKKEYLLEERDINRF SVQKDF INRNLDVTR YATRGLMNNLLRSYFRVNNLDVKVKS INGGFTSFLRRKWKFKKERNKGYKHAEDALI IA NADFIKKEWKLLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKD YKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKL LMYHHDPTQYQKLLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNK LNAHLDI TDDYPNSRNKVVKLSLKP YRFDVYLDNGVYKFVTVKNLDVIKKENYEVNSK CYEEAKKLLKISNQAEIFIASFYNNDLIKINGELRYVIGVNNDDLNR I EVNMIDITYREY LENMNDKRPPRI I KTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKKG	dSaCas9 (aa)
99	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR RRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGV HNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRSINRFKTSYV KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGH CTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYEYKFOI IENVFQKQKKPT LKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEI IENAELLDQIAKIL LTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDN QIAIFNRLKLVPPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQS IKVINAI IKKYGLPN DII IELAREKNSKDAQKMINEMQKRNRQTNERIEE I RTTGKENAKYLIEKIKLHDMQEG GKCLYSLEAIPLEDLLNNPFNYEVDHI I PRSVSFDNSFNKVLVKQEEASKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRI SKTKKEYLLEERDINRF SVQKDF INRNLDVT RYATRGLMNNLLRSYFRVNNLDVKVKS INGGFTSFLRRKWKFKKERNKGYKHAEDALI I ANADFIKKEWKLLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKD DYKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEK LLMYHHDPTQYQKLLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGN KLNAHLDI TDDYPNSRNKVVKLSLKP YRFDVYLDNGVYKFVTVKNLDVIKKENYEVNS KCYEEAKKLLKISNQAEIFIASFYNNDLIKINGELRYVIGVNNDDLNR I EVNMIDITYRE YLENMNDKRPPRI I KTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKKG	WT Sa Cas9 (aa)
100	AAGCGACCTGCCGCCACAAAGAAGGCTGGACAGGCTAAGAAGAAGAACTGGAC	NLS sequence 1 (nt)
101	KRPAATKKAGQAKKKLD	NLS sequence 1 (aa)
102	CCAAAGAAGAAGCGGAAGGTC	NLS sequence 2 (nt)
103	PKKKRKV	NLS sequence 2 (aa)
104	AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCCAAAAAGAAAAAG	NLS sequence 3 (nt)
105	KRPAATKKAGQAKKKK	NLS sequence 3 (aa)
106	GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGA TGACAAG	FLAG epitope tag (nt)
107	DYKDHDGDYKDHDIDYKDDDDK	FLAG epitope tag (aa)
108	GCAACAACTTCTCTCTGCTGAAACAAGCCGGAGATGTGGAAGAGAATCCTGGACCG	P2A sequence (nt)

109	ATNFSLLKQAGDVEENPGP	P2A sequence (aa)
110	ATGACCGAGTACAAGCCACGGTGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCCGCCGTTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGTACCCGAGCTGCAAGAACTCTTCCTCACGCGCGTCCGGCTCGACATCGGCAAGGTGTGGGTGCGCGACGACGGCGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCAAGCGGGGGCGGTGTTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGACGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGACCCGGCCC AAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTGTTCGCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCTTCTACGAGCGGCTCGGCTTCACCGTACCCGCGGACGTGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGA	Puromycin resistance gene (nt)
111	MTEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIERVTELQELFLTRVGLDIGKVVWADDGAAVAVWTTPEVSVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAPHRPKEPAWFLATVGVSPDHQKGLGSAVVLPVGEAAERAGVPAFLETSAPRNLPFYERLGFVTADVEVPEGPRTWCMTRKPGA*	Puromycin resistance gene (aa)
112	DKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAQLQSKDYYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGSASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKYAHLFDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIIDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILOTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQE LDINRLSDYDVHIVPQSFLKDDSIDNKVLTNRSDKNRGSNDNVPSEEVVKMKMYRQQLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIENNYHHAHDAYLNAVVGTA LIKKYPKL ESEFVYGDYKVVYDVRKMIKAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIAR KKDWDPKKYGGFDSPTVAYSVLVVAKEVKGKSKKLKSVKELLGITIMERSSEFEKNPIDF LEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALP SKYVNLYLAS HYEKLGSPEDNEQKQLFVEQHKKHYLDEIEIQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD	WT Sp Cas9 (aa)
113	KRNYILGLDIGITSVGYGIIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVVRGSINRFKTSYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETTRTYEYEGPGEPSFGWKDIKEYEMLMGHC TYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYEYKFI IENVFQKQKPTL KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDKDITARKEI IENAELLDQIAKIL TTYQSSEDIQEELTNLSELTOEEIEQISNLKGYTGTHNLSLKA INLI LDELWHTNDNQ IAIFNRLKLVPKKVDLSQOKEIPTTLVDDFILSPVVKRSFIQS I KVINAI I KKYGLPND I I IELAREKNSKDAQKMINEMQKRNRQTNERIEE I IRTTGKENAKYLI EKIKLHDMQEG KCLYSLEAIPLEDLLNPNFNVEVDHIIPRSVSFDNSFNKVLVKQEENS KKGNRTPFPQY LSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAAEDALIIA NADFIKWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHKIDFKDYKYSHRVDKKNRELINDTLYSTRKDDKGNLTVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLLIMEQYGDEKNPLYKYEEETGNLYLTKYSKKDNGPVIKKIKYYGNK LNAHLDI TDDYPNSRNKVVKLSLKP YRFDVYLDNGVYKFTVVKNLDV I KKENYYEVNSK CYEEAKLKKIISNQAEIFASFYNNDLIKINGELYRVI GVNNDLLNRIEVMIDITYREY LENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	WT Sa Cas9 (aa)
114	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGCACGTTAAGCGACCTGCCGCCACAAAGAAGGCTGGACAGGCTAAGAAGA	TET1-extended linker-dSpCas9

<p>AGAAACTGGACGTTCTGCCACCTgcagctgtcttgatcgagttatacaaaaagacaaa  ggcccatattatcacacaccttggggcaggaccaagtgttgctgctgtcagggaaatcat  ggagaataggtatggtcaaaaaggaacgcaataaggatagaaatagtgtacaccg  gtaagaaggggaaaagctctcatgggtgtccaattgctaagtgggttttaagaagaagc  agtgatgaagaaaaagttctttgtttggctcggcagcgtacaggccaccactgtccaac  tgctgtgatgggtggtgctcatcatgggtgtgggatggcatccctctccaatggccgacc  ggctatacacagagctcacagagaatctaaagtcatacaatgggcaccctaccgacaga  agatgcaccctcaatgaaaatcgtacctgtacatgtcaaggaattgatccagagacttg  tggagcttcattctcttttggctgttcctggagtgtactttaatggctgtaagtttg  gtagaagcccaagcccagaagatltagaattgatccaagctctcccttaccatgaaaaa  aaccttgaagataacttacagagtttggctacacgattagctccaatttataagcagta  tgctccagtagcttaccaaaatcaggtggaatatgaaaatgttgcccagaaatgtcggc  ttggcagcaaggaaggtcgacccttctctggggctactgcttgctggacttctgtgct  catccccacagggacattcacacatgaataatggaagcactgtgggttgtaacctaac  tcgagaagataaccgctctttgggtgttattcctcaagatgagcagctccatgtgctac  ctctttataagctttcagacacagatgagtttggctccaaggaaggaatggaagccaag  atcaaatctggggccatcgaggtcctggcaccctggcgcgcaaaaaaagaacgtgttcac  tcagcctgttccccgttctggaaagaagagggctgogatgatgacagaggttcttgcac  ataagataagggcagtggaagaaacctattccccgaatcaagcggagaataactca  acaacaacaacaacagtaagcctctcgtcactgcaacccttagggagtacactgagac  cgtgcaacctgaagtaaaaagtgaaccgaaccccattttatcttaaaaagttcagaca  aactaaaacttattcgtctgatgccatccgctcctcaaccagtgaagagggcatctcca  ggcttctcctggctcccgaagactgcttcagccacaccagctccactgaagaatgagc  aacagcctcatcggggttttcagaaagaagcagcactccccactgtacgatgcttctgg  gaagactcagtggtgccaatgctgcagctgctgatggccctggcatttcacagcttggc  gaagtggctcctctccccacctgtctgctcctgtgatggagccctcattaattctga  gccttccactggtgtgactgagccgctaaccgctcatcagccaaaccaccagccctcct  tctcacctctcctcaagaccttgctcttctccaatggaagaagatgagcagcattct  gaagcagatgagcctccatcagacgaacccctatctgatgacccctgtcactgctga  ggagaattgccccacattgatgagtattggtcagacagtgagcacatcttttggatg  caaatattggtgggtggccatcgcaactgctcaggtcteggttttgattgagtgtgc  cggcgagagctgcacgtaccactcctgttgagcacccecaaccgtaatacatccaaccg  cctctcccttgtcttttaccagcacaaaaacctaataagccccaacatggttttgaa  taacaagattaagtttgaggctaaagaagctagaataagaaaatgaaggcctcagag  caaaaagaccaggcagctaatgaaggtccagaacagctcctctgaagtaaatgaattgaa  ccaaattccttctcataaagcattaacattaacccatgacaatgttgtcacctgtccc  cttatgctctcacacagttgcgggggcCCTATAACCATTGGGTCGTTAACGGAGGCGG  AGCTCTGGCGCACCCACCAAGTGGAGGGTCTCCTGCGGGTCCCAACATCTACTGA  AGAAGGCACCAGCGAATCCGCAACGCCGAGTCAGGCCCTGGTACCTCCACAGAACCAT  CTGAAGGTAGTGCCTGGTTCCCCAGCTGGAAGCCCTACTTCCACCGAAGAAGGCACG  TCAACCGAACCAAGTGAAGGATCTGCCCTGGGACCAGCACTGAACCATCTGAGCCAAA  GAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACTCCATTG  GGCTCGCCATCGGCACAAACAGCGTCCGGCTGGGCGGTCAATACGGACGAGTACAAGGTG  CCGAGCAAAAAATTCAAAGTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCT  CATTGGCGCCTCCTGTTTCGACTCCGGGAAACCGCCGAAGCCACGCGGCTCAAAAGAA  CAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTgcaGGAGATCTTT  AGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCTTTTT  GGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGG  TGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGT  ACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTTCG  GGGACACTTCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTGCAAAACTCT  TTACTCAACTGGTTCAGACTTACAATCAGCTTTTTCGAAGAGAACCCGATCAACGCATCC  GGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAA  CCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCC  TGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAG  CTTCAACTGAGCAAAGACACCTACGATGATGATCTUGACAATCTGCTGGCCAGATCGG  CGACCAGTACGCAGACCTTTTTTTGGCGGCAAGAACCTGTGACAGCCATTTCTGCTGA  GTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATC  AAGCGTATGATGAGCACCAACCAAGACTTGACTTTGCTGAAGGCCCTTGTGACAGCA  ACTGCCTGAGAAGTACAAGGAAATTTCTTCGATCAGTCTAAAAATGGCTACGCCGAT  ACATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCATCTTGGAA</p>	<p>(nt)</p>
--	-------------

	<p>AAAATGGACGGCACCCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAA  ACAGCGCACTTTTCGACAATGGAAGCATCCCCACCAGATTACCTGGGCGAACTGCACG  CTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAAGATAACAGGGAAAAAGATT  GAGAAAATCCTCACATTTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTC  CAGATTTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAGG  AAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGAT  AAAAATCTGCCTAACGAAAAGGTGCTTCTTAAACACTCTCTGCTGTACGAGTACTTCAC  AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAAGCCAGCAT  TCCTGTCTGGAGAGCAGAAGAAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAA  GTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGT  TGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCC  TGAAAATCATTAAAGACAAGGACTTCTGGACAATGAGGAGAACGAGGACATTCTTGAG  GACATTGTCTCACCCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAA  AACTTACGCTCATCTCTTCGACGACAAAGTCATGAAAACAGCTCAAGAGGGCGCCGATATA  CAGGATGGGGCGGCTGTCAAGAAAATGATCAATGGgatcCGAGACAAGCAGAGTGGGA  AAGACAATCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACCTCATGCAGTT  GATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAAGTTCTGGCC  AGGGGGACAGTCTTACGAGCACATCGCTAATCTTGCAGGTAGCCAGCTATCAAAAAG  GGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAA  GCCCCGAGAATATCGTTATCGAGATGGCCCAGAGAACCAAACCTACCCAGAAGGGACAGA  AGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCCAA  ATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTA  CTACCTGCAGAACCGCAGGACATGTACGTGGATCAGGAACCTGGACATCAATCGGCTCT  CCGACTACGACGTGGATGCCATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGAT  AATAAAGTGTGACAAGATCCGATAAAAATAGAGGGAAAGAGTGATAACGTCCCCTCAGA  AGAAGTTGTCAAGAAAATGAAAATATTGGCGGCAGCTGCTGAACGCCAAACTGATCA  CACAAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGAT  AAAGCCGGCTTCATCAAAGGCAGCTTGTGAGACACGCCAGATCACCAA<sub>gca</sub>cGTGGC  CCAAATCTCGATTACGTCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAG  AGGTGAAAGTTATTACTCTGAAGTCTAAGTCTGAGTGGTCTCAGATTTGAGAAAGGACTTCAG  TTTTATAAGGTGAGAGAGATCAACAATACCACCATGCGCATGATGCC<sub>T</sub>ACTGAAATGC  AGTGGTAGGCACTGCACTTATCAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACG  GAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGC  AAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGAT  TACACTGGCCAATGGAGAGATTCCGGAAGCGACCCTTATCGAAACAAACGGAGAAAACAG  GAGAAATCGTGTGGGACAAGGGTAGGGATTTGCGGACAGTCCGGAAGGTCTGTCCATG  CCGCAAGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAAG  TATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAGATTGGGACCCCA  AGAAATACGGCGGATTTCGATTCTCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAA  GTGGAGAAAGGGAAGTCTAAAAAACTCAAAGCGTCAAGGAACCTGCTGGGCATCACAAAT  CATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTCTCGAGGCGAAAGGATATA  AAGAGGTCAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAA  AACGGCCGGAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAAGGTAACGAGCTGGC  ACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAG  GGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTT  GATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAA  CCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCATCAGGGAGCAGG  CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAG  TACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCTTGACGC  CACACTGATTATCATGATCAATTACGGGGCTCTATGAAAACAAGAATCGACCTCTCTCAGC  TCGGTGGAGACAAAAGGCCGGCGGCCACGAAAAGGCCGGCCAGGCAAAAAGAAAAAG  TGA</p>	
<p>115</p>	<p>MDYKDHDGDYKDHDIDYKDDDDKHVKRPAATKKAGQAKKKKLDVLPCTC SCLDRVIQKDK  GPYYTHLGAGPSVAAVREIMENRYGQKGNIRIEIVVYTGKEGKSSHGCP IAKWVLRRS  SDEEKVLC LVRQRTGHHCP TAVMVVLMVWDGIP LPMADRLYTELTENLKS YNGHPTDR  RCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEK  NLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTA CLDFCA  HPRHD IHNMNNGSTVVCTLTREDNRS LGVIPQDEQLHVLPLYKLSDTDFEFGSKEGMEAK  IKSGAIEVLPARRKKRTCTFQPVPRSGKKRAAMTEVLAHKIRAVEKKPIPRIKRKNNS  TTTNSKPSLPTLGSNTETVQPEVKSETEPHFILKSDNTKTYSLMP SAPHVKEASP  GFSWSPKTASATPAPLKN DATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLG</p>	<p>TET1-  extinded  linker-dSpCas9  (aa)</p>

	<p>EVAPLPTLSAPVMEPLINSEPTSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHS                  EADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECA                  RRELHATTPVEHPNRNHPTRLSLVFYQHKNLNKPKQHGFELNKIKFEAKEAKNKKMKASE                  QKDQANEGPEQSSEVNELNQIPSHKALTLTHDNVTVSPYALTHVAGPYNHVVVNGGP                  SSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT                  STEPSEGSAPGTSTEPSEPKKKRKVG IHGVPAAADKKYSIGLAIGTNSVGVAVITDEYKV                  PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF                  SNEMAKVDDSFHRLEESFLVEEDKKHERHP IFGNI VDEVAYHEKYPTIYHLRKKLVDS                  TDKADRLRIYLA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEFNINAS                  GVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAEDAK                  LQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI                  KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE                  KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFILKDNREKI                  EKILTFRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGGASQSFIERMTNFD                  KNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTRNK                  VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILE                  DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSG                  KTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKK                  GILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKQKNSRERMKRIEKGKELGSQ                  ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSID                  NKVLTRSDKNRGSNDNVPSEEVVKMKNYWRQLLNAKLI TQRKF'DNLTKAERGGGLSELD                  KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQ                  FYKVRINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVVYDVRKMI AKSEQEIG                  KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLMS                  PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK                  VEKGSKLLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIKLPKYSLFELE                  NGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYL                  DEIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFK                  YFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDKRPAATKKAGQAKKKK</p>	
116	<p>GGAGGGCCGAGCTCTGGCGCACCCCCACCAAGTGGAGGGTCTCCTGCCGGGTCCCCAAC                  ATCTACTGAAGAAGGCACCAGCGAATCCGCAACGCCCGAGTCAGGCCCTGGTACCTCCA                  CAGAACCATCTGAAGGTAGTGCGCCTGGT TCCCCAGCTGGAAGCCCTACTTCCACCGAA                  GAAGGCACGTCAACCGAACCAAGTGAAGGATCTGCCCTGGGACCAGCACTGAACCATC                  TGAG</p>	extended linker (nt)
117	<p>GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE                  EGTSTEPSEGSAPGTSTEPSE</p>	extended linker (aa)
118	<p>AACCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC</p>	linker (nt)
119	<p>NPKKKRKVG IHGVPAA</p>	linker (aa)
120	<p>ATGGACTACAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGA                  CGTAGACAACGCTTAAGCGACCTGCCGCCACAAAGAAGGCTGGACAGGCTAAGAAGA                  AGAAACTGGACGTTCTGCCACCTgcagctgtcttgatcgagttatacaaaaagacaaa                  ggcccatattatacacaccttggggcagggaaccaagtggtgctgctgcagggaatcat                  ggagaataggtatggtcaaaaaggaaacgcaataaggatagaaatagtagtgcacccg                  gtaaagaagggaagctctcatgggtgtccaattgctaagtggttttaagaagaagc                  agtgatgaagaaaaagttctttggttggctcgggcagcgtacaggccaccactgtccaac                  tgctgtgatggtggtgctcatcatggtgtgggatggcatccctctccaatggccgacc                  ggctatacacagagctcacagagaatctaaagtatacaatgggcacctaccgacaga                  agatgcaacctcaatgaaaatcgtacctgtacatgtcaaggaattgatccagagacttg                  tggagcttcattctcttttggctgttcatggagtagtactttaatggctgtaagtttg                  gtagaagcccaagcccagaagatttagaattgatccaagctctcccttacatgaaaaa                  aaccttgaagataacttacagagtttggctacacgattagctccaatttataagcagta                  tgctccagtagcttaccaaaatcaggtggaatatgaaaatggtgcccgagaatgtcggc                  ttggcagcaaggaaggtcgaccttctctggggtcactgcttgctggacttctgtgct                  catccccacagggacattcacacatgaataatggaagcactgtggtttgtataac                  tcgagaagataaacgctctttgggtgttattcctcaagatgagcagctccatgtgctac                  ctctttataagctttcagacacagatgagtttggctccaaggaaggaatggaagccaag                  atcaaatctggggccatcgaggtcctggcaccgcccgcaaaaaaagaacgtgtttcac                  tcagcctgttccccgttctggaaagaagagggctgcgatgatgacagaggttcttgcac                  ataagataagggcagtggaagaagaacctattccccgaatcaagcggagaataactca                  acaacaacaacaacagtaagccttctgctcactgccaaccttagggagtaacactgagac                  cgtgcaacctgaagtaaaaagtgaaaccgaaccccattttatcttaaaaagttcagaca</p>	TET1-dSpCas9-573-N intein (nt)

	<p>acactaaaacttattcgcgtgatgccatccgctcctcaaccagtgaagaggcatctcca  ggcttctcctgggtccccgaagactgcttcagccacaccagctccactgaagaatgacgc  aacagcctcatgcggttttcagaaagaagcagcactcccactgtacgatgccttogg  gaagactcagtggtgccaatgctgcagctgctgatggcctggcatttcacagcttggc  gaagtggctcctctccccaccctgtctgctcctgtgatggagcccctcattaattctga  gccttccactggtgtgactgagccgctaacgcctcatcagccaaaccaccagccctcct  tcctcacctctcctcaagacettgctcttctccaatggaagaagatgagcagcattct  gaagcagatgagcctccatcagacgaaccctatctgatgaccccctgtcactctgtga  ggagaaattgccccacattgatgagatattgggtcagacagtgagcacatcttttggatg  caaatattgggtgggtggccatcgcacctgctcacggctcggttttgattgagtgtcc  cggcgagagctgcacgctaccactcctggtgagcaccocaaaccgtaataatccaaccg  cctctccttgtcttttaccagcacaaaaacctaataagccccaacatggttttgaa  taaacagattaagtttgaggctaaagaagctaaagaataagaaaatgaaggcctcagag  caaaaagaccaggcagctaatgaaggtccagaacagtcctctgaagtaaatgaattgaa  ccaaattccttctcataaagcattaacattaacccatgacaatgttgtcacctgtccc  cttatgctctcacacacgttgcggggcCCTATAACCATTGGGTCAACCCAAAGAAGAAG  CGGAAGGTCGGTATCCACGGAGTCCAGCAGCCGACAAGAAGTACTCCATTGGGCTCGC  CATCGGCACAAACAGCGTCGGCTGGGCGTCAATTACGGACGAGTACAAGGTGCCGAGCA  AAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGC  GCCCTCCTGTTGACTCCGGGGAAACCGCCGAAGCCACGCGGCTCAAAAAGAACAGCAGC  GCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTgcaggAGATCTTTAGTAATG  AGATGGCTAAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCTTTTTGGTGGAG  GAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTA  CCATGAAAAGTACCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGATGATA  AGGCTGACTTGGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCCGGGGACAC  TTCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCA  ACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTG  ACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATC  GCACAGCTCCCTGGGGAGAAGAAGAAGCCGCTGTTTGGTAATCTTATCGCCCTGTCACT  CGGGCTGACCCCCAACCTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGTTC AAC  TGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCAGATCGGGCAGCAG  TACGCAGACCTTTTTTTGGCGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATAT  TCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCT  ATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTGACAGCAACTGCCT  GAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGA  CGGCGGAGGAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAATGG  ACGGACCCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC  ACTTTCGACAATTGGAAGCATCCCCACCAGATTACCTGGGCGAACTGCACGCTATCCT  CAGGCGGCAAGAGGATTTCTACCCCTTTTTTGAAGATAACAGGGAAAAAGATTGAGAAAA  TCCTCACATTTCCGATACCCTACTATGTAGGCCCCCTCGCCCGGGAAATTCAGATTC  GCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACCTCGAGGAAGTCGT  GGATAAGGGGGCCTCTGCCAGTCTTCATCGAAAGGATGACTAACTTTGATAAAAAATC  TGCCIAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTAT  AACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTC  TGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCG  TGAAACAGCTCAAAGAAGACTATTTCAAAGGATTGAATGCCTGTCTACGAGACAGAG  ATCCTGACAGTGGAGTATGGCCTGCTGCCAATCGGCAAGATCGTGGAGAAGAGGATCGA  GTGTACCGTGTACTCTGTGGATAACAATGGCAACATCTATACACAGCCCGTGGCACAGT  GGCACGATAGGGGAGAGCAGGAGGTGTTTCAGTATTGCCTGGAGGACGGCAGCCTGATC  AGGGCAACCAAGGACCACAAGTTTCATGACAGTGGATGGCCAGATGCTGCCCATCGACGA  GATTTTCGAGCGGGAGCTGGACCTGATGAGAGTGGATAACCTGCCTAATAGCGGAGCA  GTA AAAACACAGCAGACGGGAGT'GAGT'TTGACCCCAAGAAAAAGAGAAAAGT'GA</p>	
<p>121</p>	<p>MDYKDHDGDYKDHDIDYKDDDDKHVKRPAATKKAGQAkkkkLDVLP TC SCLDRVIQDKD  GPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCP IAKWVLRRS  SDEEKVLC LVRQRTGHHCTAVMVVLMVWDGIP LPMADRLYTELTENLKS YNGHPTDR  RCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEK  NLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCA  HPRHDIHNMNNGSTVVCTLTREDNRS LGVIPQDEQLHVLPLYKLSDTDFEFGSKEGMEAK  IKSGAIEVLPARRKKRTCTFQPVPRSGKKRAAMTEVLAHKIRAVEKKPIPRIKRKNNS  TTTNNKPSLPTLGSNTETVQPEVKSETEPHFILKSDNTKTYSLMP SAPHVKEKASP  GFSWSPKTASATPAPLKN DATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLG</p>	<p>TET1- dSpCas9-573- N intein (aa)</p>



	<p>EVAPLPTLSAPVMEPLINSEPTSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHS          EADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSLIECA          RRELHATTPVEHPNRNHPTRLSLVFYQHKNLNKPKQHGFELNKIKFEAKEAKNKKMKASE          QKDQANEGPEQSSEVNELNQIPSHKALTLTHDNVVTVSPYALTHVAGPYNHWVNPKKK          RKGVIHGVPAAADKKYSIGLAIGTNSVGVAVITDEYKVP SKKFKVLGNTDRHS IKKNLIG          ALLFDSGETAEATRLKRTARRRYTRRNRI CYLQEIFSNEMAKVDDSF FHRLEESFLVE          EDKKHERHPIFGNIVDEVAHYHEKYPTIYHLRKKLVDS TDKADLRLIYLALAHMIKFRGH          FLIEGDLNPDNSDVKLF IQLVQTYNQLFEEENP INASGVDAKA ILSARLSKSRRENLI          AQLPGEKKNLFGNLI AL SIGLTPNFKSNFDLAEDAQLQLSKDTYDDDLNLLA QIGDQ          YADLFLAAKNLSDA ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP          EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLNREDLLRKQR          TFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRF          AWMTRKSEETITPWNFEVVDKGGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYF TVY          NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECLSYETE          ILTVEYGLLP IGKIVEKRIECTVYSVDNNGNIYTOPVAQWHDRGEQEVFEYCLEDGSLI          RATKDHKFMFTVDGQMLPIDEIFERELDLMRVDNLPNSGGSKRTADGSEFEPKKKRKV</p>	
122	<p>GGCCGCGGCGCCGTCGCGCGCC</p>	<p>Cpf1 gRNA J spacer</p>
123	<p>MLSGKKA AAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAE VGP GAVGERT          PRKKEPPRASPPGGLAEP PGSAGPQAGPTVVVPGSATPMETGIAETPEGRRT SRRKRAKV          EYREMD E SLANLSEDEYYSEEERNAKAEKEKLP P P P P P QAPPEEENESEPEEP SGVEGA          AFQSR LPHDRMTS QEAACFPDIISGPQQTQKVLFIRNRTLQLWLDNPKIQLTFEATLQ          QLEAPYNSD TVLVHRVHSYLERHGLINFGIYKRIKPLPTKKTGKVI I I GSGVSGLAAAR          QLQSF GMDVTLLEARDRVGGRVATFRKGNVYADLGAMVVTGLGGNPM AVVSKQVMELA          KIKQKCP LYEANGQAVPKEKDEMEVQEFNRLLEATSYLSHQLD FNVLNKPVSLGQALE          VVIQLQEKHVKDEQIEHWKIVKTQEELKELLNKMVNLKEKIKELHQYKEASEVKPPR          DITAEFLVSKSHRDLTALCKEYDELAETQGKLEEKLQELEANPPSDVYLSRRDRQILDW          HFANLEFANATPLSTLSLKHWDQDDDFEFTGSHLTVRNGYSCVPVALAEGLDIKLN TAV          RQVRYTASGCEVIAVNTRSTSQTFIYKCDAVLCTLPLGLVKQOPP AVQFVPPLEPWKTS          AVQRMGFGNLNKVVLCFDRVFWDP SVNLF GHVGT T ASRGELFLFWNLYKAP ILLALVA          GEAAGIMENI SDDVIVGRCLA I LKGI FGS SAVPQPKETVVSRRWRADPWARGSYSYVAAG          SSGNDYDLMAQPI TPGPS IPGAPQIPRLFFAGEHTIRNYPATVHGALLSGLREAGRIA          DQFLGAMYTLPRQATPGVPAQQSPSM</p>	<p>LSD1 (aa)</p>
124	<p>CTGCCACCTgcagctgtcttgatcgagttatcaaaaagacaaaggccatattatac          acaccttggggcaggaccaagtgttgetgctgtcagggaaatcatggagaataggtatg          gtcaaaaaggaaacgcaataaggatagaaatagtagtgtacaccggtaaagaagggaaa          agctctcatgggtgtccaattgctaagtggttttaagaagaagcagtgatgaagaaa          agttctttgtttgggtcgggcagcgtacaggccaccactgtccaactgctgtgatggtgg          tgctcatcaggtgtgggatggcatccctctccaatggcgcaccggtatcacagag          ctcacagagatctaaagtcatacaatgggcaccctaccgacagaagatgcaccctcaa          tgaaaatcgtacctgtacatgtcaaggaattgatccagagacttgtggagcttcattct          cttttggctgttcattggagtagtactttaatggctgtaagtttggtagaagcccaagc          cccagaagatttagaattgatccaagctctcccttacatgaaaaaaccttgaagataa          cttacagagtttggtacacgattagctccaatttataaagcagtagtccagtagctt          accaaaatcaggtggaat atgaaaatgttgcccagagaatgtcggcttggcagcaaggaa          ggtcgacctctctggggctcactgcttgcctggactctctgctcatccccacagggga          cattcacaacatgaataatggaagcactgtggtttgtacctaaactcgagaagataacc          gctctttgggtgttattcctcaagatgagcagctccatgtgctacctctttataagctt          tcagacacagatgagtttggtccaaggaaggaatggaagccaagatcaaatctggggc          catcgaggtcctggcaccgccgcgcaaaaaaagaacgtgttccactcagcctgttcccc          gttctggaagaagagggctgcgatgatgacagaggttcttgcacataagataagggca          gtggaaaagaacctattccccgaatcaagcggagaagaataactcaacaacaacaacaa          cagtaagcctctcgtcactgccaaccttagggagtaacactgagaccgtgcaacctgaag          taaaaagtgaaaccgaaccctttatcttaaaaagttcagacaacactaaaacttat          tcgctgatgccatccgctcctcaccagtgaaagagggcatctccagcttctcctggtc          cccgaagactgcttcagccacaccagctccactgaagaatgacgcaacagcctcatgcg          ggttttcagaaagaagcagcactccccactgtacgatgecttcgggaagactcagtggt          gccaatgctgcagctgctgatggcctggcatttcacagcttggcgaagtggtcctct          ccccacctgtctgctcctgtgatggagccctcattaatctgagccttccactggtg          tgactgagccgctaaccgctcatcagccaaaccaccagcctccttctcactcctcct          caagaccttgccctcttctccaatggaagaagatgagcagcattctgaagcagatgagcc</p>	<p>N term TET1 catalytic domain (nt)</p>

	<p>tccatcagacgaaccctatctgatgacccctgtcaoctgctgaggagaaattgcccc                  acattgatgagatttgggtcagacagtgagcacatcttttggatgcaaatattggtggg                  gtggccatcgcacctgctcacggctcggttttgattgagtgtgcccgcgagagctgoa                  cgctaccactcctggtgagcaccceaacgtaatcatccaaccgctctcccttgtct                  tttaccagcacaacaaacctaataagccccaacatggttttgaaactaaacaagattaag                  tttgaggctaagaagctaagaataagaaaatgaaggcctcagagcaaaaagaccaggc                  agctaataagaggtccagaacagtcctctgaagtaaatgaattgaaccaaattccttctc                  ataaagcattaacattaaccatgacaatggtgtcacogtgtccccttatgctctcaca                  cacgttgcggggcCCTATAACCATTGGGTC</p>	
<p>125</p>	<p>LPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGK                  SSHGCP IAKWVLRSSDEEKVLCVLRQRTGHHCP TAVMVVLIMVWDG I PLPMADRLYTE                  L TENLKS YNGHPTDRRCTLNENRTCTCQGIDPETCGASF SFGCSWSMYFNGCKFGRSPS                  PRRFRIDPSSPLHEKNLEDNLQSLATRLAP IYKQYAPVAYQNQVEYENVARECRLGSKE                  GRPFSGVTA CLDFCAHPHRDIHNMNNGSTVVCTLTREDNRS LGVIPQDEQLHVLPL YKL                  SDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTCTFTQPVPRSGKKRAAMMTEVLAKIRA                  VEKKPIPRIKRKNNSITTTNSKPSLPTLGSNTETVQPEVKSETEPHF ILKSSDNTKTY                  SLMP SAPHVPKEASPGFSWSPKTASATPAPLKN DATASCGF SERSSTPHCTMPSGRLSG                  ANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPTSTGVTEPLTPHQPNHQPSFLTSP                  QDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWS DSEHIFLDANIGG                  VAIAPAHGSVLIIECARRELHATTPVEHPNRNHP TRLSLVFYQHKNLNK PQHGFELNKIK                  FEAKEAKNKKMKASEQKDQAA NEGPEQSSEVNELNQIPSHKALTLTHDNVVTVSPYALT                  HVAGPYNHWV</p>	<p>N term TET1 catalytic domain (aa)</p>
<p>126</p>	<p>GACAAGAAGTACTCCATTGGGCTCGCCATCGGCACAAACAGCGTCGGCTGGGCCGT CAT                  TACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCC                  ACAGCATAAAGAAGAACCTCATTGGCGCCCTCTGTTGACTCCGGGGAAAACCGCGAA                  GCCACGCGGCTCAAAGAACAGCACGGCGCAGATATACCCGAGAAAAGAAATCGGATCTG                  CTACCTgcaggagatcttttagtaatgagatggctaaaggtggatgactctttctccata                  ggctggaggagtccttttgggtggaggagataaaaagcacgagcgccaccaatcttt                  ggcaatatcgtggacgaggtggcgtaccatgaaaagtacccaaccatataatcatctgag                  gaagaagctttagacagtagctgataaggctgacttgcggttgatctatctcgcgctgg                  cgcaatagatcaaatttcggggacacttctcactcgagggggacctgaaccacgacacaac                  agcgatgtcgacaaactctttatccaactgggttcagacttacaatcagcttttcgaaaga                  gaacccgatcaacgcattcggagttgacgccaagcaatcctgagcgctaggctgtcca                  aatccccggcggctcgaaaacctcactcgacagctccctggggagaagaagaacggcctg                  tttggtaactttatcgccctgtcactcgggctgacccccaaactttaaacttaacttcga                  cctggccgaagatgccaagcttcaactgagcaaaagacacctacgatgatgatctcgaca                  atctgctgcccagatcggcgaccagtagcagaccttttttggcggcaaaagaacctg                  tcagacgctattctgctgagtgatattctgagtgaaacacggagatcaccaaagctcc                  gctgagcgctagtatgatcaagcgctatgatgagcaccaccaagacttgactttgctga                  aggcccttgtcagacagcaactgcctgagaagtacaaggaaattttcttcgatcagctct                  aaaaatggctacgccgatacattgacggcggagcaagccaggaggaattttacaaatt                  tattaagccatcttgaaaaaatggacggcaccgaggagctgctggtaaagcttaaca                  gagaagatctgttgcgcaaacagcgcactttcgacaatggaagcatccccaccagatt                  cacctggcgaactgcacgctatcctcaggcggcaagaggatttctacccctttttgaa                  agataacagggaaaagattgagaaaaatcctcacatttcggataccctactatgtaggcc                  cctcgccccgggaaaattccagattcgcgtggatgactcgcaaatcagaagagaccatc                  actccctggaacttcgaggaagtcgtggataagggggcctctgcccagctccttcacga                  aaggatgactaaactttgataaaaaatctgcctaacgaaaaggctgcttctaaactctc                  tgctgtacgagtacttcacagtttataacgagctcaccaggtcaaatacgtcacagaa                  gggatgagaaagccagcattcctgtctggagagcagaagaaagctatcgtggacctcct                  tttcaagacgaaccggaaagttaccgtgaaaacagctcaaaagaagactatttcaaaaaaga                  ttgaa</p>	<p>N term dSpCas9 N terminal fragment (nt)</p>
<p>127</p>	<p>DKKYSIGLAI G TNSVGVAVITDEYKVP SKKFKVLGNTDRHS IKKNLIGALLFDSGETAE                  ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHP IF                  GNIVDEVAYHEKYPTIYHLRKKLVSDTKADRLRIYLALAHMIKFRGHFLIEGDLNPDN                  SDVDKLF IQLVQTYNQLFEENP INASGVDAKA ILSARLSKSRRL ENLIAQLPGEKKNGL                  FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDYDDDLNLLAQIGDQYADLFLAAKNL                  SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS                  KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI                  HLGELHA I LRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEETI                  TPWNFEVVDK GASAQSFIERMTNFDKNLPNEKVLPKHSLLEYEFTVYNELTKVKYVTE</p>	<p>N term dSpCas9 N terminal fragment</p>

	GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE	
128	TGCCTGTCTACGAGACAGAGATCCTGACAGTGGAGTATGGCCTGCTGCCAATCGGCAA GATCGTGGAGAAGAGGATCGAGTGTACCGTGTACTCTGTGGATAACAATGGCAACATCT ATACACAGCCCGTGGCACAGTGGCACGATAGGGGAGAGCAGGAGGTGTTTCGAGTATTCG CTGGAGGACGGCAGCCTGATCAGGGCAACCAAGGACCACAAGTTCATGACAGTGGATGG CCAGATGCTGCCCATCGAC	N term Npu Intein (nt)
129	CLSYETEILTVEYGLLP I GKIV EKRIE CTVYSVDNNGNIYTPVAQWHRGEQEVFEYC LEDGSLIRATKDHKFM TVDQMLPID	N term Npu Intein (aa)
130	ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCAT CAAGATTGCTACACGGAAAATACCTGGGAAAAGCAGAACGTGTACGACATCGGCCTGGAGC GGGATCACAACCTTCGCCCTGAAGAATGGCTTTATCGCCAGCAATTGTTTCGACTCTGTT GAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCAGCATCTCCT GAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGG ACATTGTCCTCACCCITACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAA ACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATAC AGGATGGGGGGCGGTGTCAAGAAAATGATCAATGGgat cCGAGACAAGCAGAGTGGAA AGACAATCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACCTCATGCAGTTG ATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAAGCACAAGTTTCTGGCCA GGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGG GAATACTGCAGACCGTTAAGGTCTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAG CCCAGAAATATCGTTATCGAGATGGCCGAGAGAACCAAACTACCCAGAAGGGACAGAA GAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCCAAA TCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTAC TACCTGCAGAAGCGCAGGGACATGTACGTGGATCAGGAACCTGGACATCAATCGGCTCTC CGACTACGACGTGGATGCCATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATA ATAAAGTGTGACAAGATCCGATAAAAATAGAGGGAAGAGTGATAACGTCCCTCAGAA GAAGTTGTCAAGAAAATGAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTGATCAC ACAACGGAAGTTTGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATA AAGCCGGCTTCATCAAAAGGCAGCTTGTGAGACACGCCAGATCACCAAagcacGTGGCC CAAATTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTTCGAGA GGTAAAGTTATTACTCTGAAGTCTAAGTCTGGTCTCAGATTTTCAGAAAGGACTTTCAGT TTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCA GTGGTAGGCACTGCACCTATCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGG AGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAAATAGGCA AGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATT ACACTGGCCAATGGAGAGATTCCGAAGCGACCATTATCGAAAACAAACGGAGAAAACAGG AGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCTGTCCATGC CGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCCGAGGCTTCTCCAAGGAAAGT ATCCTCCCAGATAAATGAGCAGAAGCAGCTGATCGCACGCAAAAAAAGATTGGGACCCCAA GAAATACGGCGGATTCGATTCTCTACAGTTCGCTTACAGTGTACTGGTTGTGGCCAAAG TGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATC ATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTCTCGAGGCGAAAGGATATAA AGAGGTCAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTGAGCTTGAAA ACGGCCGGAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAAGGTAACGAGCTGGCA CTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGG GTCTCCCAGATAAATGAGCAGAAGCAGCTGTTTCGTGGAACAACACAAACACTACCTTG ATGAGATCATCGAGCAAATAAGCGAATTTCCAAAAGAGTGATCCTCGCCGACGCTAAC CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGC AGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGT ACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCTCGGACGCC ACACTGATTTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCT CGGTGGAGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGT GA	dSpCas9-573- C Intein (nt)
131	MKRTADGSEFESPKKKRKVIKIATRKYLGKQNVYDIGVERDHNFAKNGFIASNCFDVS EISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDI LEDIVLTLTLFEDREMIERLK TYAHLFDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDFANRNFML IHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA I KKGILQTVKVVDLVKVMGRHK PENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLRSDKNRKGSDNVPSE EVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGSELDKAGFI KRQLVETROITKHVA QILDSRMNTKYDENDKLI REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA	dSpCas9-573- C Intein (aa)

	VVGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLKSVELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK	
132	ATCAAGATTGCTACACGGAAATACCTGGGAAAGCAGAACGTGTACGACATCGGCGTGGA CGGGATCACAACTTCGCCCTGAAGAATGGCTTTATCGCCAGCAAT	C term Npu Intein (nt)
133	IKIATRKYLGKQNVYDIGVERDHNFA LKNGFIASN	C term Npu Intein (aa)
134	TGTTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGACTTCTTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCTTACGTTGTTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCATCTCTTCGACGACAAAAGTCATGAAAACAGCTCAA GAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACCTGATCAATGGgat cCGAGACAAGCAGATGGAAAGACAATCCTGGATTTTCTTAAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAACACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACCTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACCTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGGAAGCTTACCTGTACTACCTGCAGAACGGCAGGACATGTACGTGGATCAGGAACCTGACATCAATCGGCTCTCCGACTACGACGTGGATGCCATCGTGCCCGAGCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAAATAGAGGGAAAGAGTGATAACGCTCCCTCAGAAGAAGTTGTCAAGAAAATGAAAATTTATTGGCGGCAGCTGCTGAA CGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCC TGCTGAGTTGGATAAAGCCGGCTTCATCAAAAAGGCAGCTTGTGAGACACGCCAGATC ACCAAGc acGTGGCCCAAATTTCTCGATTACGCATGAACACCAAGTACGATGAAAAATGACAAACTGATTCCGAGAGGTGAAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCA GAAAGGACTTTTCAGTTTTATAAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACCTGCACCTTATCAAAAAATATCCCAAGCTTGAATC TGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTT TCAAGACCGAGATTACACTGGCCAATGGAGAGATTCCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCCGCGACAGTCCGGAAGCTCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGC TTCTCCAAGGAAAGTATCCTCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAAGGGAAGTCTAAAAAACTCAAAAAGCGTCAAGGAACTGCTGGGCATCACAAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGA GGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAACGAATGCTCGCTAGTGGGGCGAGCTGCAGAAA GGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTA TGAAGAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCTGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTC CCAAAAAGAGTATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTTGGCGCGC CTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGTCGGTGGAGAC	C term dSpCas9 C terminal fragment (nt)
135	CFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI EERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKV MGRHKPENIV IEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNE KLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLRSDKNRGS DNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKFDFQFYKVIENNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGG	C term dSpCas9 C terminal fragment (aa)

	FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLKSVKEL LGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV LADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVL DATLIHQSI TGLYETRIDLSQLGGD	
136	DALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSLP SASVEFEGSGGPPSGQISNQALALAPSSAPVLAQTMVPPSSAMVPLAQPAPAPVLT P GPPQSLSA PVPKSTQAGEGTLSEALLHLQFDADL GALLGNSTDPGVFTDLASVDN SEFQQLLNQG VSMSTAEPLMEYPEAITRLVTGSQRPPDPAPTPLGTSGLPNGLSG DEDFSS IADMF SALLSQISSSGQGGGGSGFSVDT SALLDLFSPSVTVPDMSLP DL DSSLASIQELLSPQ EPPRPPEAENSSPDSGKQLVHYTAQPLFLLDPGSVDT GSNDLPVLFELGEGSYFSEGDG FAEDPTISLLTGSEPPKAKDPTVS	VPH (aa)
137	LLPKNYHLENEVARLKKL VGER	SunTag GCN4 peptide (aa)
138	GGSGG	GGSGG linker (aa)
139	PAAKRVKLD	c-myc NLS (aa)
140	RQRRNELKRSP	c-myc NLS (aa)
141	NQSSNFGPMKGGNF GGRSSGPYGGGGQYFAKPRNQQGY	hRNP A1 M9 NLS (aa)
142	RMRI ZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV	importin-alpha IBB domain NLS (aa)
143	VSRKRPRP	myoma T protein NLS (aa)
144	PPKKARED	myoma T protein NLS (aa)
145	POPKKKPL	human p53 NLS (aa)
146	SALIKKKKKMAP	mouse c-abl IV NLS (aa)
147	DRLRR	influenza virus NS1 NLS (aa)
148	PKQKKRK	influenza virus NS1 NLS (aa)
149	RKLKKKIKKL	Hepatitis virus delta antigen NLS (aa)
150	REKKKFLKRR	mouse Mx1 protein MNLS (aa)
151	KRKGDEVDGVDEVAKKSKK	human poly(ADP-ribose) polymerase NLS (aa)
152	RKCLQAGMNLEARKTKK	steroid hormone receptors (human) glucocorticoid NLS (aa)
153	GGGGG	GGGGG linker (aa)
154	GGAGG	GGAGG linker

		(aa)
155	GGGGSSS	GGGGSSS linker (aa)
156	GGGGAAA	GGGGAAA linker (aa)
157	GGGGS	(GGGGS) <sub>n</sub> linker (aa), n is 1 to 10
158	NGG, where N is any nucleotide	PAM - SpCas9, F. Novicida Cas9
159	NNGRRT, where N is any nucleotide, and R is G or A	PAM - SaCas9
160	NNNNGATT, where N is any nucleotide	PAM - N. meningitidis Cas9
161	NNNNRYAC, where N is any nucleotide, R is G or A, and Y is C or T	PAM - C. jejuni Cas9
162	NNAGAAW, where N is any nucleotide and W is A or T	PAM - S. thermophilus Cas9
163	NAAAAC, where N is any nucleotide	PAM - T. denticola Cas9
164	TTTV, where V is A, C, or G	PAM - Cas12a/Cpf1 Cas9
165	NGAN, where N is any nucleotide	Variant PAM - SpCas9 variant
166	NGNG, where N is any nucleotide	Variant PAM - SpCas9 variant
167	NGAG, where N is any nucleotide	Variant PAM - SpCas9 variant
168	NGCG, where N is any nucleotide	Variant PAM - SpCas9 variant
169	MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPLPERAHPEVNGDTKWHSPKSYGIPCMKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLLKQDQKANGERRNFGVSGERNPGESSQPNVSDLSDKKESVSSVAQENAVKDFTSFSTHNCSSGPEPENLEQILNEQEGKSANYHDKNIVLLKNAVLMPNGATVSASSVEHTHGELLEKTLSQYYPD CVSIYVQKTTSHINAINSQATNELSCEITHPSHTSGQINSAQTSNSELPPKPAAVVSEACDADDADNASKLAAMLNTCSFQKPEQLQQQKSVFEICPSPAENNIQGTTKLASGEEFCSGSSSNLQAPGGSSERYLKQNEMNAGAYFKQSSVFTKDSFSATTTPPPPSQQLLSPPPLPQVPQLPSEGKSTLNGGVLEEHHPNQSNTTLLREVKIEGKPEAPPSQSPNPSTHVCSPSPMLSERPQNNVCVNRNDIQTAGTMTVPLCSEKTRPMSEHLKHNPPIFGSSGELQDNCQQLMRNKEQEILKGRDKEQTRDLVPPTQHLYLKPGWIELKAPRFHQAESHLKRNEASLPSILQYQPNLSNQMTSKQYTGNSNMPGGLPROAYTOKTTOLEHKSQMYQVEMNQGSQGTVDQHLQFQKPSHQVHFSKTDHLPKAHVQSLCGTRFHFQQRADSQTEKLMSPVLKQHLNQQAS ETEPFNSHLLQHKPHKQAAQTOPSQSSHLPQONQQQQKQIKNKEEILQTFPHPQSNN DQQREGSFFGQTKVEECFHGENQYSKSSEFETHNVQMGLEEVQINRRNSPYSQTMKSS ACKIQVSCSNTHLVSENKEQTTHELFAGNKTQNLHMQYFPNNVIPKQDLLHRCFQEQEQKSQQASVQLQGYKRNQDMSGQAAQLAQQRYLIIHNHANVFPVPDQGGSHQTQTPPQKDTQKHAALRWHLQKQEQQQTQPPQTESCHSQMHRPIKVEPGCKPHACMHTAPPENKTKVTKQENPPASCDNVQKSI IETMEQHLKQFHAKSLFDHKALTLKSQKQVVMESGPFVTLTRQTAAELDSHTPALEQQTTSSEKPTKRRTAASVNLNFIIESPSKLLDTPIKNLLDTPVKIQYDFPSCRVCVEQIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGAIRIERV IYTGKEGKSSQGCP IAKWVVRSSSEKLLCLVRERAGHTCEAAVIVILILVWEGIPLS LADKLYSELTTETLRKYGTLNRRCALNEERTCACQGLDPETCGASFSGCSWSMYNGC KFAFSKIPRKFLLGDDPKKEEKLESHLQNLSTLMAPTYKKLAPDAYNNQIIBEYHRAPE CRLGLKEGRPFSGVTACLDFCAHAHRDLHNMQNGSTLVCTLTREDNREFGGKPEDEQLH VLPPLYKVSVDVDFGVSVEAQEEKRSQAIQVLSFRRKVRMLAEPVKTCRQRKLEAKKAA AEKLSLENSNKNKEKESAPSRTKQTENASOAKOLAEALLRLSGPVMQSOQOPOPLOKQ PPQPQQQQRPPQQQPHHPQTESVNSYSASGSTNPNMRRPNPVPSPYPNSHTSDIYGSTS	TET2

	PMNFYSTSSQAAGSYLNSNPMNYPYGLLNQNTQYP SYQCNGNL SVDNCSPYLGSYSPO SQPMDLYRYP SQDPLSKLSLPP IHTLYQPRFGNSQSFTSKYLGYGNQNMQGDGFS SCTI RPNVHHVGLKPPYP THEMDGHFMGATSRLPPNLSNPNMDYKNGEHHSP SHI IHNYSAAP GMFNSSLHALHLQNKENDMLSHTANGLSKMLPALNHDRTACVQGGLHKLSDANGQEKQP LALVQGVASGAEDNDEVWSDSEQSFLDPD IGGVAVAP THGS ILIECAKRELHATTP LKN PNRNHP TRI SLV FYQH KSMNEPKHGLALWEAKMAEKAREKEEECEKYGPDYVPQKSHGK KVKREPAEPHETSEPTYLRFIKSLAERTMSVTTDSTVTTSPYAFTRVTPYNYRI	
170	MSQFQVPLAVQPDLPGLYDFPQRQVMVGSFPGSLSMAGSE SQRGGGDGRKKRRCGT CEPCRELSGACTSCTNRRTHQICKLRKCEVLKKKVGLLKEVEIKAGEGAGPWQOGAA VKTGSELSPVDGVPVQMDSGPVYHGDSRQLSASGVPVNGAREPAGP SLLGTGGPWRVD QKPDWEAAPGPAHTARLEDAHDLVAFSAVAEAVSSYGALSTRLYETFNREMSREAGNNS RGPRPGPEGCSAGSEDLDTLQ TALALARHGMKPPNCNCDGPECPDYLEWLEGKIKSVVM EGGEERPLRPGPLPPGEAGLPAPSTRPLLSSEVPQISPQEGLELSQSALSIAKEKNI SL QTAIAIEALTQLSSALPQPSHSTPQASCPLPEALSPPAPFRSPQSYLRAPSWPVVPEE HSSFAPDSSAFPPATPRTEFPEAWGTDTPPATPRSSWPMRPPSPDPAELEQLLGSASD YIQSVFKREALPTKPKVKVEAPSSSPAPAPSPVLQREAPTPSSEPDTHQKAQTALQOH LHHKRSFLLEQVHDTSFAPSEPSAPGWWPPSPVPRLPDRPPKEKKKLP TPAGGPV GTEKAAPGIKPSVRKPIQIKKSRPREAQPLFPPVRQIVLEGLRSPASQEVQAHPPAPLP ASQGSAPVLPPEPSLALFAPSPSRDSLLPPTQEMRSPSPMTALQPGSTGPLPADDKLE ELIRQFEAEFGDSFGLPGPPSVP IQDPENQOTCLPAPESPATRSPKQIKIESSGAVTV LSTTCFHSEEGQEATPTKAENPLPTLSGFLESPLKYLDTPTKSLLDTPAKRAQAEFP TCDCVEQIVEKDEGPYYTHLGSQPTVASIRELMEERYGEKGAIRIEKVIYTGKEGKSS RGCP IAKWVIRRHTEELKLLCLVRHRAGHHCQNAVIVILILAWEGIPRSLGDTLYQELT DTLRKYGNPTSRRCGLNDDRTACQCKDNTCCGASFSGCSWSMYFNGCKYARSKTRK FRLAGDNPKEEEVLRKSFQDLATEVAPLYKRLAPQAYQNQVTNEEIAIDCRGLLKEGRP FAGVTACMDFCAHAHKDQHNLINGCTVVC TLTKEDNRCVGIPEDEQLHVLPLYKMANT DEFGSEFNQNAKVGSGA IQVI.TAFPRFVRRIPFPAKSCRQRQI.FARKAAAFK KKI QKFK LSTPEKIKQEALELAGITSDPGLSLKGGLSQQGLKPSLKVEPQNHFS SFKYSGNAVVES YSVLGNCRPSDPYSMNSVYSYHSYYAQPSLTSVNGFHSKYALPFSYGFPPSSNPVFP QFLGPGAWGHS GSGSFEKKPDLHALHNSLSPAYGAEFAELP SQAVP TDAHHP TP HHQ QPAYPGPKEYLLPKAPLLHSVSRDP SPFAQSSNCYNRSIKQEPVDPLTQAEVPVRDAGK MGKTPLESEVSONGGP SHLWQYSGGP SMSPKRTNGVGGSWGVS SSGESPAI VPDKLS SF GASCLAP SHFTDQWGLFPGEGQQAASHSGGRLRGKPWSPCKFGNSTSALAGPSLTEKP WALGAGDFNSALKGPSGFQDKLWNPMKGEEGRI PAAGASQLDRAWQSFGLPLGSSEKLF GALKSEEKLDWPFSL EEGPAEPPSKGAVKEEKG GGGGAE EEEEEELWSD SEHNFLDENIG GVAVAPAHGSI LIECARRELHATTP LKKPNRCHPTRI SLV FYQH KNLNQP NHGLALWEA KMKQLAERARARQEEAARLGLGQOEAKLYGKKRKGWGTVAEPQOKEKKGVPVTRQALA VPTDSAVTVSSYAYTKVTGPYSRWI	TET3
171	DALDDFDLMDLGS DALDDFDLMDLGS DALDDFDLMDLGS DALDDFDLMD	VP64 (aa)
172	PTQAGEGTLSEALLQLQF DDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAP HTTEPMLMEYPEAI TRLV TGAQRPPDPAPAPL GAPGLPNGLLSGDEDF SSIAD MDF SAL LSQISS	p65AD (aa)
173	RDSREGMFLPKPEAGSAI SDVFE GREVCQPKRI RPFHPPGSPWANRPLPASLAP TPTGP VHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPEETSQAVKALREMADTVIPQKEEA AICGQMDLSHPPPRGHLDLDTTLESMTEDLNLD SPLTPELNEI LD TFLNDECLLHAMH ISTGLSIFDTSLF	Rta domain (aa)
174	MAENVVEPGPPSAKRPKLSSPALSASASDGTDFGSLFDLEHDL PDELINSTE LGLTNGG DINQLQTS LGMVQDAASKHKQLSELRLSGSSPNLNMVGGPGQVMASQAQQSSPGLGLI NSMVKSPMTQAGLTPSNMGMGTSGPNQGP TQSTGMMNSPVNQ PAMGMNTGMNAGMNP GM LAAGNGQGIMPVQVMNGS IGAGRGRQNMQYPNPGMGSAGNLLTEPLQQGSPQMGQTGL RGPQPLKMGMMNPNPYGSPYTQNPQQQIGASGLGLQIQTKTVLSNNLSPFAMDKKA VP GGMPNMGQQPAPVQVQPGLVTPVAQGMGSGAHTADPEKRKLIQQQLVLLLHAHKCQR EQANGEVRCNLPHCRTMKNVNLNMTHCQSGKSCQVAHCAS SRQI ISHWKNCTRHD CPV CLPLKNAGDKRNQQP ILTGAPVGLGNPSSLGVGQASAPNLS TVSQIDP SSI ERAYAALG LPYQVNQMP TQPQVQAKNQONQPGQSPQGMRPMNSMSASPMGVNGGVQVTPSLLSDS MLHSAINSQNPMMSENASVPSMGPMPTAAQPS TTGIRKOWHEDI TODLRNHLVHKLVQA IFPTPDPAALKDRR MENLVAYARKVEGDMYESANNRAEYHLLAEKIYKIQKELEEKRR TRLQKQNMLPNAAGMVPVSMNPGPNMQPQPGMT SNGPLPDP SMIRGSVPNQMMPRI TP QSGLNQFGQMSMAQPP I VPRQTPPLQHHGQLAQPGALNPPMGYGPRMQQP SNQGF LPQ TQFP SQGMNVTNIPLAPSSGQAPVSAQMS SSSCPVNSP IMPPGSQGSHI HCPQLPQPA LHQNSPSPVP SRTPTPHHTPPSIGAQQPPATTIPAPVPTPPAMP PGPQSALHPPRQT	p300 domain (aa)

	<p>PTPPTTQLPQQVQPSLPAAPSADQPQQQPRSQOSTAASVPTPTAPLLPPQPATPLSQPA                  VSIIEGQVSNPPSTSSSTEVNSQAIIEKQPSQEVKMEAKMEVDQPEPADTQPEDISESKVE                  DCKMESTETEERSTELKTEIKEEEDQPSTSATQSSPAPGQSKKIFKPEELRQALMPTL                  EALYRQDPESLFFRQPVDPQLLGGIPDYFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDI                  WLMFNNAWLYNRKT SRVYKYCSKLSSEVFEQEIDPVMQSLGYCCGRKLEFSPQTLCCYGK                  QLCTIPRDAYYSYQNRHYHFCEKCFNEIQGESVSLGDDPSQPQT TINKEQFSKRKNDTL                  DPELFVECTEGRKMHQICVLHHEI IWPAGFVCDGCLKKSARTRKENKFSKARLPSTRL                  GTFLENRVNDFLRRQNHPESGEVTVRVVHASDKTVEVKPGMKARFVDSGEMAEFPPYRT                  KALFAFEEIDGVDLCCFFGMHVQEYGSDCPPPNQRRVYISYLDVHFFFRPKCLRTAVYHE                  ILIGYLEYVKKLGYTTGHIWACPPSEGDDYIFHCHPPDQKIPKPKRLQEWYKMLDKAV                  SERIVHDYKDFIKQATEDRLTSAKELPYFEGDFWPNVLEESIKELEQEEEEERKREENTS                  NESTDVTKGD SKNAKKNKKT SKNKSSLSRGNKKKPGMPNVSNDLSQKLYATMEKHKE                  VFFVIRLIAGPAANSLLP IVDPDPLIPCDLMDGRDAFLTLARDKHLEFSSSLRRAQWSTM                  CMLVELHTQSQDRFVYTCNECKHHVETRWHCTVCEYDLCITCYNNTKNHDKMEKLGGLG                  LDDESNNQAAATQSPGDSRRLS IQRCIQSLVHACQCRNANCSLP SCQKMKRVVQHTKG                  CKRKTNGGCPICKQLIALCCYHAKHCQENKCPVFFCLNLIKQKLRQQQLQHRLQQAQMLR                  RRMASMQRTGVVGGQQGLPSPTPATPTTP TGQQPTTPQTPQPTSQPQP TFPNSMPPYLP                  RTQAAGPVSQGAAGQVTPPTPPQTAQPLPGPPPAAVEMAMQIQRAAETQRQMAHVQI                  FQRP IQHQMP PMTPMAPMGMNPPMTRGP SGHLEPGMGPTGMQQQPPWSQGGLPQFQQL                  QSGMPPRAMPMSVAQHGGQLNMAPQPLGLQVGI SPLKPGTVSQQALQNLRLTLRSPSSPL                  QQQQVLSILHANPQLLAAF IKQRAAKYANSNPQPIPGQPGMPQGGPGLQPP TMSGQQGV                  HSNPAMQNNPMQAGVQRAGLPQQQPQQQLQPPMGGMSPQAQQMNMNHTMP SQFRDIL                  RRQQMMQQQQOAGGPGIGPGMANHNQFOQPOGVGYPQQOORMQHMHMQOQNGMGOI                  GQLPQALGAEAGASLQAYQQRLLQQQMGSPVQPNMSPQQHMLPNQAQSPHLQGGQIPN                  SLSNQVRSPQVPSRPRQSQPPHSSPSPRMQPQPSPHHVSPQTS SPHPGLVAAQANPME                  QGHFASPDQNSMLSQLASNPGMANLHGASATDLGLSTDNSDLNLSQSTLIDH</p>	
175	<p>GFSVDTSAI.I.DI.FSPSVTVPMDSI.PDI.DSSIASTQF.LI.SPQFPPRPPFAFNSSPDSGKQ                  LVHYTAQPLFLLDPGSVDTGSNDLPVLFELGEGSYFSEGDFGAEDPTI SLLTGSEPPKA                  KDPTVS</p>	<p>HSF1 domain (aa)</p>
176	<p>DALDDFDLMDLGS DALDDFDLMDLGS DALDDFDLMDLGS DALDDFDLMDLGS PPKKRRKV                  GSQYLPD TDDRHRIEEKRRKRTYETFKSIMKKS PFSGPTDPRPPRRRIAVPSRSSASVPK                  PAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSA                  LAQAPAPVPLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDLGALLGNSTDPFA                  VFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGA                  PGLPNGLLSGDEDFSSIADMDFSALLSQISSGSGSRSRSREGMFLPKPEAGSAISDVF                  EGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGVPHEPVGSLTPAPVPQPLDFAPA                  VTEASHLLEDLDEETSQAVKALREMAADVIPQKEEAACGQMDLSHPPPRGHDELTT                  TLESMTLEDLNDSPLTPELNEILDFTFLNDECLLHAMHISTGLSIFDTSLF</p>	<p>VPR (aa)</p>
177	<p>MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEA                  GKAETSEGSGSAPAVPEASAPKQRRS IIRDRGPMYDDPTLPEGWTRKLRKRSGRSAG                  KYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKPKP                  SPKAPGTGRGRGRPKGSGTTRPKAATSEGQVQKRVLEKSPGKLLVKMPFQTSPGGKAEG                  GGATTSTQVMVIKRPGRKRKAADPQAIPKGRGRKPGSVAAAAAEAKKAVKESIRS                  VQETVLP IKKRKTRETVSIEVKEVVKPLLVLSTLGEKSGKGLKTKCKSPGRKSKESSPKGR                  SSSASSPPKKEHHHHHHHSESPKAPVLLPPLPPPPPEPESSEDFTSPPEPQDLSSSVC                  KEEKMPRGGSLESDGCPKEPAKTQPAVATAATAAEKYKHRGEGERKDIVSSSMRPNRE                  EPVDSRTPVTERVS</p>	<p>Human MeCP2 isoform A</p>
178	<p>SGSETPGTSESATPES</p>	<p>Linker (aa)</p>
179	<p>MKRNYILGLAIGITSVGYGIIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGAARLRKRR                  RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGV                  HNVNEVEEDTGNELSTKEQISRNKALEEKYVAELQLERLKKDGEVRSINRFKTSYV                  KEAKQLLKVQKAYHQLDQSFIDTYIDLLETRTYEYEGPGEPSFGWKDIKEWYMLMGH                  CTYFPEELRSVKYAYNADLYNALNDLNNLVI TRDENEKLEYEYEFQI IENVFKQKKKPT                  LKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI TARKEI IENAELLDQIAKI                  LTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKA INLILDELWHTNDN                  QIAIFNRLKLVPKKVDLSQOKEIPTTLVDDF ILSPVVKRSF IQS IKVINAI IKKYGLPN                  DII IELAREKNSKDAQKMINEMQKRNROTNERIEEI IRTTGKENAKYLIEKIKLHDMQE                  GKCLYSLEAIPLEDLNNPFNYEVDHIIPRSVSFDNSFNKVLVKQEEASKKNRTPFQ                  YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERD INRFVQKDF INRNLDVT                  RYATRGLMNLRSYFRVNNLDVVKVKS INGGFTSFLRRKWKFKKERNKGYKHAEDALI I                  ANADFIKFEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFK</p>	<p>dSaCas9 (aa)</p>



	DYKYSHRVDKPKNRELINDTLYSTRKDDKGNLTVNINLNGLYDKDNDKLLKLLINKSPEK LLMYHHDPQTYQKLLIMEQYGDENPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGN KLNALHDITDDYPNSRNKVVKLSLKPYPYRFVYLDNGVYKFVTVKNLDV IKKENYEVNS KCYEEAKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDDLNR IEVNMIDITYRE YLENMNDKRPPRI I KTIASKTQS I KKYSTD I LGNLYEVKSKKHPQ I I KKG	
180	MDKKYSIGLAIGTNSVGVAVITDEYKVP SKFKVVLGNTDRHS I KKNL I GALLFDSGETA EATRLKRTARRRYTRRNRI CYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP I FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVKLFIQLVQTYNQLFEENP INASGVDAKAIL SARLSKSRLENL I AQLPGEKNG LFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAACKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQ IHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYVVGPLARGNSRFAMTRKSEET ITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNMS LGTYHDLK I IKDKDFLDNEENEDILEDIVLTLTFEDREMIEERLKYAHLFDDKVMK QLKRRRYTGWGRLSRKLIINGIRDKQSGK I LDFLKSDGFANRNFQLIHDDSLTFKEDI QKAQVSGQGD SLHEHIANLAGSPA I KKG I LQTVKVVDELVKVMGRHKPEN I VIEMAREN QTTQKGQKNSRERMKRIE EGIKELGSQLKEHPVENTQLONEKLYLYLQNGRDMYVDQ ELDINRLSDYDVDAIVPQSFLKDDS IDNKVLRSDKNRGKSDNVPSEEVVKKMKNYWRQ LLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYD ENDKLIREVKVI TLKSKLVSDFRKDFQFYKVI INNYHHAHDAYLNAVVG TALIKKYPK LESEFVYGDYKVDVRKMI AKSEQE I GKATAKYFFYSNIMNFFKTEITLANGEIRKRPL IETNGETGEI VWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKES I LPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVAKVEK GKSKKLKSVKELLGITIMERS SFEKNPID FLEAKGYKEVKDL I IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYFKI.KGSPEDNEQKQI.FVFEQHKHYI.DEITFQISFESKRVTI.ADANI.DKVI.SAYNKHR DKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYE TRIDLSQLGGD	dSpCas9 (aa)
181	MTQFEGFTNLVQVSKTLRFELIPQGKTLKHIQEQQGFI EEDKARN DHYKELKPI IDRIYK TYADQCLQLVQLD WENLSAAIDSYRKEKTEETRNL I EEQATYRNAIHDYF IGRDNL I DAINKRHA EIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK NVFSAEDI STAIPHRI VQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTS IEEVFSFPFYNQLLTQTQIDLYNQLLGGI SREAGTEKIKGLNEVLNLAIQKNDETAHI I ASLPHRFIPLFKQILSDRNTLSF ILEEFKSDEEVIQSFCKYKTL LRNENVLETAEALFN ELNSIDLTHIF I SHKKLET I SSALCDHWDTLRNALYERRI SELTGK I TKSACEKVQRSL KHEDINLQEI I SAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLD SLGLYHLLDWF AVDESNEVDPEFSARLTGIKLEMESLSFYNKARNYATKKPYSVEKFK KLNFMPTLASGWDVNKEKNNGA I L FVKNGLYLLGIMPQKQKRYKALSFEPTKTEKSEGF DKMYDYFPDAAKMIPKCS TQLKAVTAHFQHTHTP ILLSNNFIEPLEITKEIYDLNNPE KEPKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPS SQYKDLG EYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTG LFSPENLAKTSIKLNGQAEFYRPKSRMMAHRLGEKMLNKKLKDQKTP I PDTLYQEL YDYVNHRLSHDLSDEARALLPNVITKEVSHEI I KDRRFTSDKFFFHVPITLNYQAANSF SKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSNTIQQFDYQKKL DNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLANLNF GFKSKR TGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQGF FYVPAPYTSKIDPLTGFVDPFVWKT IKNHESRKHFLGDFDLHYDVKTGDFILHFKMNR NLSFQRLPGFMPAWDIVFEKNETQFDAQGTPFIAGKRIVPVIENHRFTGRYRDLYPAN ELIALLEEKIVFRDGSNILPKLLENDSDHAIDTMVALIRSVLQMRNSNAATGEDYINS PVRDLNGVCFDSRFQNP EWPMADANGAYH IALKGQLLLNHLKESKDLKLQNGI SNQDW LAIQELRN	dAsCas12a (aa)
182	TQFEGFTNLVQVSKTLRFELIPQGKTLKHIQEQQGFI EEDKARN DHYKELKPI IDRIYK YADQCLQLVQLD WENLSAAIDSYRKEKTEETRNL I EEQATYRNAIHDYF IGRDNL I AINKRHA EIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRK VFSAE DI STAIPHRI VQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTS I EEVFSFPFYNQLLTQTQIDLYNQLLGGI SREAGTEKIKGLNEVLNLAIQKNDETAHI I A SLPHRFIPLFKQILSDRNTLSF ILEEFKSDEEVIQSFCKYKTL LRNENVLETAEALFNE LNSIDLTHIF I SHKKLET I SSALCDHWDTLRNALYERRI SELTGK I TKSACEKVQRSLK HEDINLQEI I SAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLD LLGLYHLLDWF AVDESNEVDPEFSARLTGIKLEMESLSFYNKARNYATKKPYSVEKFK	dAsCas12a (aa)

	<p>LNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGI MPKQKGRYKALSFEPTTEKTSEGF                  KMYDYDFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEK                  EPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGE                  YYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL                  FSPENLAKTISKLNQAEALFYRPKSRMKRMAHRLGKMLNKKLKDQKTPIPDTLYQELY                  DYVNHRLSHDLSDEARALLPNVITKEVSHEI IKDRRFTSDKFFFHVPITLNYQAANS                  SFNFQVRVAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKL                  NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLNLFNGFKSKRT                  GIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQSGFL                  YVPAPYTSKIDPLTGFVDPFVWTKIKNHESRKHFLGEGDFLHYDVKTGDFILHFKMNRN                  LSFQRGLPGFMPAWDIVFEKNETQFQDAKGTPIIAGKRIVPVIENHRFTGRYRDLYPANE                  LIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSP                  VRDLNGVCFDSRFQNPPEWPMADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL                  AYIQELRN</p>	
<p>183</p>	<p>MTQFEGFTNLYQVSKTLRFELIPOGKTLKHIQEQGFIEEDKARNHDHYKELKPIIDRIYK                  TYADQCLQLVQLDWNENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYIGRTDNL                  TDAINKRHAIEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENR                  NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTS                  IEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHI                  IASLPHRFIPLFKQILSDRNTLSFIEEFKSDEEVIQSFCKYKTLRNENVLETAELFN                  ELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSL                  KHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQD                  SLLGLYHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKF                  KLNFMPTLASGWDVNKEKNNGAILFVKNGLYYLGI MPKQKGRYKALSFEPTTEKTSEGF                  DKMYDYDFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPE                  KEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLG                  EYAFI.NPI.I.YHSFQRIAEKEIMDAVETGKI.YL.FQIYNKDFAKGHHGKPNLHTLYWTG                  LFSPENLAKTISKLNQAEALFYRPKSRMKRMAHRLGKMLNKKLKDQKTPIPDTLYQEL                  YDYVNHRLSHDLSDEARALLPNVITKEVSHEI IKDRRFTSDKFFFHVPITLNYQAANS                  SFNFQVRVAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKL                  DNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLNLFNGFKSKR                  TGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQSGFL                  FYVPAPYTSKIDPLTGFVDPFVWTKIKNHESRKHFLGEGDFLHYDVKTGDFILHFKMNR                  NLSFQRGLPGFMPAWDIVFEKNETQFQDAKGTPIIAGKRIVPVIENHRFTGRYRDLYPAN                  ELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSP                  PVRDLNGVCFDSRFQNPPEWPMADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDW                  LAYIQELRN</p>	<p>WT AsCas12a (AsCpf1) (aa)</p>
<p>184</p>	<p>TQFEGFTNLYQVSKTLRFELIPOGKTLKHIQEQGFIEEDKARNHDHYKELKPIIDRIYK                  YADQCLQLVQLDWNENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYIGRTDNL                  TDAINKRHAIEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENR                  NVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTS                  IEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHI                  IASLPHRFIPLFKQILSDRNTLSFIEEFKSDEEVIQSFCKYKTLRNENVLETAELFNE                  LNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLK                  HEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQD                  SLLGLYHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKF                  LNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGI MPKQKGRYKALSFEPTTEKTSEGF                  KMYDYDFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEK                  EPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGE                  YYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL                  FSPENLAKTISKLNQAEALFYRPKSRMKRMAHRLGKMLNKKLKDQKTPIPDTLYQELY                  DYVNHRLSHDLSDEARALLPNVITKEVSHEI IKDRRFTSDKFFFHVPITLNYQAANS                  SFNFQVRVAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKL                  NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLNLFNGFKSKRT                  GIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQSGFL                  YVPAPYTSKIDPLTGFVDPFVWTKIKNHESRKHFLGEGDFLHYDVKTGDFILHFKMNRN                  LSFQRGLPGFMPAWDIVFEKNETQFQDAKGTPIIAGKRIVPVIENHRFTGRYRDLYPANE                  LIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSP                  VRDLNGVCFDSRFQNPPEWPMADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL                  AYIQELRN</p>	<p>WT AsCas12a (AsCpf1) (aa)</p>
<p>185</p>	<p>MPKPAVESEFSKVLKHFPGERFRSYMKRGGKILAAQGEAAVAYLQGGKSEEEPPNFQ</p>	<p>dCasPhi-2 (aa)</p>

	PPAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGV SNHGYSHVQGLNLI FDHTLGRYDGVLLKKVQLRNEKARARLESINASRADEGLPEIKAE EEVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAP I PLGV VRNRCDIQKGCPCGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ DALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF TYTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIALGQTNPISAGISRVTQENG ALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEALPEAQQAEVRALDGVSK TARTQLCADFGLDPKRLPWDKMSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPV EVMRKDRTWARAYKPRLSVEAQKLNKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIE KTRRRTOCQIVIPVIEDLNVRFFHSGKRLPGWDNFF TAKKENRWF IQGLHKAFSDLRT HRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVAL TGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS GSGP KKKRKY EDPKKKRKY	
186	PKPAVESEFSKVLKHHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQ PAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVS NHGYSHVQGLNLI FDHTLGRYDGVLLKKVQLRNEKARARLESINASRADEGLPEIKAE EVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAP I PLGV VRNRCDIQKGCPCGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ ALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF TYTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIALGQTNPISAGISRVTQENGA LQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEALPEAQQAEVRALDGVSKET ARTQLCADFGLDPKRLPWDKMSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVE VMRKDRTWARAYKPRLSVEAQKLNKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIEK TRRRTOCQIVIPVIEDLNVRFFHSGKRLPGWDNFF TAKKENRWF IQGLHKAFSDLRTH RSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVALT GKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS GSGP KKKRKY DPKKKRKY	dCasPhi-2 (aa)
187	MPKPAVESEFSKVLKHHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQ PPAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGV SNHGYSHVQGLNLI FDHTLGRYDGVLLKKVQLRNEKARARLESINASRADEGLPEIKAE EEVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAP I PLGV VRNRCDIQKGCPCGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ DALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF TYTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIDLQTNPISAGISRVTQENG ALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEALPEAQQAEVRALDGVSK TARTQLCADFGLDPKRLPWDKMSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPV EVMRKDRTWARAYKPRLSVEAQKLNKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIE KTRRRTOCQIVIPVIEDLNVRFFHSGKRLPGWDNFF TAKKENRWF IQGLHKAFSDLRT HRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVAL TGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS	WT CasPhi-2 (aa)
188	PKPAVESEFSKVLKHHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQ PAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVS NHGYSHVQGLNLI FDHTLGRYDGVLLKKVQLRNEKARARLESINASRADEGLPEIKAE EVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAP I PLGV VRNRCDIQKGCPCGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ ALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF TYTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIDLQTNPISAGISRVTQENGA LQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEALPEAQQAEVRALDGVSKET ARTQLCADFGLDPKRLPWDKMSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVE VMRKDRTWARAYKPRLSVEAQKLNKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIEK TRRRTOCQIVIPVIEDLNVRFFHSGKRLPGWDNFF TAKKENRWF IQGLHKAFSDLRTH RSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVALT GKTMPEKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS	WT CasPhi-2 (aa)
189	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAY CTTQVERNACLFCKARKLDDKIFYQKLRGQFPDAVFWEISEIFRQLQKQAAE IYNQSLI ELYEYIFIKGKGIANASSVEHYLSDVCYTRAELEFKNAAIASGLRSKIKSNFRLKELKN MKSGLPTTKSDNFP I PLVKQKGGQYTGFEISNHNSDFI IKIPFGRWQVKKEIDKYRWE KDFEQVQKSPKPI SLLSTQRRKRNGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSK IGEKSAWMLNLSIDVPKIDKGVDPISIGGIDVGVKSPVCAINNAFSRYSISDNDLFHF NKKMFARRRILLKKNRHKRAGHGAKNKLKPIITILTEKSERFRKKLIERWACEIADFFIK NKVGTVQMENLES MKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGI EIRKVAPNNTS	WT Un1Cas12f1 (aa)

	KTCSKCGHLNNYFNFEYRKKNKFPHFKEKCNFKENADYNAALNINSNPKLKSTKEEP	
190	AKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYC TTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEI SE IFRQLQKQAAE IYNQSLIE LYYE I F I K G K G I A N A S S V E H Y L S D V C Y T R A A E L F K N A A I A S G L R S K I K S N F R L K E L K N M K S G L P T T K S D N F P I P L V K Q K G G Q Y T G F E I S N H N S D F I I K I P F G R W Q V K K E I D K Y R P W E K F D F E Q V Q K S P K P I S L L L S T Q R R K R N K G W S K D E G T E A E I K K V M N G D Y Q T S Y I E V K R G S K I G E K S A W M L N L S I D V P K I D K G V D P S I I G G I D V G V K S P L V C A I N N A F S R Y S I S D N D L F H F N K K M F A R R R I L L K K N R H K R A G H G A K N K L K P I T I L T E K S E R F R K K L I E R W A C E I A D F F I K N K V G T V Q M E N L E S M K R K E D S Y F N I R L R G F W P Y A E M Q N K I E F K L K Q Y G I E I R K V A P N N T S K T C S K C G H L N N Y F N F E Y R K K N K F P H F K E K C N F K E N A D Y N A A L N I S N P K L K S T K E E P	WT Un1Cas12f1 (aa)
191	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAY CTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEI SE IFRQLQKQAAE IYNQSLI ELYE I F I K G K G I A N A S S V E H Y L S D V C Y T R A A E L F K N A A I A S G L R S K I K S N F R L K E L K N M K S G L P T T K S D N F P I P L V K Q K G G Q Y T G F E I S N H N S D F I I K I P F G R W Q V K K E I D K Y R P W E K F D F E Q V Q K S P K P I S L L L S T Q R R K R N K G W S K D E G T E A E I K K V M N G D Y Q T S Y I E V K R G S K I G E K S A W M L N L S I D V P K I D K G V D P S I I G G I A V G V K S P L V C A I N N A F S R Y S I S D N D L F H F N K K M F A R R R I L L K K N R H K R A G H G A K N K L K P I T I L T E K S E R F R K K L I E R W A C E I A D F F I K N K V G T V Q M E N L E S M K R K E D S Y F N I R L R G F W P Y A E M Q N K I E F K L K Q Y G I E I R K V A P N N T S K T C S K C G H L N N Y F N F E Y R K K N K F P H F K E K C N F K E N A A Y N A A L N I S N P K L K S T K E E P	dUn1Cas12f1 (aa)
192	AKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYC TTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEI SE IFRQLQKQAAE IYNQSLIE LYYE I F I K G K G I A N A S S V E H Y L S D V C Y T R A A E L F K N A A I A S G L R S K I K S N F R L K E L K N M K S G L P T T K S D N F P I P L V K Q K G G Q Y T G F E I S N H N S D F I I K I P F G R W Q V K K E I D K Y R P W E K F D F E Q V Q K S P K P I S L L L S T Q R R K R N K G W S K D E G T E A E I K K V M N G D Y Q T S Y I E V K R G S K I G E K S A W M L N L S I D V P K I D K G V D P S I I G G I A V G V K S P L V C A I N N A F S R Y S I S D N D L F H F N K K M F A R R R I L L K K N R H K R A G H G A K N K L K P I T I L T E K S E R F R K K L I E R W A C E I A D F F I K N K V G T V Q M E N L E S M K R K E D S Y F N I R L R G F W P Y A E M Q N K I E F K L K Q Y G I E I R K V A P N N T S K T C S K C G H L N N Y F N F E Y R K K N K F P H F K E K C N F K E N A A Y N A A L N I S N P K L K S T K E E P	dUn1Cas12f1 (aa)
193	MARILAF A I G I S S I G W A F S E N D E L K D C G V R I F T K V E N P K T G E S L A L P R R L A R S A R K R L A R R K A R L N H L K H L I A N E F K L N Y E D Y Q S F D E S L A K A Y K G S L I S P Y E L R F R A L N E L L S K Q D F A R V I L H I A K R R G Y D D I K N S D D K E K G A I L K A I K Q N E E K L A N Y Q S V G E Y L Y K E Y F Q K F K E N S K E F T N V R N K K E S Y E R C I A Q S F L K D E L K L I F K K Q R E F G F S F S K K F E E E V L S V A F Y K R A L K D F S H L V G N C S F F T D E K R A P K N S P L A F M F V A L T R I I N L L N N L K N T E G I L Y T K D D L N A L L N E V L K N G T L T Y K Q T K K L L G L S D D Y E F K G E K G T Y F I E F K K Y K E F I K A L G E H N L S Q D D L N E I A K D I T L I K D E I K L K K A L A K Y D L N Q N Q I D S L S K L E F K D H L N I S F K A L K L V T P L M L E G K K Y D E A C N E L N L K V A I N E D K K D F L P A F N E T Y Y K D E V T N P V V L R A I K E Y R K V L N A L L K K Y G K V H K I N I E L A R E V G K N H S Q R A K I E K E Q N E N Y K A K K D A E L E C E K L G L K I N S K N I L K L R L F K E Q K E F C A Y S G E K I K I S D L Q D E K M L E I D A I Y P Y S R S F D D S Y M N K V L V F T K Q N Q E K L N Q T P E A F G N D S A K W Q K I E V L A K N L P T K K Q K R I L D K N Y K D K E Q K N F K D R N L N D T R Y I A R L V L N Y T K D Y L D F L P L S D D E N T K L N D T Q K G S K V H V E A K S G M L T S A L R H T W G F S A K D R N N H L H H A I D A V I I A Y A N N S I V K A F S D F K K E Q E S N S A E L Y A K K I S E L D Y K N K R K F F E P F S G F R Q K V L D K I D E I F V S K P E R K K P S G A L H E E T F R K E E E F Y Q S Y G G K E G V L K A L E L G K I R K V N G K I V K N G D M F R V D I F K H K K T N K F Y A V P I Y T M D F A L K V L P N K A V A R S K K G E I K D W I L M D E N Y E F C F S L Y K D S L I L I Q T K D M Q E P E F V Y Y N A F T S S T V S L I V S K H D N K F E T L S K N Q K I L F K N A N E K E V I A K S I G I Q N L K V F E K Y I V S A L G E V T K A E F R Q R E D F K K	dCjCas9 (aa)
194	ARILAF A I G I S S I G W A F S E N D E L K D C G V R I F T K V E N P K T G E S L A L P R R L A R S A R K R L A R R K A R L N H L K H L I A N E F K L N Y E D Y Q S F D E S L A K A Y K G S L I S P Y E L R F R A L N E L L S K Q D F A R V I L H I A K R R G Y D D I K N S D D K E K G A I L K A I K Q N E E K L A N Y Q S V G E Y L Y K E Y F Q K F K E N S K E F T N V R N K K E S Y E R C I A Q S F L K D E L K L I F K K Q R E F G F S F S K K F E E E V L S V A F Y K R A L K D F S H L V G N C S F F T D E K R A P K N S P L A F M F V A L T R I I N L L N N L K N T E G I L Y T K D D L N A L L N E V L K N G T L T Y K Q T K K L L G L S D D Y E F K G E K G T Y F I E F K K Y K E F I K A L G E H N L S Q D D L N E I A K D I T L I K D E I K L K K A L A K Y D L N Q N Q I D S L S K L E F K D H L N I S F K A L K L V T P L M L E G K K Y D E A C N E L N L K V A I N E D K K D F L P A F N E T Y Y K D E V T N P V V L R A I K E Y R K V L N A L L K K Y G K V H K I N I E L A R E V G K N H S Q R A K I E K E Q N E N Y K A K K D A E L E C E K L G L K I N S K N I L K L R L F K E Q K E F C A Y S G E K I K I S D L Q D E K M L E I D A I Y P Y S R S F D D S Y M N K V L V F T K Q N Q E K L N Q T P F E A F G N D S A K W Q K I E V L A K N L P T K K Q K R I L D K N Y K D K E Q K N F K D R N L N D T R Y I A R L V L N Y T K D Y L D F L P L S D D E N T K L N D T Q K G S K V H V E A K S G M L T S A L R H T W G F S A K D R N N H L H H A I D A V I I A Y A N N S I V K A F S D F K K E Q E S N S A E L Y A K K I S E L D Y K N K R K F F E P F S G F R Q K V L D K I D E I F V S K P E R K K P S G A L H E E T F R K E E E F Y Q S Y G G K E G V L K A L E L G K I R K V N G K I V K N G D M F R V D I F K H K K T N K F Y A V P I Y T M D F A L K V L P N K A V A R S K K G E I K D W I L M D E N Y E F C F S L Y K D S L I L I Q T K D M Q E P E F V Y Y N A F T S S T V S L I V S K H D N K F E T L S K N Q K I L F K N A N E K	dCjCas9 (aa)

	EVIAKSIGIQNLKVFKEYIVSALGEVTKAEFRQREDFKK	
195	MSSGSGHMARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLAR SARKRLARRKARLNHLKHLIANEFKLNIEDYQSFDES LAKAYKGS LI SPYELRFRALNE LLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKONEEKLANYQSVGEYLYKEY FQKFKENSKEFTNVRNKESYERCIAQSFLKDELKLIFFKKQREFGFSF SKKFEEVLSV AFYKRALKDFS HLVGNC SFFTDEKRAPKNSPLAFMFVALTRI INLLNNLKNTEGILYTK DDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYF IEFKKYKEFIKALGEHNL SQDDLNEIAKDI TLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTP LMLEGKKYDEACNELNLKVAINEDKKDFLP AFNETYYKDEV TNPVVLRAIKEYRKVLNA LLKKYGVHKNINIELGGGSGGYIARLVLN YTKDYLD FLP LPSDDENTKLN D TQKGSKVHV EAKSGLM TSALRHTWGFSAKDRNNHLHHAIDAVI IAYANNS IVKAFSDFKKEQESNSAE LYAKKISELDYKNKRKFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEE FYQSYGGKEGV LKALELGKIRKVN GKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFAL KVLPNKAVARSKKGEIKDWILMDENYEFCSLYKDSLILIQTKDMQEPFVYNAFTSS TVSLIVSKHDNK FETLSKNQKILFKNANEKEVIAKSIGIQNLKVFKEYIVSALGEVTKA EFRQREDFKK	WT CjCas9 (aa)
196	SSGSGHMARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARS ARKRLARRKARLNHLKHLIANEFKLNIEDYQSFDES LAKAYKGS LI SPYELRFRALNEL LSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKONEEKLANYQSVGEYLYKEYF QKFKENSKEFTNVRNKESYERCIAQSFLKDELKLIFFKKQREFGFSF SKKFEEVLSVA FYKRALKDFS HLVGNC SFFTDEKRAPKNSPLAFMFVALTRI INLLNNLKNTEGILYTKD DLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYF IEFKKYKEFIKALGEHNL SQDDLNEIAKDI TLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTP MLEGKKYDEACNELNLKVAINEDKKDFLP AFNETYYKDEV TNPVVLRAIKEYRKVLNAL LKKYGVHKNINIELGGGSGGYIARLVLN YTKDYLD FLP LPSDDENTKLN D TQKGSKVHVE AKSGLM TSALRHTWGFSAKDRNNHLHHAIDAVI IAYANNS IVKAFSDFKKEQESNSAE YAKKISELDYKNKRKFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEF YQSYGGKEGV LKALELGKIRKVN GKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALK VLPNKAVARSKKGEIKDWILMDENYEFCSLYKDSLILIQTKDMQEPFVYNAFTSST VSLIVSKHDNK FETLSKNQKILFKNANEKEVIAKSIGIQNLKVFKEYIVSALGEVTKAE FRQREDFKK	WT CjCas9 (aa)
197	MSSTSGKNPTNP IVRTI SCNLSAKEDVLRKVWEEMS QKNTPLIVQLLKS VSEQPEFEAN KENGITITKKEITELRRDITEDSDLKKQSGRLRSADSLVTEVYSSWLRLYQVRKNKKEG KEYFLNNILKSDVELVEQSNCDLQTI RCKAKEILSQVEEFIEQVNNPKINKTTSAKKK INKSNKNNKAI E EILNRFFIGNIDKTLTNTLYE IHRKSPDILTQCAVAYLIKNDNKVSE AEENLTQLNKRSSAKEIEIKRLETQIQNTRLPNGRDI TGEKYSQAFEKLVNQVPQDNEE FAEWIAILLKKVSSLPYPILYSSGDL SWYKDEKGNIFVYFNGWAEYHFQICCDKRQLRF FERFLKYYKALKASEKGEKLSGSLVTLRS AHLLWRQGGKGEKGPWKVNKLALHCTYDAR LWTAEGTEEV RQEKTDKAQAEVNQAESNENIDSKQQKLTKNKSSLSRLKNSFARP SKP LYRGQSNII VGISFHPVELATLVVVDINTKEILICKTVKELLGDAFP LLSRRRRQVHF RKEREKAQKSDPCDLGESKLG EYVDRLLAKRIVEVAKEYQASCIVLPGLKGIREIRTS VIQAKAETKFPGDINAQELYVKEYNRQIHNWSYSRLQESIKSRAAELKISIKFGKQPSH STLQEQAINLALSA	WT Cas12k (aa)
198	SSTSGKNPTNP IVRTI SCNLSAKEDVLRKVWEEMS QKNTPLIVQLLKS VSEQPEFEANK ENGTITKKEITELRRDITEDSDLKKQSGRLRSSADSLVTEVYSSWLRLYQVRKNKKEGK EYFLNNILKSDVELVEQSNCDLQTI RCKAKEILSQVEEFIEQVNNPKINKTTSAKKKI NKSNNKAI E EILNRFFIGNIDKTLTNTLYE IHRKSPDILTQCAVAYLIKNDNKVSEA EENLTQLNKRSSAKEIEIKRLETQIQNTRLPNGRDI TGEKYSQAFEKLVNQVPQDNEEF AEWIAILLKKVSSLPYPILYSSGDL SWYKDEKGNIFVYFNGWAEYHFQICCDKRQLRFF ERFLKYYKALKASEKGEKLSGSLVTLRS AHLLWRQGGKGEKGPWKVNKLALHCTYDARL WTAEGTEEV RQEKTDKAQAEVNQAESNENIDSKQQKLTKNKSSLSRLKNSFARP SKPL YRGQSNII VGISFHPVELATLVVVDINTKEILICKTVKELLGDAFP LLSRRRRQVHF KEREKAQKSDPCDLGESKLG EYVDRLLAKRIVEVAKEYQASCIVLPGLKGIREIRTSV IQAKAETKFPGDINAQELYVKEYNRQIHNWSYSRLQESIKSRAAELKISIKFGKQPSH TLQEQAINLALSA	WT Cas12k (aa)
199	MGPKKRKRKVGSGSAKNTITKTLKLRIVRPYN SAEVEKIVADEKNNREKIALEKNKDKVK EACSKHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEI SEIFRQLQ KQAAEIYNQSLIELYYEIFIKGKGIANASVEHYLSRVCYRRAAELFKNAAIASGLRSK IKSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFI IKIPFGRWQ VKKEIDKYRPWEKFD FEQVQKSPKISL LLS TQRKRKNKGWSKDEGTEAEIKKVMNGDY QTSYIEV KRGSKICEKSAWMLNLSIDV PKIDKGVDP SII GGIAGVRSPLVCAINNAFS	CasMini

	RYSISDNDLDFHFNKMMFARRRILLKKNRHKRAGHGAKNKLKPIITILTEKSERFRKKLIE RWACEIADFFIKNKVGTVMENLES MKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYG IEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAAYNAALNIN PKLKSTKERPAKRPAATKKAGQAKKKK	
200	GPKKRKRKVGSGSAKNTITKTLKLRIVRPNYNSAEVEKIVADEKNNREKIALEKNKDKVKE ACSKHLKVAAYCTTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEI SEIFRQLQK QAAEIYNQSLIELYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKNAAIASGLRSKI KSNFRLKELKNMKSGLPTTKSDNFP IPLVKQKGGQYTGFEISNHNSDFI IKIPFGRWQV KKEIDKYRPWEKFDFEQVQKSPKPI SLLLSSTQRRKRNGWSKDEGTEAEIKKVMNGDQ TSYIEVGRGSKICEKSAWMLNLSIDVPKIDKGVDP S IGGIAVGVRSPLVCAINNAF SR YSISDNDLDFHFNKMMFARRRILLKKNRHKRAGHGAKNKLKPIITILTEKSERFRKKLIER WACEIADFFIKNKVGTVMENLES MKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGI EIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENAAAYNAALNIN PKLKSTKERPAKRPAATKKAGQAKKKK	CasMini
201	gaaacaccguaauuucuaucucuuguagau	dAsCas12a gRNA scaffold
202	CAACGAUUGCCCCUCACGAGGGGAC	dCasPhi-2 gRNA scaffold
203	ACCGCUUCACUUAGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUU UCUUCGGAAAGUAACCCUCGAAACAAAGAAA	Un1Cas12f1 gRNA scaffold
204	GGAAUGAAC	Un1Cas12f1 gRNA scaffold
205	UUUUUUUU	Un1Cas12f1 gRNA scaffold
206	GCGUUUUAGGGGAUUGUAACCCCGCAGAGUCCCGCAAACUCUUUAUUUUAGUCCUUU UCAGGGACUAAAAC	CjCas9 gRNA scaffold
207	AUAUUAAUAGCGCCGCAAUUAUGCUGCUUGCAGCCUCUGAAUUUUUGUAAAUGAGGGU UAGUUUGACUGUAUAAAUCAGUCUUGCUUUCUGACCCUGGUAGCUGCUCACCCUGAUG CUGCUGUCAAUAGACAGGAUAGGUGCGCUCCAGCAAUAAGGGCGCGGAUGUACUGCUG UAGUGGCUACUGAAUCACCCCGAUC AAGGGGGAACCCUCCAAAAGGUGGGUUGAAAG	Cas12k gRNA scaffold
208	GGGCUUCACUGAUAAGUGGAGAACCGCUUCACCAAAGCUGUCCCUUAGGGGAUUAGA ACUUGAGUGAAGGUGGGCUGCUUGCAUCAGCCUAAUGUCGAGAAGUGCUUUUCUUCGAA AGUAACCCUCGAAACAAAUUCAUUUGAAUGAAGGAAUGCAAC	CasMini gRNA scaffold
209	GSGSG	linker (aa)
210	MPKPAVESEFSKVLKHHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQ PPAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGV SNHGYSHVQGLNLI FDHTLGRYDGVLLKVKQLRNEKARARLESINASRADEGLPEIKAE EEVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGV VRNRCDIQKGC PGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ DALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF YTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIALGQTNPISAGISRVTQENG ALQCEPLDRFTLPDDLKDISAYRIAWDRNEEELRARSVEALPEAQQA EVRALDGVSK TARTQLCADFGLDPKRLPDKMS SNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPV EVMRKDRTWARAYKPRLSVEAQKLKNEALWALKRTSPEYLLKLSRRKEELCRRSINYVIE KTRRRITQCQIVIPVIEDLNVRFFHSGKRLPGWDNFFTAKKENRWFIOGLHKAFSDLRT HRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVAL TGKTMKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS	dCasPhi-2
211	PKPAVESEFSKVLKHHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQ PAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVS NHGYSHVQGLNLI FDHTLGRYDGVLLKVKQLRNEKARARLESINASRADEGLPEIKAE EVATNETGHLLQPPGINPSFYVYQTI SPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGV VRNRCDIQKGC PGYIPEWQREAGTAI SPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQ ALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTF YTLDACGTYARKWTLKKGQTKATLDKLTATQTVALVAIALGQTNPISAGISRVTQENGA LQCEPLDRFTLPDDLKDISAYRIAWDRNEEELRARSVEALPEAQQA EVRALDGVSK ARTQLCADFGLDPKRLPDKMS SNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPV VMRKDRTWARAYKPRLSVEAQKLKNEALWALKRTSPEYLLKLSRRKEELCRRSINYVIE KTRRRITQCQIVIPVIEDLNVRFFHSGKRLPGWDNFFTAKKENRWFIOGLHKAFSDLRTH RSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVALT GKTMKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS	dCasPhi-2

212	NNNNGNNN, where N is any nucleotide	N. meningitidis Cas9 PAM
213	AKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYC TTQVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAE IYNQSLIE LYYEIFIKGKGIANASSVEHYLSRVCYRRAAELFKNAATIASGLRSKIKSNFRLKELKNM KSGLPPTTKSDNFPIPLVKQKGGQYTGFEISNHNSDFI IKIPFGRWQVKKEIDKYRPWEK FDFEQVQKSPKPI SLLLSTQRRKRKNGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKI CEKSAWMLNLSIDVPKIDKGVDPSTIGGIAVGVRSPLVCAINNAF SRY S I S D N D L F H F N KKMFARRRILLKKNRHKRAGHGAKNKLPITILTEKSERFRKKLIERWACEIADFFIKN KVGTVQMENLESMKRKEDSYFNIRLRGFWPYAEMQNKIEFKLKQYGI EIRKVAPNNTSK TCSKCGHLNNYFNFEYRKKNKFPHPFKCEKCNFKENAAAYNAALNINPPLKLSKERP	CasMini Cas12f/D326A/ D510A/D143R /T147R/K330R /E528R (aa)
214	TBN, where N is any nucleotide, and B is G, T, or C	PAM - CasPhi
215	TTN, where N is any nucleotide	PAM - Un1Cas12f1
216	NNNNACAC, where N is any nucleotide	PAM - CjCas9
217	GGTT	PAM - Cas12k
218	GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCG TCAACTTGGTTGGCGAGA	SaCas9 gRNA scaffold DNA
219	GUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCG UCAACUUGUUGCGAGA	SaCas9 gRNA scaffold sequence
220	GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAA CTTGAAAAAGTGGCACCGAGTCGGTGC	SpCas9 gRNA scaffold DNA
221	MAAAAAAAPS GGGGGGEEERLEEKSEDDQLQGLKDKPLKFKKVKKDKKKEEKEGKHEPVQ PSAHHSAEPAEAGKAETSESGSAPAVPEASASPKQRRS I IRDRGPMYDDP TLPEGWTR KLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDF TVTGRGSP SRREQPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKM PFQTSPPGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVAAAAAAEA KKKAVKESSIRSVQETVLP IKKRKTRET V S IEVKEVVKPLL V S T L G E K S G K L K T C K S P GRKSKESSPKGRSSSASSPPKKEHHHHHHHSESPKAPVPLLPPLPPPPPEPESSEDPTS PPEPQDLSSSVCKEEKMPRGGSLSDGCPKEPAKTQPAVATAATAAEKYKHRGEGGERKD IVSSSMRPNREEPVDSRTPVTERVS	Human MeCP2 isoform B
222	TTTR, where R is G or A	PAM - Cas12f
223	GCCGCGCCGAGCGGAGGAGG	gRNA m1 spacer
224	CTGGCGTTGTTCCAAGCCAA	gRNA m2 spacer
225	AGTGGGACCGCCAAGGCCGC	gRNA m3 spacer
226	TGCTGACTGGTATCAGGGTA	gRNA m4 spacer
227	GACCGCCAAGGCCGCGGGCG	gRNA m5 spacer
228	TTGAAAAAAGAGGCGGCTA	gRNA m6 spacer
229	GGGGCAAAAAGTCACGGAAT	gRNA m7 spacer
230	GATCGGTTATGTTTAGGGTT	Non-targeting gRNA spacer
231	CGGGGTCCGACGACACGGCTG	SaCas9 gRNA A spacer
232	TTCACTTGCCCCAGCATCCGC	SaCas9 gRNA B spacer
233	GTTAAGGATTAATGGACCCTT	SaCas9 gRNA C spacer
234	AGTGGGCGGAATTTGAATGTT	SaCas9 gRNA D spacer
235	GACTGGTTTAGTGGGCGGAAT	SaCas9 gRNA E spacer

236	CGGGGACCCTTGCCGGGGGGC	SaCas9 gRNA F spacer
237	CGAGGACGGTCACCCGCGAGC	SaCas9 gRNA G spacer
238	CGGAGGGACTGGTTTAGTGGG	SaCas9 gRNA H spacer
239	AGGAGGCGAGGAGGAGAGACT	SaCas9 gRNA I spacer
240	CTCGGCCCGTCACCCCTGCTC	SaCas9 gRNA J spacer
241	GTGGGCGGAATTTGAATGTTAAG	Cpf1 gRNA A spacer
242	CCTGGCCGAAATGGACAGGAAAT	Cpf1 gRNA B spacer
243	AATGTTAAGGATTAATGGACCCT	Cpf1 gRNA C spacer
244	CCCAAACGACGGCCGAAAGCAGC	Cpf1 gRNA D spacer
245	CCACAGCCCTCTCTCCGAGAGGA	Cpf1 gRNA E spacer
246	GGGAAAAGAGGCGGCTTGGGCGC	Cpf1 gRNA F spacer
247	CTGTCCATTTTCGGCCAGGGAAAA	Cpf1 gRNA G spacer
248	CGCTGCTCTGAGGGGCGATTGAC	Cpf1 gRNA H spacer
249	GGCCGTCGTTTGGGGAAAAGAGG	Cpf1 gRNA I spacer



Claims

1. A DNA-targeting system comprising a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.
2. The DNA-targeting system of claim 1, further comprising at least one effector domain that increases transcription of the MeCP2 locus.
3. A DNA-targeting system comprising:
  - (a) a DNA-targeting domain that binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and
  - (b) at least one effector domain that increases transcription of the MeCP2 locus.
4. The DNA-targeting system of any of claims 1-3, wherein binding of the DNA-targeting domain to the target site does not introduce a genetic disruption or a DNA break at or near the target site.
5. The DNA-targeting system of any of claims 1-4, wherein the DNA-targeting domain comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination comprising (a) a Cas protein or a variant thereof, optionally wherein the Cas protein or a variant thereof is a deactivated Cas (dCas) protein, and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.
6. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:
  - (a) a deactivated Cas (dCas) protein; and
  - (b) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus or is complementary to the target site.
7. The DNA-targeting system of claim 5 or 6, wherein the at least one gRNA is

capable of complexing with the Cas protein or variant thereof or the dCas protein.

8. The DNA-targeting system of any of claims 5-7, wherein the at least one gRNA comprises a gRNA spacer sequence that is capable of hybridizing to the target site or is complementary to the target site.

9. The DNA-targeting system of any of claims 5-8, wherein the Cas protein or variant thereof is a deactivated Cas9 (dCas9) protein, optionally a *Staphylococcus aureus* dCas9 (dSaCas9) protein or a *Streptococcus pyogenes* dCas9 (dSpCas9) protein.

10. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

- (a) a *Streptococcus pyogenes* dCas9 (dSpCas9) protein;
- (b) at least one effector domain that increases transcription of a methyl-CpG-binding protein 2 (MeCP2) locus; and
- (c) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a MeCP2 locus or is complementary to the target site.

11. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

- (a) a *Streptococcus pyogenes* dCas9 (dSpCas9) protein; and
- (b) at least one gRNA comprising a gRNA spacer sequence that is capable of hybridizing to a target site in a regulatory DNA element of a MeCP2 locus or is complementary to the target site.

12. The DNA-targeting system of claim 11, further comprising at least one effector domain that increases transcription of a methyl-CpG-binding protein 2 (MeCP2) locus.

13. The DNA-targeting system of any of claims 5 and 7-12, wherein the Cas protein or a variant thereof is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96, and/or the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%,

96%, 97%, 98%, or 99% sequence identity thereto.

14. The DNA-targeting system of any of claims 5 and 7-9, wherein the Cas protein or a variant thereof is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99, and/or the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

15. The DNA-targeting system of any of claims 5 and 7-14, wherein the Cas protein or variant thereof is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein, wherein when the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein.

16. The DNA-targeting system of claim 15, wherein:

the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and

the N-terminal fragment of the variant Cas protein comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and/or the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and

the C-terminal fragment of the variant Cas protein comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of

the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

17. The DNA-targeting system of any of claims 1-16, wherein the target site comprises the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

18. The DNA-targeting system of any of claims 1-17, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

19. The DNA-targeting system of any of claims 1-18, wherein the target site comprises the sequence set forth in SEQ ID NO: 9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

20. The DNA-targeting system of any of claims 1-19, wherein:  
the target site comprises the sequence set forth in SEQ ID NO: 9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing; and/or  
the at least one gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

21. The DNA-targeting system of claim 20, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30, and/or wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:69, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:69.

22. The DNA-targeting system of any of claims 1-19, wherein:  
the target site comprises the sequence set forth in SEQ ID NO: 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing; and/or  
the at least one gRNA comprises a gRNA that comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least

14 nt.

23. The DNA-targeting system of claim 22, wherein the at least one gRNA further comprises the sequence set forth in SEQ ID NO:30; and/or wherein the at least one gRNA comprises a gRNA that comprises the sequence set forth in SEQ ID NO:87, optionally wherein the at least one gRNA is the gRNA sequence set forth in SEQ ID NO:87.

24. The DNA-targeting system of any of claims 6-23, wherein the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length, optionally wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

25. The DNA-targeting system of any of claims 5-24, wherein the gRNA comprises modified nucleotides for increased stability.

26. The DNA-targeting system of any of claims 1-25, wherein the DNA-targeting system further comprises at least one effector domain, optionally wherein the DNA-targeting domain or a component thereof is fused to the at least one effector domain, optionally wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA, and the component thereof fused to the at least one effector domain is the Cas protein or a variant thereof.

27. The DNA-targeting system of claim 26, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

28. The DNA-targeting system of claims 26 or 27, wherein the effector domain induces transcription de-repression.

29. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

30. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein set forth in SEQ ID NO:95 fused to at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

31. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

32. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a first polypeptide of a split variant Cas9 protein comprising an N-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an N-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

33. The DNA-targeting system of claim 31 or 32, further comprising a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of the dSpCas9 fused to a C-terminal Intein.

34. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment

of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to an C-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:39.

35. A DNA-targeting system comprising a DNA-targeting domain that is a Cas-guide RNA (gRNA) combination comprising:

(a) a second polypeptide of a split variant Cas9 protein comprising a C-terminal fragment of a *Streptococcus pyogenes* deactivated Cas9 protein (dSpCas9) protein fused to a C-terminal intein and at least one effector domain that induces transcription de-repression; and

(b) at least one gRNA that is a gRNA comprising a gRNA spacer sequence set forth in SEQ ID NO:57.

36. The DNA-targeting system of claim 34 or 35, further comprising a first polypeptide of a split variant Cas9 protein an N-terminal fragment of the dSpCas9 fused to an N-terminal Intein.

37. The DNA-targeting system of any of claims 31-36, wherein when the first polypeptide and the second polypeptide of the split variant Cas9 are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas9 to form a full-length variant Cas9 protein.

38. The DNA-targeting system of claim 37, wherein:

the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and

the N-terminal fragment of the variant Cas9 comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and/or the C-terminal Intein comprises a C-terminal Npu Intein,

or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and

the C-terminal fragment of the variant Cas9 comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

39. The DNA-targeting system of any of claims 2-38, wherein the effector domain comprises a catalytic domain of a ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof.

40. The DNA-targeting system of any of claims 2-39, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof, and/or the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

41. The DNA-targeting system of any of claims 2-40, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof.

42. The DNA-targeting system of any of claims 2-41, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

43. The DNA-targeting system of any of claims 1-42, wherein the DNA-targeting system comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

44. A combination, comprising:



a first DNA-targeting domain comprising the DNA targeting domain of any of claims 1-43, and

one or more second DNA-targeting domains, optionally wherein the one or more second DNA-targeting domains comprises the DNA targeting domain of any of claims 1-43.

45. The combination of claim 44, wherein:

the first DNA-targeting domain binds a first target site in the MeCP2 locus; and

the second DNA-targeting domain binds a second target site in the MeCP2 locus.

46. The combination of claim 44 or 45, wherein the first target site and the second target site independently are located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

47. The combination of any of claims 44-46, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is a deactivated Cas9 (dCas9) protein, optionally a *Staphylococcus aureus* dCas9 (dSaCas9) protein or a *Streptococcus pyogenes* dCas9 (dSpCas9) protein.

48. The combination of claim 47, wherein the first variant Cas protein and/or the second variant Cas protein is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96; or comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

49. The combination of claim 47, wherein the first variant Cas protein and/or the second variant Cas protein is a *Staphylococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; or comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

50. The combination of any of claims 44-49, wherein the first variant Cas protein and/or the second variant Cas protein is a split variant Cas9 protein, wherein the split variant

Cas9 protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas9 and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas9 and a C-terminal Intein.

51. The combination of any of claims 44-50, wherein the first Cas protein and the second Cas protein are the same.

52. The combination of any of claims 44-50, wherein the first Cas protein and the second Cas protein are different.

53. The combination of any of claims 44-52, wherein the first Cas protein or a variant thereof and/or the second Cas protein or a variant thereof is fused to at least one effector domain, optionally wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation, optionally wherein the effector domain induces transcription de-repression.

54. The combination of any of claims 44-53, wherein the first Cas protein and the second Cas protein are encoded in a first polynucleotide and/or the first gRNA and the second gRNA are encoded in a first polynucleotide.

55. The combination of any of claims 44-53, wherein the first Cas protein is encoded in a first polynucleotide and the second Cas protein is encoded in a second polynucleotide; and/or wherein the first gRNA is encoded in a first polynucleotide and the second gRNA is encoded in a second polynucleotide, optionally wherein the first Cas protein and the first gRNA are encoded in a first polynucleotide, and the second Cas protein and the second gRNA are encoded in a second polynucleotide.

56. A guide RNA (gRNA) that binds a target site located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

57. A guide RNA (gRNA) that binds a target site comprising the sequence set forth in any one of SEQ ID NOs: 1-29, a contiguous portion thereof of at least 14 nt, or a

complementary sequence of any of the foregoing.

58. The gRNA of claim 56 or 57, wherein the target site comprises the sequence set forth in SEQ ID NO: 9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

59. The gRNA of any of claims 56-58, wherein:  
the target site comprises the sequence set forth in SEQ ID NO: 9, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing; and/or  
the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:39, or a contiguous portion thereof of at least 14 nt.

60. The gRNA of any of claims 56-59, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30, optionally wherein the gRNA comprises the sequence set forth in SEQ ID NO:69, optionally wherein the gRNA sequence is set forth in SEQ ID NO:69.

61. The gRNA of any of claims 56-58, wherein:  
the target site comprises the sequence set forth in SEQ ID NO: 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing; and/or  
the gRNA comprises a gRNA spacer sequence comprising the sequence set forth in SEQ ID NO:57, or a contiguous portion thereof of at least 14 nt.

62. The gRNA of any of claims 56-58, and 61, wherein the gRNA further comprises the sequence set forth in SEQ ID NO:30, optionally wherein the gRNA comprises the sequence set forth in SEQ ID NO:87, optionally wherein the gRNA sequence is set forth in SEQ ID NO:87.

63. The gRNA of any of claims 56-62, wherein the gRNA spacer sequence is between 14 nt and 24 nt, or between 16 nt and 22 nt in length, optionally wherein the gRNA spacer sequence is 18 nt, 19 nt, 20 nt, 21 nt or 22 nt in length.

64. The gRNA of any of claims 56-63, wherein the gRNA comprises modified nucleotides for increased stability.

65. The gRNA of any of claims 56-64, wherein the gRNA is capable of complexing with the Cas protein or variant thereof.

66. A combination, comprising a first gRNA comprising the gRNA of any of claims 56-65, and one or more second gRNAs that binds to a second target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

67. The combination of claim 66, wherein the second gRNA comprises the gRNA of any of claims 56-65.

68. A combination, comprising:  
a first gRNA that binds a first target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus, wherein the first target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158; and  
a second gRNA that binds a second target site in a regulatory DNA element of a MeCP2 locus, wherein the second target site is located within the genomic coordinates hg38 chrX:154,097,151-154,098,158.

69. A fusion protein comprising (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein:

the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and  
the effector domain increases transcription of the MeCP2 locus.

70. A fusion protein comprising (1) a DNA-targeting domain or a component thereof and (2) at least one effector domain, wherein:

the DNA-targeting domain or a component thereof binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus; and  
the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

71. The fusion protein of claim 69 or 70, wherein the DNA-targeting domain

comprises a Clustered Regularly Interspaced Short Palindromic Repeats associated (Cas)-guide RNA (gRNA) combination comprising (a) a Cas protein or a variant thereof and (b) at least one gRNA; a zinc finger protein (ZFP); a transcription activator-like effector (TALE); a meganuclease; a homing endonuclease; or an I-SceI enzyme or a variant thereof, optionally wherein the DNA-targeting domain comprises a catalytically inactive variant of any of the foregoing.

72. The fusion protein of any of claims 69-71, wherein the DNA-targeting domain comprises a Cas-gRNA combination comprising a Cas protein or a variant thereof and at least one gRNA, and the component of the DNA-targeting domain is a Cas protein or a variant thereof.

73. A fusion protein comprising (1) a Cas protein or a variant thereof and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

74. A fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

75. A fusion protein comprising (1) a first polypeptide of a split variant Cas protein comprising an N-terminal fragment of a Cas protein and an N-terminal Intein, and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

76. A fusion protein comprising (1) a second polypeptide of a split variant Cas protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain induces transcription activation, transcription co-activation, transcription elongation, transcription de-repression, histone modification, nucleosome remodeling, chromatin remodeling, reversal of heterochromatin formation, DNA demethylation, or DNA base oxidation.

77. A fusion protein comprising (1) a second polypeptide of a split variant Cas

protein comprising a C-terminal fragment of a Cas protein and a C-terminal Intein and (2) at least one effector domain, wherein the effector domain increases transcription of the MeCP2 locus.

78. The fusion protein of any of claims 71-77, wherein the Cas protein or a variant thereof is capable of complexing with at least one gRNA.

79. The fusion protein of any of claims 71-77, wherein the gRNA binds to a target site in a regulatory DNA element of a methyl-CpG-binding protein 2 (MeCP2) locus.

80. The fusion protein of any of claims 69-79, wherein binding of the DNA-targeting domain or a component thereof targeted to the target site does not introduce a genetic disruption or a DNA break at or near the target site.

81. The fusion protein of any of claims 71-80, wherein the Cas protein or variant thereof is a deactivated Cas (dCas) protein.

82. The fusion protein of any of claims 71-81, wherein the Cas protein or variant thereof is a deactivated Cas9 (dCas9) protein, optionally a *Staphylococcus aureus* dCas9 (dSaCas9) protein or a *Streptococcus pyogenes* dCas9 (dSpCas9) protein.

83. The fusion protein of any of claims 71-82, wherein the Cas protein or variant thereof is a *Streptococcus pyogenes* dCas9 (dSpCas9) protein that comprises at least one amino acid mutation selected from D10A and H840A, with reference to numbering of positions of SEQ ID NO:96; and/or the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

84. The fusion protein of any of claims 71-82, wherein the Cas protein or variant thereof is a *Streptococcus aureus* dCas9 protein (dSaCas9) that comprises at least one amino acid mutation selected from D10A and N580A, with reference to numbering of positions of SEQ ID NO:99; and/or the variant Cas9 protein comprises the sequence set forth in SEQ ID NO:98, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

85. The fusion protein of any of claims 71-84, wherein the Cas protein or variant thereof is a split variant Cas protein, wherein the split variant Cas protein comprises a first polypeptide comprising an N-terminal fragment of the variant Cas protein and an N-terminal Intein, and a second polypeptide comprising a C-terminal fragment of the variant Cas protein and a C-terminal Intein, wherein when the first polypeptide and the second polypeptide of the split variant Cas protein are present in proximity or present in the same cell, the N-terminal Intein and C-terminal Intein self-excite and ligate the N-terminal fragment and the C-terminal fragment of the variant Cas protein to form a full-length variant Cas protein.

86. The fusion protein of claim 85, wherein:

the N-terminal Intein comprises an N-terminal Npu Intein, or the sequence set forth in SEQ ID NO:129, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and

the N-terminal fragment of the variant Cas protein comprises:

the N-terminal fragment of variant SpCas9 from the N-terminal end up to position 573 of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:127, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and/or the C-terminal Intein comprises a C-terminal Npu Intein, or the sequence set forth in SEQ ID NO:133, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing; and the C-terminal fragment of the variant Cas protein comprises:

the C-terminal fragment of variant SpCas9 from position 574 to the C-terminal end of the dSpCas9 sequence set forth in SEQ ID NO:95, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto; or

the sequence set forth in SEQ ID NO:135, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, or a portion of any of the foregoing.

87. The fusion protein of any of claims 69-86, wherein the target site is located within the genomic coordinates human genome assembly GRCh38 (hg38) chrX:154,097,151-154,098,158.

88. The fusion protein of any of claims 69-87, wherein the target site comprises the sequence set forth in any one of SEQ ID NOS:1-29, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

89. The fusion protein of any of claims 69-88, wherein the target site comprises the sequence set forth in SEQ ID NO:9 or 27, a contiguous portion thereof of at least 14 nt, or a complementary sequence of any of the foregoing.

90. The fusion protein of any of claims 69-89, wherein the effector domain induces transcription de-repression, DNA demethylation or DNA base oxidation.

91. The fusion protein of any of claims 69-90, wherein the effector domain comprises a catalytic domain of a Ten-eleven translocation (TET) family methylcytosine dioxygenase or a portion or a variant thereof, or the effector domain comprises a catalytic domain of a Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) or a portion or a variant thereof, optionally wherein the effector domain comprises the sequence set forth in SEQ ID NO:93, or a portion thereof, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the foregoing.

92. The fusion protein of any of claims 69-91, wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus, of the DNA-targeting domain or a component thereof, optionally wherein the at least one effector domain is fused to the N-terminus, the C-terminus, or both the N-terminus and the C-terminus of the Cas protein or a variant thereof.

93. The fusion protein of any of claims 69-92, further comprising one or more linkers connecting the DNA-targeting domain or a component thereof, optionally the Cas protein or variant thereof, to the at least one effector domain, and/or further comprising one or more nuclear localization signals (NLS).

94. The fusion protein of any of claims 69-93, wherein the fusion protein comprises the sequence set forth in SEQ ID NO:91, or an amino acid sequence that has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.



95. A combination comprising the fusion protein of any of claims 69-94, and at least one gRNA, optionally wherein the at least one gRNA is a gRNA of any of claims 56-65.

96. A polynucleotide encoding the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, or the fusion protein of any of claims 69-94, or a portion or a component of any of the foregoing.

97. A polynucleotide encoding a first DNA-targeting system, a first Cas protein and/or a first gRNA of the DNA-targeting system of any of claims 1-43 or the combination of any of claims 44-55, 66-68, and 95.

98. A polynucleotide encoding a second DNA-targeting system, a second Cas protein and/or a second gRNA of the DNA-targeting system of any of claims 1-43 or the combination of any of claims 44-55, 66-68, and 95.

99. A plurality of polynucleotides, comprising the polynucleotide of any of claims 96-98, and one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, or the fusion protein of any of claims 69-94, or a portion or a component of any of the foregoing.

100. A plurality of polynucleotides, comprising:  
a first polynucleotide comprising the polynucleotide of claim 97; and  
a second polynucleotide comprising the polynucleotide of claim 98.

101. A vector comprising the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, or a first polynucleotide or a second polynucleotide of the plurality of polynucleotides of claim 99 or 100, or a portion or a component of any of the foregoing.

102. The vector of claim 101, wherein the vector is a viral vector, optionally wherein the viral vector is an AAV vector.

103. The vector of claim 102, wherein the viral vector, optionally the AAV vector, exhibits tropism for a nervous system cell, optionally a neuron, a heart cell, optionally a cardiomyocyte, a skeletal muscle cell, a fibroblast, an induced pluripotent stem cell, or a cell derived from any of the foregoing, and/or wherein the viral vector is an AAV vector and the AAV vector is selected from among AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or AAV-DJ vector, optionally an AAV9 vector.

104. The vector of claim 101, wherein the vector is a non-viral vector selected from: a lipid nanoparticle, a liposome, an exosome, or a cell penetrating peptide.

105. A plurality of vectors, comprising the vector of any of claims 101-104, and one or more additional vectors comprising one or more additional polynucleotides encoding an additional portion or an additional component of the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, or the fusion protein of any of claims 69-94, or a portion or a component of any of the foregoing.

106. A plurality of vectors, comprising:  
a first vector comprising the polynucleotide of claim 97; and  
a second vector comprising the polynucleotide of claim 98.

107. A cell comprising the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106, or a portion or a component of any of the foregoing.

108. The cell of claim 107, wherein is a nervous system cell, optionally a neuron, a heart cell, optionally a cardiomyocyte, a skeletal muscle cell, a fibroblast, an induced pluripotent stem cell, or a cell derived from any of the foregoing, optionally a nervous system cell, optionally a neuron, or an induced pluripotent stem cell, optionally a nervous system cell, optionally a neuron, or an induced pluripotent stem cell.

109. The cell of claim 107 or 108, wherein the cell is from a subject that has or is suspected of having that has or is suspected of having Rett syndrome, MeCP2-related severe

neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally wherein the cell is from a subject that has or is suspected of having Rett syndrome.

110. A pharmaceutical composition comprising the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106, or a portion or a component of any of the foregoing.

111. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a cell, the method comprising:

introducing the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106 the pharmaceutical composition of claim 110, or a portion or a component of any of the foregoing, into the cell.

112. The method of claim 111, wherein the cell is from a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, optionally wherein the cell is from a subject that has or is suspected of having Rett syndrome.

113. A method for modulating the expression of methyl-CpG-binding protein 2 (MeCP2) in a subject, the method comprising:

administering the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106 the pharmaceutical composition of claim 110, or a portion or a component of any of the foregoing, to the subject.

114. The method of claim 112 or 113, wherein the subject has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or

PPM-X syndrom, optionally wherein the subject has or is suspected of having Rett syndrome.

115. A method of treating Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome, the method comprising:  
administering the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106 the pharmaceutical composition of claim 110, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome, MeCP2-related severe neonatal encephalopathy, Angelman syndrome, or PPM-X syndrome.

116. A method of treating Rett syndrome, the method comprising:  
administering the DNA-targeting system of any of claims 1-43, the gRNA of any of claims 56-65, the combination of any of claims 44-55, 66-68, and 95, the fusion protein of any of claims 69-94, the polynucleotide of any of claims 96-98, the plurality of polynucleotides of claim 99 or 100, the vector of any of claims 101-104, the plurality of vectors of claim 105 or 106 the pharmaceutical composition of claim 110, or a portion or a component of any of the foregoing, to a subject that has or is suspected of having Rett syndrome.

117. The method of any of claims 111-116, wherein:  
a cell in the subject comprises a mutant MeCP2 allele in the active X chromosome, optionally wherein the mutant MeCP2 allele comprises a mutation corresponding to R255X; and/or a cell in the subject comprises a wild-type MeCP2 allele in the inactive X chromosome; and/or  
a cell in the subject exhibits reduced or minimal expression of the wild-type MeCP2 compared to a cell from a normal subject.

118. The method of any of claims 111-116, wherein the cell is a nervous system cell, optionally a neuron, a heart cell, optionally a cardiomyocyte, a skeletal muscle cell, a fibroblast, an induced pluripotent stem cell, or a cell derived from any of the foregoing, optionally a nervous system cell, optionally a neuron, or an induced pluripotent stem cell.

119. The method of any of claims 111-118, wherein the introducing, contacting or

administering is carried out *in vivo* or *ex vivo*.

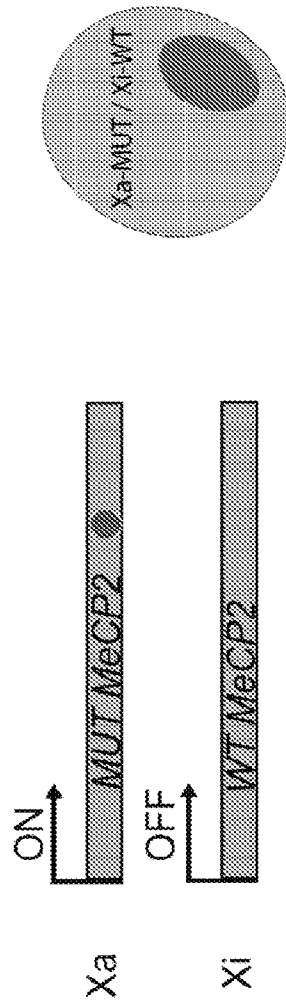
120. The method of any of claims 111-119, wherein following the introducing, contacting or administering, the expression of the wild-type MeCP2 allele from the inactive X chromosome is increased in the cell or the subject, optionally wherein:

the expression is increased at least about 2-fold, 2.5-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 75-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 25-fold, or 30-fold; and/or the expression is increased by less than about 200-fold, 150-fold, or 100-fold; and/or

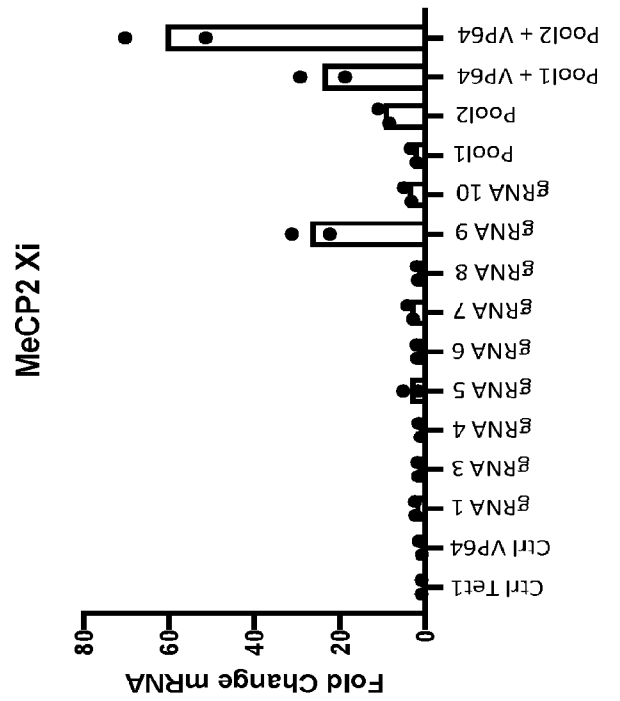
the expression of the wild-type MeCP2 allele is increased to at least 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression of the wild-type MeCP2 of a cell from a normal subject.

121. The method of any of claims 111-120, wherein the subject is a human.

**FIG. 1A** MUT cells (R255X)



**FIG. 1B**



**FIG. 1C**

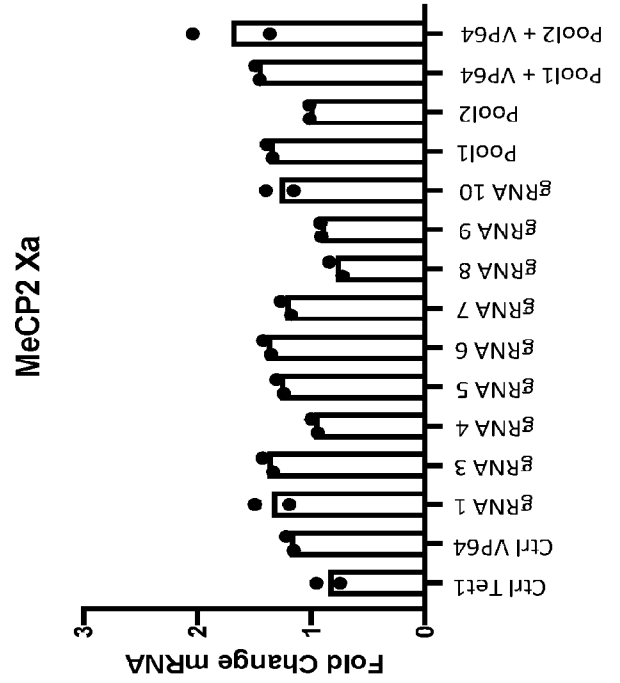


FIG. 2

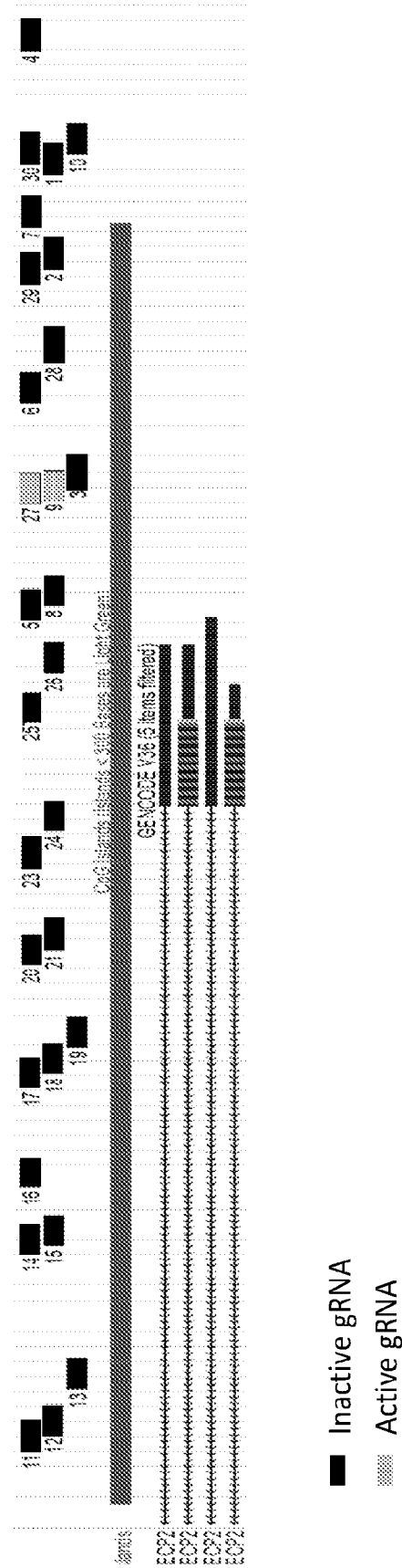


FIG. 3A

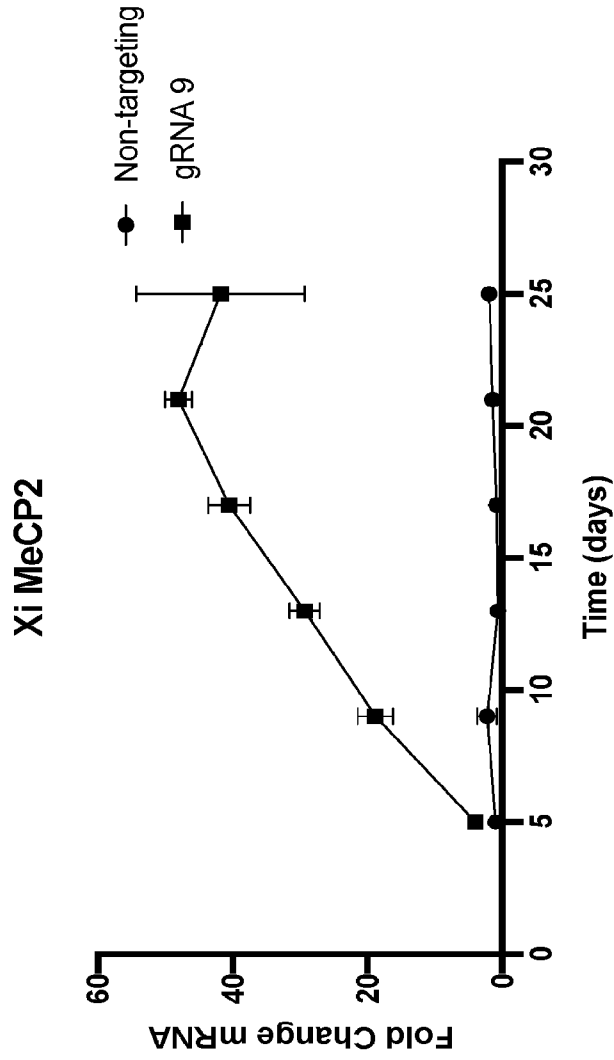


FIG. 3B

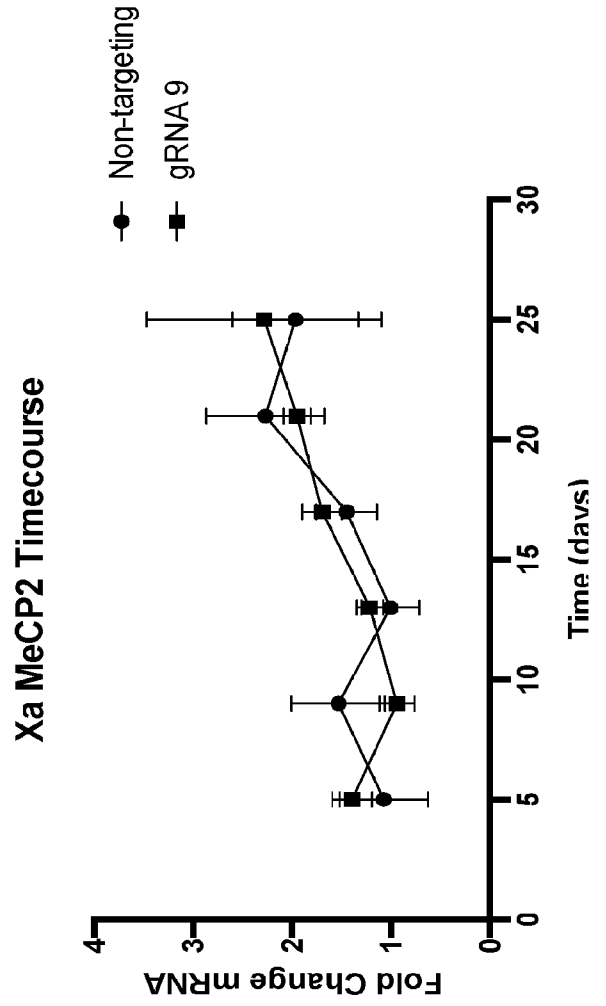




FIG. 4A

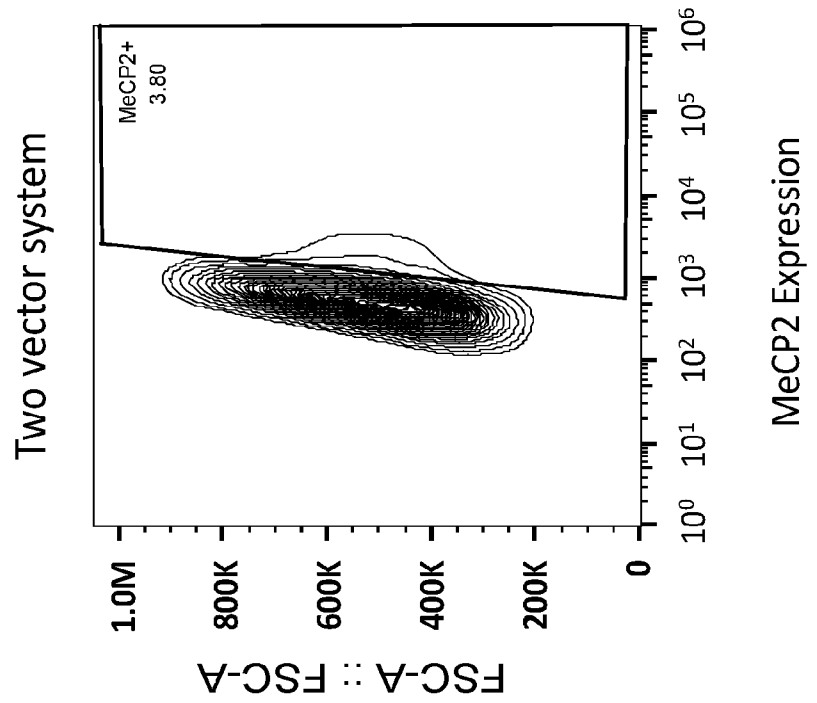


FIG. 4B

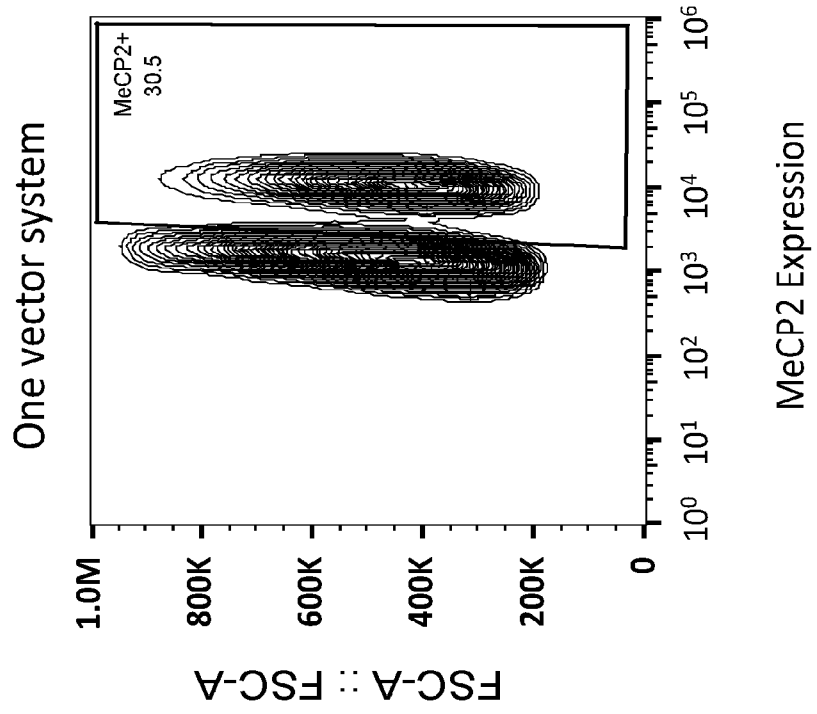
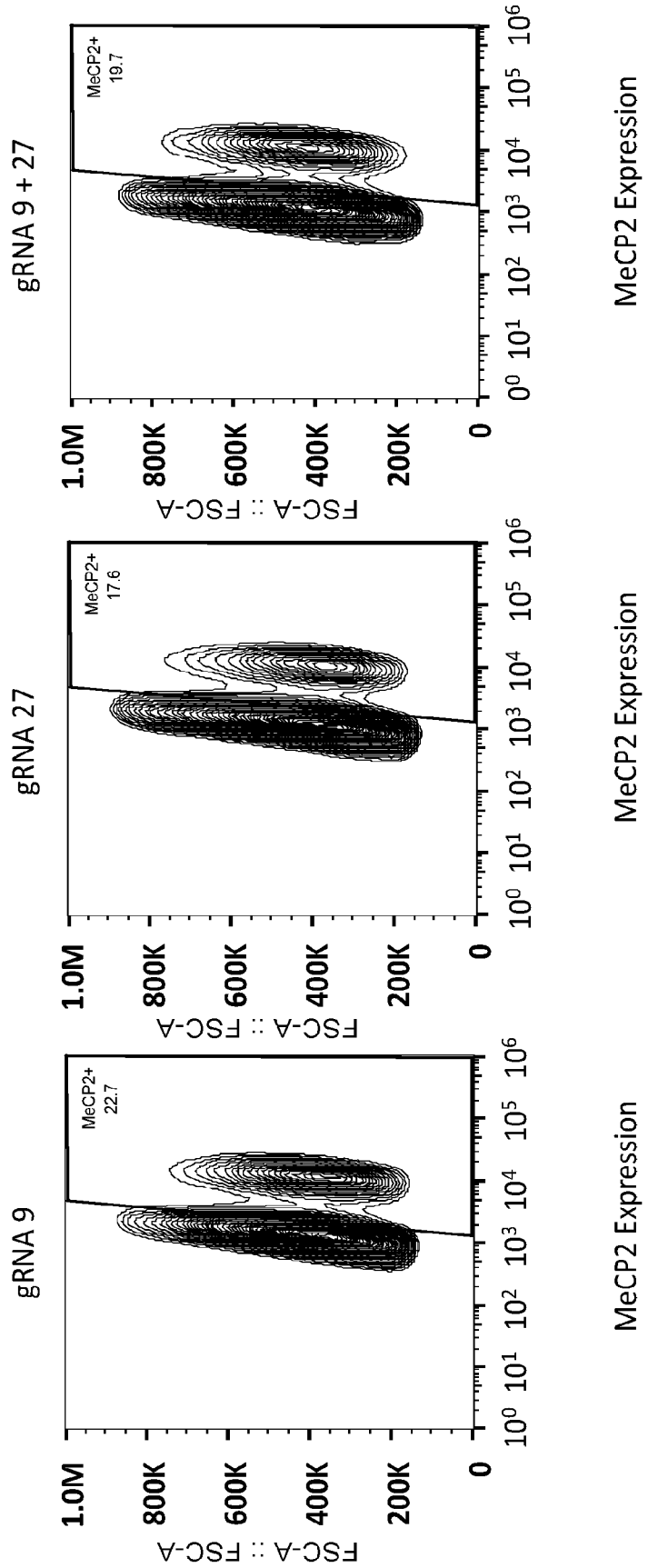


FIG. 4C



**FIG. 5**

R255X

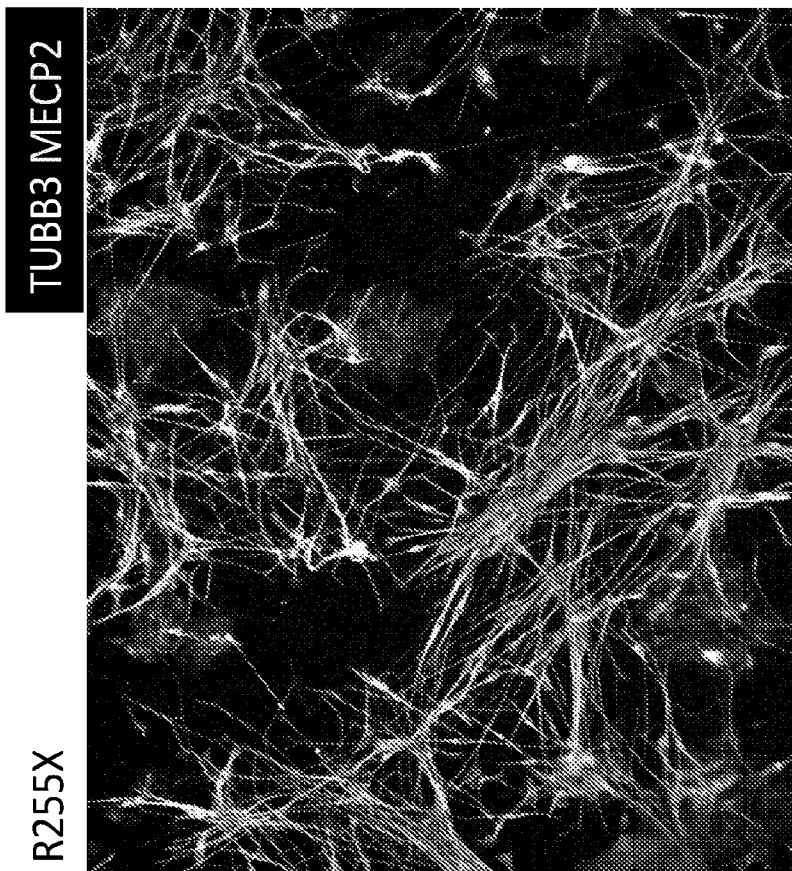


FIG. 6

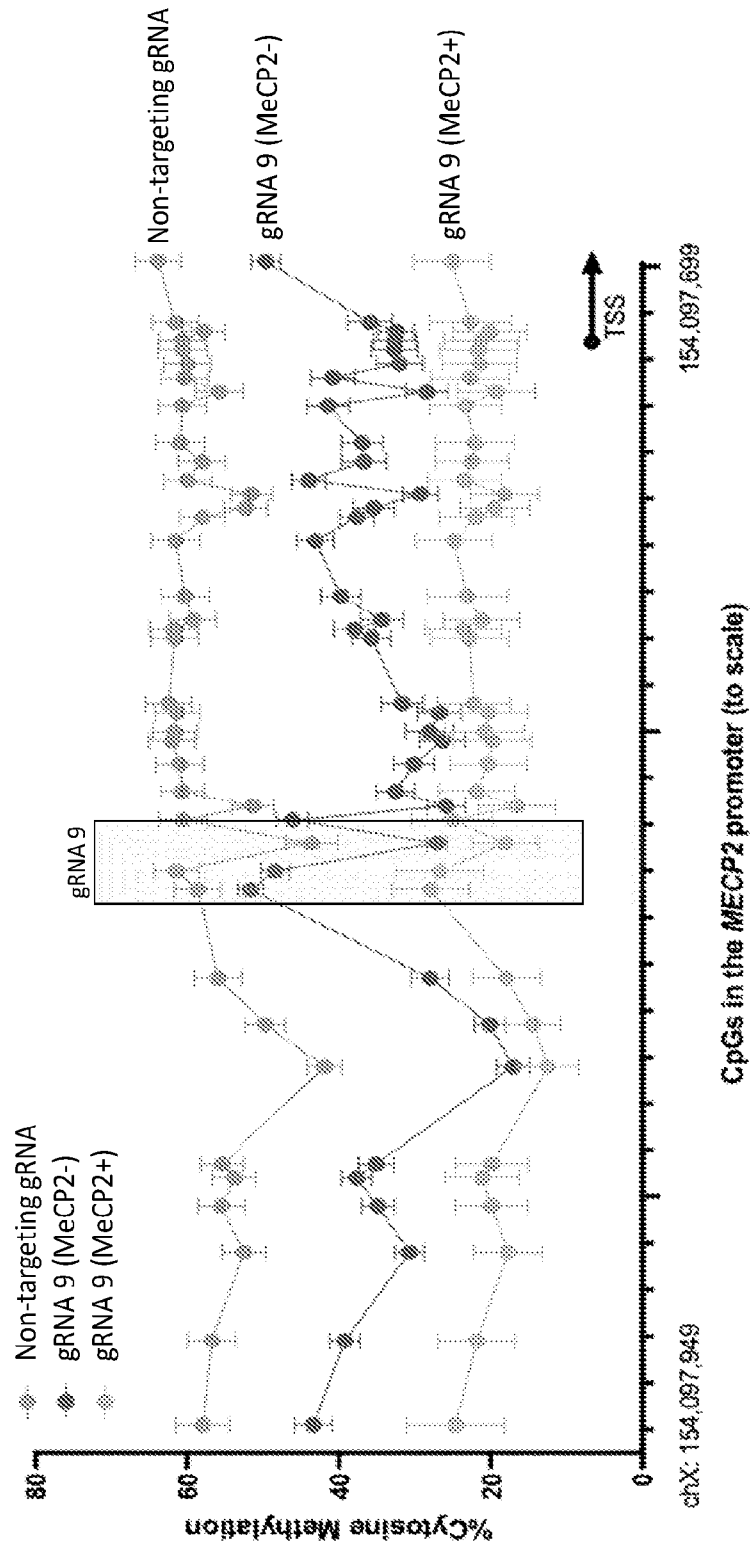


FIG. 7A

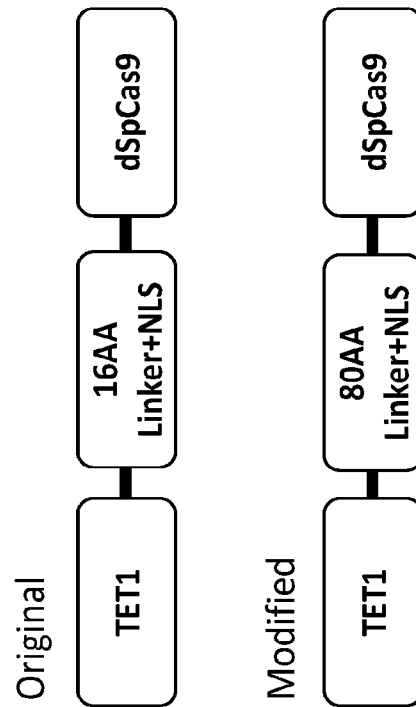
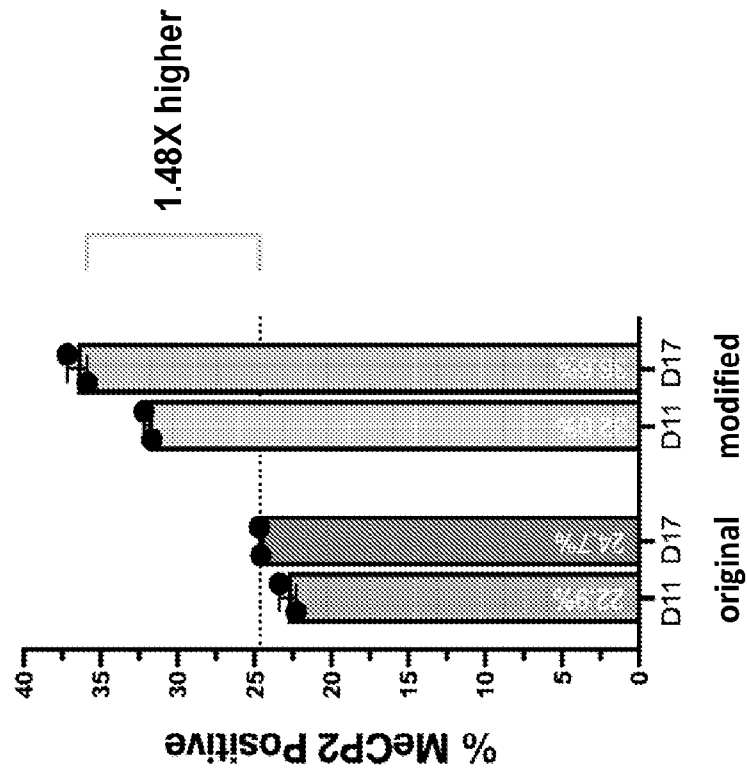


FIG. 7B



**FIG. 8**

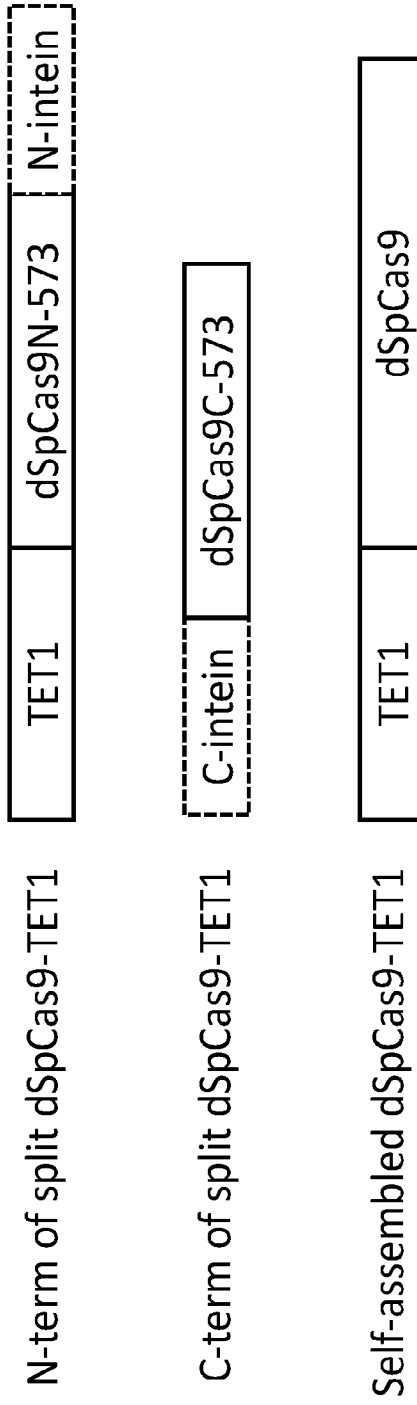


FIG. 9

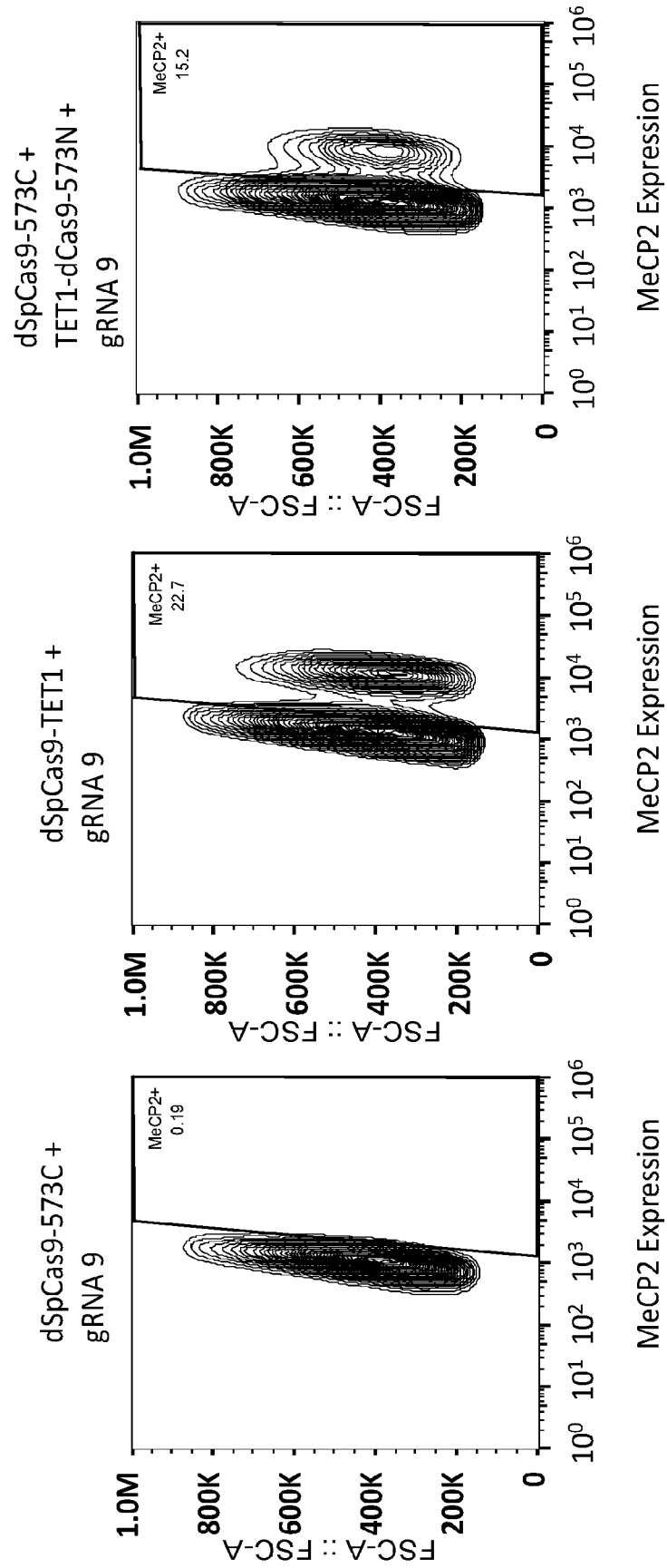


FIG. 10

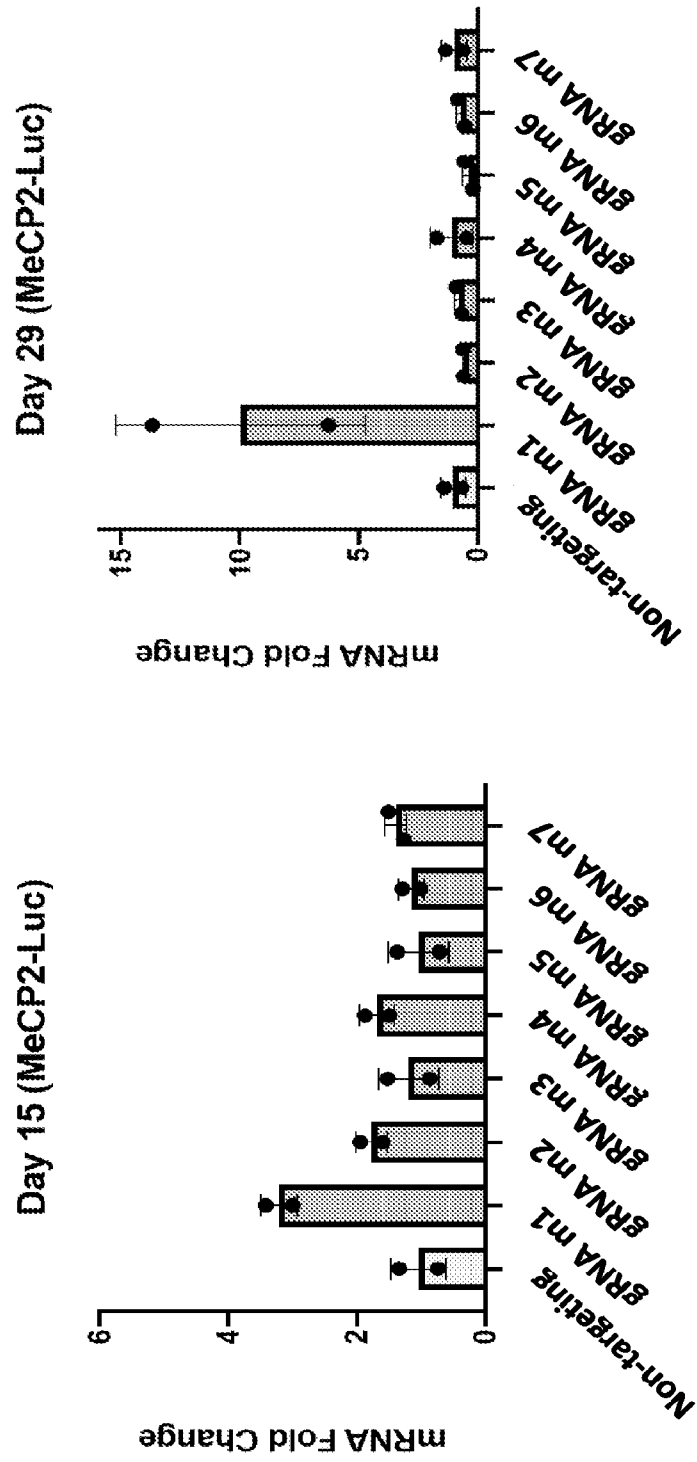




FIG. 3A

Xi MeCP2

