Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3100037 A1 2019/11/14

(21) 3 100 037

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2019/05/11
(87) Date publication PCT/PCT Publication Date: 2019/11/14
(85) Entrée phase nationale/National Entry: 2020/11/10
(86) N° demande PCT/PCT Application No.: US 2019/031899
(87) N° publication PCT/PCT Publication No.: 2019/217944
(30) Priorités/Priorities: 2018/05/11 (US62/670,588); 2018/12/17 (US62/780,838); 2019/03/13 (US62/817,986)

(51) Cl.Int./Int.Cl. *C12N 9/22* (2006.01), *C12N 15/113* (2010.01), *C12N 9/78* (2006.01)

(71) Demandeur/Applicant:
BEAM THERAPEUTICS INC., US

(72) Inventeurs/Inventors:
BRYSON, DAVID, US;
EVANS, JOHN, US;
PACKER, MICHAEL, US;
GEHRKE, JASON MICHAEL, US;
PETROSSIAN, NATALIE, US

(74) Agent: NORTON ROSE FULBRIGHT CANADA LLP/S.E.N.C.R.L., S.R.L.

(54) Titre : PROCEDES D'EDITION DE POLYMORPHISME MONONUCLEOTIDIQUE A L'AIDE DE SYSTEMES D'EDITEUR DE BASES PROGRAMMABLES

(54) Title: METHODS OF EDITING SINGLE NUCLEOTIDE POLYMORPHISM USING PROGRAMMABLE BASE EDITOR SYSTEMS

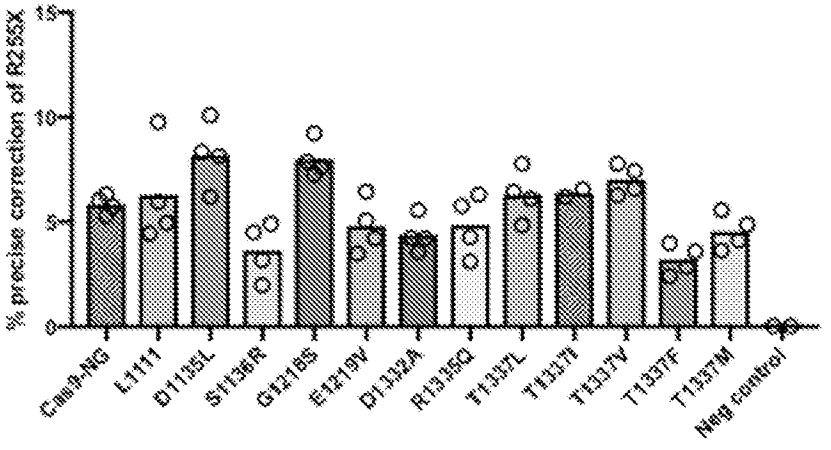

FIG. 1

(57) Abrégé/Abstract:
The invention features compositions and methods for altering mutations associated with Rett Syndrome (RTT). Provided herein are compositions and methods of using base editors comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain in conjunction with a guide polynucleotide. Also provided herein are base editor systems for editing nucleobases of target nucleotide sequences.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 November 2019 (14.11.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/217944 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 9/78* (2006.01)
C12N 15/113 (2010.01)

(21) International Application Number:
PCT/US2019/031899

(22) International Filing Date:
11 May 2019 (11.05.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/670,588	11 May 2018 (11.05.2018)	US
62/780,838	17 December 2018 (17.12.2018)	US
62/817,986	13 March 2019 (13.03.2019)	US

(71) Applicant: **BEAM THERAPEUTICS INC.** [US/US]; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **BRYSON, David**; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US). **EVANS, John**; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US). **PACKER, Michael**; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US). **GEHRKE, Jason Michael**; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US). **PETROSSIAN, Natalie**; 26 Landsdowne St., 2nd Floor, Cambridge, Massachusetts 02139 (US).

(74) Agent: **GAO, HAN**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (US).

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

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

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

(54) Title: METHODS OF EDITING SINGLE NUCLEOTIDE POLYMORPHISM USING PROGRAMMABLE BASE EDITOR SYSTEMS

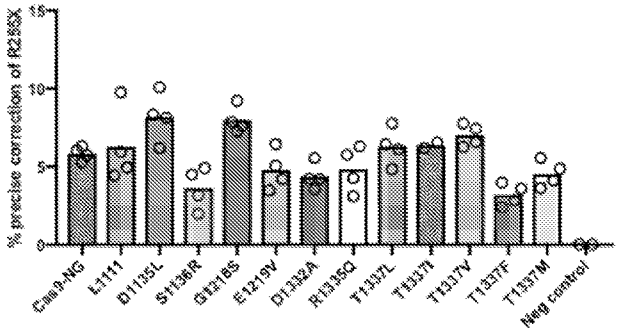

FIG. 1

(57) Abstract: The invention features compositions and methods for altering mutations associated with Rett Syndrome (RTT). Provided herein are compositions and methods of using base editors comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain in conjunction with a guide polynucleotide. Also provided herein are base editor systems for editing nucleobases of target nucleotide sequences.

WO 2019/217944 A1

METHODS OF EDITING SINGLE NUCLEOTIDE POLYMORPHISM USING PROGRAMMABLE BASE EDITOR SYSTEMS

RELATED APPLICATIONS

[0001] This application claims priority to and benefit of U.S. Provisional applications U.S. Ser. No. 62/670,588, filed May 11, 2018, U.S. Ser. No. 62/780,838, filed Dec. 17, 2018; and U.S. Ser. No. No. 62/817,986, filed March 13, 2019, each of which is incorporated herein by reference in their entireties.

BACKGROUND OF THE DISCLOSURE

[0002] Rett Syndrome (RTT or RETT) is caused by a heterogeneous group of mutations in the methyl-CpG-binding protein 2 (*Mecp2)* gene that impair or abrogate the encoded protein's ability to modify chromatin and transcriptional states in the central nervous system (CNS). Gene therapy to deliver functional *Mecp2* or using RNA editing to repair the endogenous *Mecp2* mRNA transcripts are promising approaches to therapeutic interventions when delivered broadly throughout the CNS. However, both approaches must overcome significant challenges to achieve therapeutic efficacy. *Mecp2* gene therapy must tightly control the dosage of the delivered gene on a per-cell basis or risk mimicking the phenotype of *Mecp2* duplication syndrome. RNA editing platforms are unable to precisely correct the most prevalent *Mecp2* mutations accounting for more than 45% of RTT diagnoses and also induce efficient, unguided off-target editing.

[0003] The genetic mutations in *Mecp2* that cause Rett Syndrome (RTT) are highly heterogeneous. As a consequence, the favored strategy for therapy has been to deliver wild-type *Mecp2* carried by recombinant adeno-associated virus (rAAV). Because this strategy is agnostic to the causal mutation in each individual, a successful gene therapy approach would provide a therapeutic option to a large portion of the RTT patient population. To date, however, this strategy has been met with limited success. RTT patients are nearly always heterozygotic females, resulting in characteristic wild-type and mutant X- linked MeCP2 mosaic expression within the central nervous system (CNS) due to random X- chromosome inactivation. Thus, rAAV delivery and expression of wild-type MeCP2 in neurons already expressing wild-type MeCP2 is likely to partially mimic the phenotype of MeCP2 duplication syndrome. Consistent with this, high transduction efficiency in the CNS of RTT-model mice resulted in approximately 2-fold greater MeCP2 expression than found in wild-type mice.

[0004] Therefore, there is a need for novel compositions and methods for treating patients with Rett Syndrome.

INCORPORATION BY REFERENCE

[0005] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. Absent any indication otherwise, publications, patents, and patent applications mentioned in this specification are incorporated herein by reference in their entireties.

SUMMARY OF THE DISCLOSURE

[0006] As described below, the present invention features compositions and methods for the precise correction of pathogenic amino acids using a programmable nucleobase editor. In particular, the compositions and methods of the invention are useful for the treatment of Rett Syndrome (RTT). Thus, the invention provides compositions and methods for treating Rett Syndrome using an adenosine (A) base editor (ABE) to precisely correct a single nucleotide polymorphism in the endogenous *Mecp2* gene to correct a deleterious mutation (e.g., R133C, T158M, R255*, R270*, R306C).

[0007] In one aspect, the invention provides a method of editing an MECP2 polynucleotide containing a single nucleotide polymorphism (SNP) associated with Rett Syndrome (RTT), the method involving contacting the MECP2 polynucleotide with a base editor in complex with one or more guide polynucleotides, where the base editor includes a polynucleotide programmable DNA binding domain and an adenosine deaminase domain, and where one or more of the guide polynucleotides target the base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

[0008] In another aspect, the invention provides a cell produced by introducing into the cell, or a progenitor thereof: a base editor, or a polynucleotide encoding the base editor, to the cell, where the base editor includes a polynucleotide programmable DNA binding domain and an adenosine deaminase domain; and one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

[0009] In another aspect, the invention provides a method of treating RTT in a subject involving administering to said subject: a base editor, or a polynucleotide encoding the base editor, to the subject, where the base editor includes a polynucleotide programmable DNA

binding domain and an adenosine deaminase domain; and one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

[0010] In another aspect, the invention provides a base editor comprising: (i) a modified SpCas9 comprising the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, and T1337M, or corresponding amino acid substitutions thereof; and (ii) an adenosine deaminase.

[0011] In another aspect, the invention provides a base editor comprising: (i) a modified SpCas9 comprising the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof; and (ii) an adenosine deaminase.

[0012] In various embodiments, the contacting is in a cell, a eukaryotic cell, a mammalian cell, or human cell. In various embodiments, the cell is in vivo or ex vivo. In various embodiments, the alteration is one or more of R106W, R168*, R133C, T158M, R255*, R270*, and R306C. In various embodiments, the A•T to G•C alteration at the SNP associated with RTT changes a cysteine to an arginine, methionine to a threonine, or stop codon to arginine in the methyl CpG binding protein 2 (MeCP2) polypeptide. In various embodiments, the SNP associated with RTT results in expression of an Mecp2 polypeptide comprising an arginine at amino acid position 168, 133, 255, 270, or 306; or a threonine at position 158. In various embodiments, the polynucleotide programmable DNA binding domain is a Streptococcus pyogenes Cas9 (SpCas9) or variant thereof. In various embodiments, the polynucleotide programmable DNA binding domain comprises a modified SpCas9 having an altered protospacer-adjacent motif (PAM) specificity. In various embodiments, the altered PAM has specificity for the nucleic acid sequence 5'-NGT-3'. In various embodiments, the modified SpCas9 comprises the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof. In various embodiments, the modified SpCas9 comprises the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N,

T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof. In various embodiments, the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant. In various embodiments, the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof. In various embodiments, the adenosine deaminase domain is capable of deaminating adenosine in deoxyribonucleic acid (DNA). In various embodiments, the adenosine deaminase is a modified adenosine deaminase that does not occur in nature. In various embodiments, the adenosine deaminase is a TadA deaminase. In various embodiments, the TadA deaminase is TadA*7.10. In various embodiments, the one or more guide RNAs comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), where the crRNA comprises a nucleic acid sequence complementary to a *Mecp2* nucleic acid sequence comprising the SNP associated with RTT. In various embodiments, the base editor is in complex with a single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an *Mecp2* nucleic acid sequence comprising the SNP associated with RTT. In various embodiments, the cell is a neuron. In various embodiments, the neuron expresses an Mecp2 polypeptide. In various embodiments, the cell is from a subject having RTT.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The features of the present disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0014] FIG. 1 is a graph depicting percentage of precise correction of R255X RTT mutation using base editor variants with specificity for NGT PAM.

[0015] FIG. 2 is a graph depicting percentage of precise correction of R255X RTT mutation using base editor variants with specificity for NGT PAM.

[0016] FIG. 3 is a graph depicting PAM variant optimization with the amino acid substitution T1337L. Percentage of precise correction of R255X RTT mutation is shown using base editor variants with specificity for NGT PAM.

[0017] FIG. 4 is a graph depicting percentage of precise correction of R255X RTT mutation by PAM base editor variants with specificity for NGT PAM generated by shuffling mutations from other characterized PAM variants. T1337Q was identified as important for editing efficiency.

[0018] FIG. 5 is a graph depicting changes in base editing efficiency when T1337 and D1332 are substituted with other amino acids. Percentage of precise correction of R255X RTT mutation is shown using base editor variants with specificity for NGT PAM.

[0019] FIG. 6 is a graph depicting the importance of E1219V and R1335Q for base editing activity associated with T1337. Percentage of precise correction of R255X RTT mutation is shown using base editor variants with specificity for NGT PAM.

[0020] FIG. 7 is a graph depicting percentage of precise correction of R255X RTT mutation using PAM variant base editors with specificity for NGT PAM variants listed at Tables 8 and 9.

[0021] FIG. 8 is a graph depicting percentage of precise correction of R255X RTT mutation using PAM variant base editors with specificity for NGT PAM variants listed at Tables 8 and 9.

[0022] FIG. 9 is a graph depicting percentage of precise correction of R255X RTT mutation using PAM variant base editors with specificity for NGT PAM variants listed at Table 10.

DETAILED DESCRIPTION OF THE DISCLOSURE

[0023] Described herein are compositions and methods providing base editing and base editing systems to precisely correct one or more mutations in the methyl-CpG-binding protein 2 (*Mecp2*) gene, which is causally related to the progressive neurodevelopmental disorder Rett Syndrome (RTT) and its symptoms. RTT is an X-linked dominant disorder that predominantly affects females, is associated in 96% of affected individuals with mutations in the *Mecp2* gene and is characterized by apparently normal early development followed by a regression with loss of fine motor skills and effective communication, stereotypic movements, and apraxia or complete absence of gait. Additional clinical features of afflicted individuals include abnormal postnatal deceleration in the rate of head growth, periodic breathing, gastrointestinal dysfunction, epilepsy, and scoliosis.

[0024] The most prevalent RTT-causing mutations are cytidine to thymidine (C→T) transition mutations, resulting in a C•G to T•A base pair substitution. This substitution may be reverted back to a wild-type, non-pathogenic genomic sequence with an adenosine base editor (ABE) which catalyzes A•T to G•C substitutions. By extension, highly prevalent RTT-causing mutations are potential targets for reversion to wild-type sequence using ABEs without the risks of inducing *Mecp2* gene overexpression, as may occur using gene therapy. Accordingly, A•T to G•C DNA base editing has the potential to precisely correct one or more of the most prevalent RTT-causing mutations in the *Mecp2* gene.

[0025] The following description and examples illustrate embodiments of the present disclosure in detail. It is to be understood that this disclosure is not limited to the particular

embodiments described herein and as such can vary. Those of skill in the art will recognize that there are numerous variations and modifications of this disclosure, which are encompassed within its scope.

[0026] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0027] Although various features of the present disclosure can be described in the context of a single embodiment, the features can also be provided separately or in any suitable combination. Conversely, although the present disclosure can be described herein in the context of separate embodiments for clarity, the present disclosure can also be implemented in a single embodiment.

DEFINITIONS

[0028] Unless defined otherwise, all technical and scientific terms as used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991).

[0029] In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, use of the term "including" as well as other forms, such as "include", "includes," and "included," is not limiting.

[0030] As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the present disclosure, and vice versa. Furthermore, compositions of the present disclosure can be used to achieve methods of the present disclosure.

[0031] The term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system. For

example, "about" can mean within 1 or more than 1 standard deviation, per the practice in the art. Alternatively, "about" can mean a range of up to 20%, up to 10%, up to 5%, or up to 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term "about" meaning within an acceptable error range for the particular value should be assumed.

[0032] Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

[0033] Reference in the specification to "some embodiments," "an embodiment," "one embodiment" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the present disclosures.

[0034] By "adenosine deaminase" is meant a polypeptide or fragment thereof capable of catalyzing the hydrolytic deamination of adenine or adenosine. In some embodiments, the deaminase or deaminase domain is an adenosine deaminase catalyzing the hydrolytic deamination of adenosine to inosine or deoxy adenosine to deoxyinosine. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in deoxyribonucleic acid (DNA). The adenosine deaminases (e.g. engineered adenosine deaminases, evolved adenosine deaminases) provided herein may be from any organism, such as a bacterium.

[0035] By "agent" is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

[0036] By "ameliorate" is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of a disease.

[0037] By "alteration" is meant a change (increase or decrease) in the expression levels or activity of a gene or polypeptide as detected by standard art known methods such as those described herein. As used herein, an alteration includes a 10% change in expression levels, preferably a 25% change, more preferably a 40% change, and most preferably a 50% or greater change in expression levels.

[0038] By "analog" is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polypeptide analog retains the biological activity of a corresponding naturally-occurring polypeptide, while having certain biochemical modifications that enhance the analog's function relative to a naturally occurring polypeptide. Such biochemical modifications could increase the analog's protease resistance, membrane permeability, or half-life, without altering, for example, ligand binding. An analog may include an unnatural amino acid.

[0039] "Administering" is referred to herein as providing one or more compositions described herein to a patient or a subject. By way of example and without limitation, composition administration, *e.g.*, injection, can be performed by intravenous (i.v.) injection, sub-cutaneous (s.c.) injection, intradermal (i.d.) injection, intraperitoneal (i.p.) injection, or intramuscular (i.m.) injection. One or more such routes can be employed. Parenteral administration can be, for example, by bolus injection or by gradual perfusion over time. Alternatively, or concurrently, administration can be by the oral route.

[0040] By "cytidine deaminase" is meant a polypeptide or fragment thereof capable of catalyzing a deamination reaction that converts an amino group to a carbonyl group. In one embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. PmCDA1, which is derived from Petromyzon marinus (Petromyzon marinus cytosine deaminase 1, "PmCDA1"), AID (Activation-induced cytidine deaminase; AICDA), which is derived from a mammal (e.g., human, swine, bovine, horse, monkey etc.), and APOBEC are exemplary cytidine deaminases.

[0041] By "methyl CpG binding protein 2 (Mecp2) protein" is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to NCBI Accession No. NP_004983. In particular embodiments, an Mecp2 protein comprises one or more alterations relative to the following reference sequence. In particular embodiments, an Mecp2 protein associated with RTT comprises one or more mutations selected from R106W, R168*, R133C, T158M, R255*, R270*, and R306C. An exemplary Mecp2 amino acid sequence is provided below.

```
  1 mvagmlglre eksedqdlqg lkdkplkfkk vkkdkkeeke gkhepvqpsa hhsaepaeag
 61 kaetsegsgs apavpeasas pkqrrsiird rgpmyddptl pegwtrklkq rksgrsagky
121 dvylinpqgk afrskvelia yfekvgdtsl dpndfdftvt grgspsrreq kppkkpkspk
181 apgtgrgrgr pkgsgttrpk aatsegvqvk rvlekspgkl lvkmpfqtsp ggkaegggat
241 tstqvmvikr pgrkrkaead pqaipkkrgr kpgsvvaaaa aeakkkavke ssirsvqetv
301 lpikkrktre tvsievkevv kpllvstlge ksgkglktck spgrkskess pkgrsssass
361 ppkkehhhhh hhsespkapv pllpplpppp pepessedpt sppepqdlss svckeekmpr
421 ggslesdgcp kepaktqpav ataataaeky khrgegerkd ivsssmprpn reepvdsrtp
```

481 vtervs

[0042] By "*Mecp2* polynucleotide" is meant a nucleic acid molecule encoding an Mecp2 protein or fragment thereof. The sequence of an exemplary *Mecp2* polynucleotide, which is available at NCBI Accession No. NM_004992, is provided below. In particular embodiments, an *Mecp2* polynucleotide comprises one or more alterations relative to the following reference sequence. In particular embodiments, an *Mecp2* polynucleotide associated with RTT comprises one or more mutations selected from 316C>T, 397C>T, 473C>T, 763C>T, 808C>T and 916C>T.

```
   1 ccggcgtcgg cggcgcgcgc gctccctcct ctcggagaga gggctgtggt aaaagccgtc
  61 cggaaaatgg ccgccgccgc cgccgccgcg ccgagcggag gaggaggagg aggcgaggag
 121 gagagactgc tccataaaaa tacagactca ccagttcctg ctttgatgtg acatgtgact
 181 ccccagaata caccttgctt ctgtagacca gctccaacag gattccatgg tagctgggat
 241 gttagggctc agggaagaaa agtcagaaga ccaggacctc cagggcctca aggacaaacc
 301 cctcaagttt aaaaaggtga agaaagataa gaaagaagag aaagagggca agcatgagcc
 361 cgtgcagcca tcagcccacc actctgctga gcccgcagag gcaggcaaag cagagacatc
 421 agaagggtca ggctccgccc cggctgtgcc ggaagcttct gcctccccca aacagcggcg
 481 ctccatcatc cgtgaccggg gacccatgta tgatgacccc accctgcctg aaggctggac
 541 acggaagctt aagcaaagga aatctggccg ctctgctggg aagtatgatg tgtatttgat
 601 caatccccag ggaaaagcct ttcgctctaa agtggagttg attgcgtact tcgaaaaggt
 661 aggcgacaca tccctggacc ctaatgattt tgacttcacg gtaactggga gagggagccc
 721 ctcccggcga gagcagaaac cacctaagaa gcccaaatct cccaaagctc caggaactgg
 781 cagaggccgg ggacgcccca aagggagcgg caccacgaga cccaaggcgg ccacgtcaga
 841 gggtgtgcag gtgaaaaggg tcctggagaa aagtcctggg aagctccttg tcaagatgcc
 901 ttttcaaact tcgccagggg gcaaggctga ggggggtggg gccaccacat ccacccaggt
 961 catggtgatc aaacgccccg gcaggaagcg aaaagctgag gccgaccctc aggccattcc
1021 caagaaacgg ggccgaaagc cggggagtgt ggtggcagcc gctgccgccg aggccaaaaa
1081 gaaagccgtg aaggagtctt ctatccgatc tgtgcaggag accgtactcc ccatcaagaa
1141 gcgcaagacc cgggagacgg tcagcatcga ggtcaaggaa gtggtgaagc ccctgctggt
1201 gtccaccctc ggtgagaaga gcgggaaagg actgaagacc tgtaagagcc ctgggcggaa
1261 aagcaaggag agcagcccca aggggcgcag cagcagcgcc tcctcacccc ccaagaagga
1321 gcaccaccac catcaccacc actcagagtc cccaaaggcc cccgtgccac tgctcccacc
1381 cctgccccca cctccacctg agcccgagag ctccgaggac cccaccagcc cccctgagcc
1441 ccaggacttg agcagcagcg tctgcaaaga ggagaagatg cccagaggag gctcactgga
1501 gagcgacggc tgccccaagg agccagctaa gactcagccc gcggttgcca ccgccgccac
1561 ggccgcagaa aagtacaaac accgagggga gggagagcgc aaagacattg tttcatcctc
1621 catgccaagg ccaaacagag aggagcctgt ggacagccgg acgcccgtga ccgagagagt
1681 tagctgactt tacacggagc ggattgcaaa gcaaaccaac aagaataaag gcagctgttg
1741 tctcttctcc ttatgggtag ggctctgaca aagcttcccg attaactgaa ataaaaaata
1801 ttttttttttc tttcagtaaa cttagagttt cgtggcttca gggtgggagt agttggagca
1861 ttggggatgt ttttcttacc gacaagcaca gtcaggttga agacctaacc agggccagaa
```

1921 gtagctttgc acttttctaa actaggctcc ttcaacaagg cttgctgcag atactactga
1981 ccagacaagc tgttgaccag gcacctcccc tcccgcccaa acctttcccc catgtggtcg
2041 ttagagacag agcgacagag cagttgagag gacactcccg ttttcggtgc catcagtgcc
2101 ccgtctacag ctcccccagc tccccccacc tcccccactc ccaaccacgt tgggacaggg
2161 aggtgtgagg caggagagac agttggattc tttagagaag atggatatga ccagtggcta
2221 tggcctgtgc gatcccaccc gtggtggctc aagtctggcc ccacaccagc cccaatccaa
2281 aactggcaag gacgcttcac aggacaggaa agtggcacct gtctgctcca gctctggcat
2341 ggctaggagg ggggagtccc ttgaactact gggtgtagac tggcctgaac cacaggagag
2401 gatggcccag ggtgaggtgg catggtccat tctcaaggga cgtcctccaa cgggtggcgc
2461 tagaggccat ggaggcagta ggacaaggtg caggcaggct ggcctggggt caggccgggc
2521 agagcacagc ggggtgagag ggattcctaa tcactcagag cagtctgtga cttagtggac
2581 aggggagggg gcaaaggggg aggagaagaa aatgttcttc cagttacttt ccaattctcc
2641 tttagggaca gcttagaatt atttgcacta ttgagtcttc atgttcccac ttcaaaacaa
2701 acagatgctc tgagagcaaa ctggcttgaa ttggtgacat ttagtccctc aagccaccag
2761 atgtgacagt gttgagaact acctggattt gtatatatac ctgcgcttgt tttaaagtgg
2821 gctcagcaca tagggttccc acgaagctcc gaaactctaa gtgtttgctg caatttttata
2881 aggacttcct gattggtttc tcttctcccc ttccatttct gccttttgtt catttcatcc
2941 tttcacttct ttcccttcct ccgtcctcct ccttcctagt tcatcccttc tcttccaggc
3001 agccgcggtg cccaaccaca cttgtcggct ccagtcccca gaactctgcc tgccctttgt
3061 cctcctgctg ccagtaccag ccccaccctg ttttgagccc tgaggaggcc ttgggctctg
3121 ctgagtccga cctggcctgt ctgtgaagag caagagagca gcaaggtctt gctctcctag
3181 gtagcccccc cttccctggt aagaaaaagc aaaaggcatt tcccaccctg aacaacgagc
3241 cttttcaccc ttctactcta gagaagtgga ctggaggagc tgggcccgat ttggtagttg
3301 aggaaagcac agaggcctcc tgtggcctgc cagtcatcga gtggcccaac aggggctcca
3361 tgccagccga ccttgacctc actcagaagt ccagagtcta gcgtagtgca gcagggcagt
3421 agcggtacca atgcagaact cccaagaccc gagctgggac cagtacctgg gtccccagcc
3481 cttcctctgc tccccctttt ccctcggagt tcttcttgaa tggcaatgtt ttgcttttgc
3541 tcgatgcaga caggggggcca gaacaccaca catttcactg tctgtctggt ccatagctgt
3601 ggtgtagggg cttagaggca tgggcttgct gtgggttttt aattgatcag ttttcatgtg
3661 ggatcccatc tttttaacct ctgttcagga agtccttatc tagctgcata tcttcatcat
3721 attggtatat ccttttctgt gtttacagag atgtctctta tatctaaatc tgtccaactg
3781 agaagtacct tatcaaagta gcaaatgaga cagcagtctt atgcttccag aaacacccac
3841 aggcatgtcc catgtgagct gctgccatga actgtcaagt gtgtgttgtc ttgtgtattt
3901 cagttattgt ccctggcttc cttactatgg tgtaatcatg aaggagtgaa acatcataga
3961 aactgtctag cacttccttg ccagtcttta gtgatcagga accatagttg acagttccaa
4021 tcagtagctt aagaaaaaac cgtgtttgtc tcttctggaa tggttagaag tgagggagtt
4081 tgccccgttc tgtttgtaga gtctcatagt tggactttct agcatatatg tgtccatttc
4141 cttatgctgt aaaagcaagt cctgcaacca aactcccatc agcccaatcc ctgatccctg
4201 atcccttcca cctgctctgc tgatgacccc cccagcttca cttctgactc ttccccagga
4261 agggaagggg ggtcagaaga gagggtgagt cctccagaac tcttcctcca aggacagaag
4321 gctcctgccc ccatagtggc ctcgaactcc tggcactacc aaaggacact tatccacgag
4381 agcgcagcat ccgaccaggt tgtcactgag aagatgttta ttttggtcag ttgggttttt

4441 atgtattata cttagtcaaa tgtaatgtgg cttctggaat cattgtccag agctgcttcc
4501 ccgtcacctg ggcgtcatct ggtcctggta agaggagtgc gtggcccacc aggcccccct
4561 gtcacccatg acagttcatt cagggccgat ggggcagtcg tggttgggaa cacagcattt
4621 caagcgtcac tttatttcat tcgggcccca cctgcagctc cctcaaagag gcagttgccc
4681 agcctctttc ccttccagtt tattccagag ctgccagtgg ggcctgaggc tccttagggt
4741 tttctctcta tttccccctt tcttcctcat tccctcgtct ttcccaaagg catcacgagt
4801 cagtcgcctt tcagcaggca gccttggcgg tttatcgccc tggcaggcag gggccctgca
4861 gctctcatgc tgcccctgcc ttggggtcag gttgacagga ggttggaggg aaagccttaa
4921 gctgcaggat tctcaccagc tgtgtccggc ccagttttgg ggtgtgacct caatttcaat
4981 tttgtctgta cttgaacatt atgaagatgg gggcctcttt cagtgaattt gtgaacagca
5041 gaattgaccg acagctttcc agtacccatg gggctaggtc attaaggcca catccacagt
5101 ctcccccacc cttgttccag ttgttagtta ctacctcctc tcctgacaat actgtatgtc
5161 gtcgagctcc ccccaggtct acccctcccg gccctgcctg ctggtgggct tgtcatagcc
5221 agtgggattg ccggtcttga cagctcagtg agctggagat acttggtcac agccaggcgc
5281 tagcacagct cccttctgtt gatgctgtat tcccatatca aaagacacag gggacaccca
5341 gaaacgccac atcccccaat ccatcagtgc caaactagcc aacggcccca gcttctcagc
5401 tcgctggatg gcggaagctg ctactcgtga gcgccagtgc gggtgcagac aatcttctgt
5461 tgggtggcat cattccaggc ccgaagcatg aacagtgcac ctgggacagg gagcagcccc
5521 aaattgtcac ctgcttctct gcccagcttt tcattgctgt gacagtgatg gcgaaagagg
5581 gtaataacca gacacaaact gccaagttgg gtggagaaag gagtttcttt agctgacaga
5641 atctctgaat tttaaatcac ttagtaagcg gctcaagccc aggagggagc agagggatac
5701 gagcggagtc ccctgcgcgg gaccatctgg aattggttta gcccaagtgg agcctgacag
5761 ccagaactct gtgtcccccg tctaaccaca gctccttttc cagagcattc cagtcaggct
5821 ctctgggctg actgggccag gggaggttac aggtaccagt tctttaagaa gatctttggg
5881 catatacatt tttagcctgt gtcattgccc caaatggatt cctgtttcaa gttcacacct
5941 gcagattcta ggacctgtgt cctagacttc agggagtcag ctgtttctag agttcctacc
6001 atggagtggg tctggaggac ctgcccggtg gggggggcaga gccctgctcc ctccgggtct
6061 tcctactctt ctctctgctc tgacgggatt tgttgattct ctccattttg gtgtctttct
6121 cttttagata ttgtatcaat ctttagaaaa ggcatagtct acttgttata aatcgttagg
6181 atactgcctc ccccagggtc taaaattaca tattagaggg gaaaagctga acactgaagt
6241 cagttctcaa caatttagaa ggaaaaccta gaaaacattt ggcagaaaat tacatttcga
6301 tgtttttgaa tgaatacgag caagctttta caacagtgct gatctaaaaa tacttagcac
6361 ttggcctgag atgcctggtg agcattacag gcaaggggaa tctggaggta gccgacctga
6421 ggacatggct tctgaacctg tcttttggga gtggtatgga aggtggagcg ttcaccagtg
6481 acctggaagg cccagcacca ccctccttcc cactcttctc atcttgacag agcctgcccc
6541 agcgctgacg tgtcaggaaa acacccaggg aactaggaag gcacttctgc ctgaggggca
6601 gcctgccttg cccactcctg ctctgctcgc ctcggatcag ctgagccttc tgagctggcc
6661 tctcactgcc tccccaaggc cccctgcctg ccctgtcagg aggcagaagg aagcaggtgt
6721 gagggcagtg caaggaggga gcacaacccc cagctcccgc tccgggctcc gacttgtgca
6781 caggcagagc ccagaccctg gaggaaatcc tacctttgaa ttcaagaaca tttggggaat
6841 ttggaaatct ctttgccccc aaacccccat tctgtcctac ctttaatcag gtcctgctca
6901 gcagtgagag cagatgaggt gaaaaggcca agaggtttgg ctcctgccca ctgatagccc

6961 ctctccccgc agtgtttgtg tgtcaagtgg caaagctgtt cttcctggtg accctgatta
7021 tatccagtaa cacatagact gtgcgcatag gcctgctttg tctcctctat cctgggcttt
7081 tgttttgctt tttagttttg cttttagttt ttctgtccct tttatttaac gcaccgacta
7141 gacacacaaa gcagttgaat ttttatatat atatctgtat attgcacaat tataaactca
7201 ttttgcttgt ggctccacac acacaaaaaa agacctgtta aaattatacc tgttgcttaa
7261 ttacaatatt tctgataacc atagcatagg acaagggaaa ataaaaaaag aaaaaaaaga
7321 aaaaaaaacg acaaatctgt ctgctggtca cttcttctgt ccaagcagat tcgtggtctt
7381 ttcctcgctt ctttcaaggg ctttcctgtg ccaggtgaag gaggctccag gcagcaccca
7441 ggttttgcac tcttgtttct cccgtgcttg tgaaagaggt cccaaggttc tgggtgcagg
7501 agcgctccct tgacctgctg aagtccggaa cgtagtcggc acagcctggt cgccttccac
7561 ctctgggagc tggagtccac tggggtggcc tgactccccc agtccccttc ccgtgacctg
7621 gtcagggtga gcccatgtgg agtcagcctc gcaggcctcc ctgccagtag ggtccgagtg
7681 tgtttcatcc ttcccactct gtcgagcctg ggggctggag cggagacggg aggcctggcc
7741 tgtctcggaa cctgtgagct gcaccaggta gaacgccagg gaccccagaa tcatgtgcgt
7801 cagtccaagg ggtcccctcc aggagtagtg aagactccag aaatgtccct ttcttctccc
7861 ccatcctacg agtaattgca tttgcttttg taattcttaa tgagcaatat ctgctagaga
7921 gtttagctgt aacagttctt tttgatcatc tttttttaat aattagaaac accaaaaaaa
7981 tccagaaact tgttcttcca aagcagagag cattataatc accagggcca aaagcttccc
8041 tccctgctgt cattgcttct tctgaggcct gaatccaaaa gaaaaacagc cataggccct
8101 ttcagtggcc gggctacccg tgagcccttc ggaggaccag ggctggggca gcctctgggc
8161 ccacatccgg ggccagctcc ggcgtgtgtt cagtgttagc agtgggtcat gatgctcttt
8221 cccacccagc ctgggatagg ggcagaggag gcgaggaggc cgttgccgct gatgtttggc
8281 cgtgaacagg tgggtgtctg cgtgcgtcca cgtgcgtgtt ttctgactga catgaaatcg
8341 acgcccgagt tagcctcacc cggtgacctc tagccctgcc cggatggagc ggggcccacc
8401 cggttcagtg tttctgggga gctggacagt ggagtgcaaa aggcttgcag aacttgaagc
8461 ctgctccttc ccttgctacc acggcctcct ttccgtttga tttgtcactg cttcaatcaa
8521 taacagccgc tccagagtca gtagtcaatg aatatatgac caaatatcac caggactgtt
8581 actcaatgtg tgccgagccc ttgcccatgc tgggctcccg tgtatctgga cactgtaacg
8641 tgtgctgtgt ttgctcccct tccccttcct tctttgccct ttacttgtct ttctggggtt
8701 tttctgtttg ggtttggttt ggtttttatt tctccttttg tgttccaaac atgaggttct
8761 ctctactggt cctcttaact gtggtgttga ggcttatatt tgtgtaattt ttggtgggtg
8821 aaaggaattt tgctaagtaa atctcttctg tgtttgaact gaagtctgta ttgtaactat
8881 gtttaaagta attgttccag agacaaatat ttctagacac tttttcttta caaacaaaag
8941 cattcggagg gaggggggatg gtgactgaga tgagagggga gagctgaaca gatgacccct
9001 gcccagatca gccagaagcc acccaaagca gtggagccca ggagtcccac tccaagccag
9061 caagccgaat agctgatgtg ttgccacttt ccaagtcact gcaaaaccag gttttgttcc
9121 gcccagtgga ttcttgtttt gcttcccctc cccccgagat tattaccacc atcccgtgct
9181 tttaaggaaa ggcaagattg atgtttcctt gaggggagcc aggaggggat gtgtgtgtgc
9241 agagctgaag agctggggag aatggggctg ggcccaccca agcaggaggc tgggacgctc
9301 tgctgtgggc acaggtcagg ctaatgttgg cagatgcagc tcttcctgga caggccaggt
9361 ggtgggcatt ctctctccaa ggtgtgcccc gtgggcatta ctgtttaaga cacttccgtc
9421 acatcccacc ccatcctcca gggctcaaca ctgtgacatc tctattcccc accctccct

```
 9481 tcccagggca ataaaatgac catggagggg gcttgcactc tcttggctgt cacccgatcg
 9541 ccagcaaaac ttagatgtga gaaaacccct tcccattcca tggcgaaaac atctccttag
 9601 aaaagccatt accctcatta ggcatggttt tgggctccca aaacacctga cagcccctcc
 9661 ctcctctgag aggcggagag tgctgactgt agtgaccatt gcatgccggg tgcagcatct
 9721 ggaagagcta ggcagggtgt ctgccccctc ctgagttgaa gtcatgctcc cctgtgccag
 9781 cccagaggcc gagagctatg gacagcattg ccagtaacac aggccaccct gtgcagaagg
 9841 gagctggctc cagcctggaa acctgtctga ggttgggaga ggtgcacttg gggcacaggg
 9901 agaggccggg acacacttag ctggagatgt ctctaaaagc cctgtatcgt attcaccttc
 9961 agtttttgtg ttttgggaca attactttag aaaataagta ggtcgtttta aaaacaaaaa
10021 ttattgattg ctttttttgta gtgttcagaa aaaaggttct ttgtgtatag ccaaatgact
10081 gaaagcactg atatatttaa aaacaaaagg caatttatta aggaaatttg taccatttca
10141 gtaaacctgt ctgaatgtac ctgtatacgt ttcaaaaaca ccccccccc actgaatccc
10201 tgtaacctat ttattatata aagagtttgc cttataaatt t
```

[0043] By "base editor (BE)," or "nucleobase editor (NBE)" is meant an agent that binds a polynucleotide and has nucleobase modifying activity. In various embodiments, the base editor comprises a nucleobase modifying polypeptide (*e.g.*, a deaminase) and a polynucleotide programmable nucleotide binding domain in conjunction with a guide polynucleotide (*e.g.*, guide RNA). In various embodiments, the agent is a biomolecular complex comprising a protein domain having base editing activity, i.e., a domain capable of modifying a base (e.g., A, T, C, G, or U) within a nucleic acid molecule (e.g., DNA). In some embodiments, the polynucleotide programmable DNA binding domain is fused or linked to a deaminase domain. In one embodiment, the agent is a fusion protein comprising a domain having base editing activity. In another embodiment, the protein domain having base editing activity is linked to the guide RNA (e.g., via an RNA binding motif on the guide RNA and an RNA binding domain fused to the deaminase). In some embodiments, the domain having base editing activity is capable of deaminating a base within a nucleic acid molecule. In some embodiments, the base editor is capable of deaminating a base within a DNA molecule. In some embodiments, the base editor is capable of deaminating a cytosine (C) or an adenosine (A) within DNA. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenosine base editor (ABE). In some embodiments, an adenosine deaminase is evolved from TadA. In some embodiments, the polynucleotide programmable DNA binding domain is a CRISPR associated (*e.g.*, Cas or Cpf1) enzyme. In some embodiments, the base editor is a catalytically dead Cas9 (dCas9) fused to a deaminase domain. In some embodiments, the base editor is a Cas9 nickase (nCas9) fused to a deaminase domain. In some embodiments, the base editor is fused to an inhibitor of base excision repair (BER). In some embodiments, the inhibitor of base excision repair is a uracil DNA glycosylase inhibitor (UGI). In some

embodiments, the inhibitor of base excision repair is an inosine base excision repair inhibitor. Details of base editors are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), and Rees, H.A., et al., "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1, the entire contents of which are hereby incorporated by reference.

By way of example, the cytidine base editor CBE as used in the base editing compositions, systems and methods described herein has the following nucleic acid sequence (8877 base pairs), (Addgene, Watertown, MA.; Komor AC, et al., 2017, Sci Adv., 30;3(8):eaao4774. doi: 10.1126/sciadv.aao4774) as provided below. Polynucleotide sequences having at least 95% or greater identity to the BE4 nucleic acid sequence are also encompassed.

```
   1 atatgccaag tacgcccccc attgacgtca atgacggtaa atggcccgcc tggcattatg
  61 cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg
 121 ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact
 181 cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa
 241 atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta
 301 ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct
 361 agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gagctcagag
 421 actggcccag tggctgtgga ccccacattg agacggcgga tcgagcccca tgagtttgag
 481 gtattcttcg atccgagaga gctccgcaag gagacctgcc tgctttacga aattaattgg
 541 gggggccggc actccatttg gcgacataca tcacagaaca ctaacaagca cgtcgaagtc
 601 aacttcatcg agaagttcac gacagaaaga tatttctgtc cgaacacaag gtgcagcatt
 661 acctggtttc tcagctggag cccatgcggc gaatgtagta gggccatcac tgaattcctg
 721 tcaaggtatc cccacgtcac tctgtttatt tacatcgcaa ggctgtacca ccacgctgac
 781 ccccgcaatc gacaaggcct gcgggatttg atctcttcag gtgtgactat ccaaattatg
 841 actgagcagg agtcaggata ctgctggaga aactttgtga attatagccc gagtaatgaa
 901 gcccactggc ctaggtatcc ccatctgtgg gtacgactgt acgttcttga actgtactgc
 961 atcatactgg gcctgcctcc ttgtctcaac attctgagaa ggaagcagcc acagctgaca
1021 ttctttacca tcgctcttca gtcttgtcat taccagcgac tgcccccaca cattctctgg
1081 gccaccgggt tgaaatctgg tggttcttct ggtggttcta gcggcagcga gactcccggg
1141 acctcagagt ccgccacacc cgaaagttct ggtggttctt ctggtggttc tgataaaaag
1201 tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa
```

1261 tacaaagtac cttcaaagaa atttaaggtg ttggggaaca cagaccgtca ttcgattaaa
1321 aagaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg
1381 aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc gaatatgtta cttacaagaa
1441 atttttagca atgagatggc caaagttgac gattctttct ttcaccgttt ggaagagtcc
1501 ttccttgtcg aagaggacaa gaaacatgaa cggcacccca tctttggaaa catagtagat
1561 gaggtggcat atcatgaaaa gtacccaacg atttatcacc tcagaaaaaa gctagttgac
1621 tcaactgata aagcggacct gaggttaatc tacttggctc ttgcccatat gataaagttc
1681 cgtgggcact ttctcattga gggtgatcta aatccggaca actcggatgt cgacaaactg
1741 ttcatccagt tagtacaaac ctataatcag ttgtttgaag agaaccctat aaatgcaagt
1801 ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgacg gctagaaaac
1861 ctgatcgcac aattacccgg agagaagaaa aatgggttgt tcggtaacct tatagcgctc
1921 tcactaggcc tgacaccaaa ttttaagtcg aacttcgact tagctgaaga tgccaaattg
1981 cagcttagta aggacacgta cgatgacgat ctcgacaatc tactggcaca aattggagat
2041 cagtatgcgg acttatttttt ggctgccaaa aaccttagcg atgcaatcct cctatctgac
2101 atactgagag ttaatactga gattaccaag gcgccgttat ccgcttcaat gatcaaaagg
2161 tacgatgaac atcaccaaga cttgacactt ctcaaggccc tagtccgtca gcaactgcct
2221 gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac
2281 ggcggagcga gtcaagagga attctacaag tttatcaaac ccatattaga gaagatggat
2341 gggacggaag agttgcttgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact
2401 ttcgacaacg gtagcattcc acatcaaatc cacttaggcg aattgcatgc tatacttaga
2461 aggcaggagg atttttatcc gttcctcaaa gacaatcgtg aaaagattga gaaaatccta
2521 acctttcgca taccttacta tgtgggaccc ctggcccgag ggaactctcg gttcgcatgg
2581 atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa
2641 ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact ttgacaagaa tttaccgaac
2701 gaaaaagtat tgcctaagca cagtttactt tacgagtatt tcacagtgta caatgaactc
2761 acgaaagtta agtatgtcac tgagggcatg cgtaaacccg cctttctaag cggagaacag
2821 aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgacagt taagcaattg
2881 aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cggggtagaa
2941 gatcgattta atgcgtcact tggtacgtat catgacctcc taaagataat taaagataag
3001 gacttcctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc
3061 ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgctca cctgttcgac
3121 gataaggtta tgaaacagtt aaagaggcgt cgctatacgg gctggggacg attgtcgcgg
3181 aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctcga ttttctaaag
3241 agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc tttaaccttc
3301 aaagaggata tacaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt
3361 gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg
3421 gatgagctag ttaaggtcat gggacgtcac aaaccggaaa acattgtaat cgagatggca
3481 cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc gagagcggat gaagagaata
3541 gaagagggta ttaaagaact gggcagccag atcttaaagg agcatcctgt ggaaaatacc
3601 caattgcaga acgagaaact ttacctctat tacctacaaa atggaaggga catgtatgtt
3661 gatcaggaac tggacataaa ccgtttatct gattacgacg tcgatcacat tgtaccccaa
3721 tccttttga aggacgattc aatcgacaat aaagtgctta cacgctcgga taagaaccga

3781 gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga aaatgaagaa ctattggcgg
3841 cagctcctaa atgcgaaact gataacgcaa agaaagttcg ataacttaac taaagctgag
3901 aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaaacc
3961 cgccaaatca caaagcatgt tgcacagata ctagattccc gaatgaatac gaaatacgac
4021 gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg
4081 gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgcg
4141 cacgacgctt atcttaatgc cgtcgtaggg accgcactca ttaagaaata cccgaagcta
4201 gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa
4261 agcgaacagg agataggcaa ggctacagcc aaatacttct tttattctaa cattatgaat
4321 ttctttaaga cggaaatcac tctggcaaac ggagagatac gcaaacgacc tttaattgaa
4381 accaatgggg agacaggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga
4441 aaagttttgt ccatgcccca agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg
4501 ttttcaaagg aatcgattct tccaaaaagg aatagtgata agctcatcgc tcgtaaaaag
4561 gactgggacc cgaaaaagta cggtggcttc gatagcccta cagttgccta ttctgtccta
4621 gtagtggcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg
4681 gggataacga ttatggagcg ctcgtctttt gaaaagaacc ccatcgactt ccttgaggcg
4741 aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt
4801 gagttagaaa atggccgaaa acggatgttg gctagcgccg gagagcttca aaaggggaac
4861 gaactcgcac taccgtctaa atacgtgaat ttcctgtatt tagcgtccca ttacgagaag
4921 ttgaaaggtt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat
4981 tatctcgacg aaatcataga gcaaatttcg gaattcagta agagagtcat cctagctgat
5041 gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag
5101 caggcggaaa atattatcca tttgtttact cttaccaacc tcggcgctcc agccgcattc
5161 aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac
5221 gcgacactga ttcaccaatc catcacggga ttatatgaaa ctcggataga tttgtcacag
5281 cttgggggtg actctggtgg ttctggagga tctggtggtt ctactaatct gtcagatatt
5341 attgaaaagg agaccggtaa gcaactggtt atccaggaat ccatcctcat gctcccagag
5401 gaggtggaag aagtcattgg gaacaagccg gaaagcgata tactcgtgca caccgcctac
5461 gacgagagca ccgacgagaa tgtcatgctt ctgactagcg acgcccctga atacaagcct
5521 tgggctctgg tcatacagga tagcaacggt gagaacaaga ttaagatgct ctctggtggt
5581 tctggaggat ctggtggttc tactaatctg tcagatatta ttgaaaagga gaccggtaag
5641 caactggtta tccaggaatc catcctcatg ctcccagagg aggtggaaga agtcattggg
5701 aacaagccgg aaagcgatat actcgtgcac accgcctacg acgagagcac cgacgagaat
5761 gtcatgcttc tgactagcga cgcccctgaa tacaagcctt gggctctggt catacaggat
5821 agcaacggtg agaacaagat taagatgctc tctggtggtt ctcccaagaa gaagaggaaa
5881 gtctaaccgg tcatcatcac catcaccatt gagtttaaac ccgctgatca gcctcgactg
5941 tgccttctag ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg
6001 aaggtgccac tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga
6061 gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg
6121 aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag gcggaaagaa
6181 ccagctgggg ctcgataccg tcgacctcta gctagagctt ggcgtaatca tggtcatagc
6241 tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca

6301 taaagtgtaa agcctagggt gcctaatgag tgagctaact cacattaatt gcgttgcgct
6361 cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac
6421 gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc
6481 tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt
6541 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg
6601 ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg
6661 agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat
6721 accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta
6781 ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct
6841 gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc
6901 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa
6961 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg
7021 taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaagaacag
7081 tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt
7141 gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta
7201 cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc
7261 agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca
7321 cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa
7381 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat
7441 ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct
7501 taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt
7561 tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat
7621 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta
7681 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg
7741 gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt
7801 tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg
7861 cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg
7921 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc
7981 ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa
8041 ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac
8101 cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt
8161 ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg
8221 gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa
8281 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
8341 aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc gacggatcgg
8401 gagatcgatc tcccgatccc ctagggtcga ctctcagtac aatctgctct gatgccgcat
8461 agttaagcca gtatctgctc cctgcttgtg tgttggaggt cgctgagtag tgcgcgagca
8521 aaatttaagc tacaacaagg caaggcttga ccgacaattg catgaagaat ctgcttaggg
8581 ttaggcgttt tgcgctgctt cgcgatgtac gggccagata tacgcgttga cattgattat
8641 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt
8701 tccgcgttac ataacttacg gtaaatggcc cgcctggctg accgcccaac gacccccgcc
8761 cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac

8821 gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatc

BE4 amino acid sequence:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY
YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPS
EEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEV
EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS
GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVM
LLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK

[0044] By way of example, the adenine base editor ABE as used in the base editing compositions, systems and methods described herein has the following nucleic acid sequence (8877 base pairs), (Addgene, Watertown, MA.; Gaudelli NM, et al., Nature. 2017 Nov 23;551(7681):464-471. doi: 10.1038/nature24644; Koblan LW, et al., Nat Biotechnol. 2018 Oct;36(9):843-846. doi: 10.1038/nbt.4172.) as provided below. Polynucleotide sequences having at least 95% or greater identity to the ABE nucleic acid sequence are also encompassed.

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT
GACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCC
ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGT
CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACA
GCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGT
ATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGT
GCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCA
CACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCC
TGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGT
GTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATG
AACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTA
GAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGG
AGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCC
GGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGG
CACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTG
GAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTG
GTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCG
GCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGG
CTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCA
GATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGG
CCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGA
GAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCC
ATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGA
AACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGT
CCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGC
CTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGAC
CTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACC
TGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGA
GGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCC

TGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAG
CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTT
CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA
AGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGC
TCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGG
ACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAA
CGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTAC
CCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCC
CTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAA
CTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGC
TGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACAT
ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA
AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCC
GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGG
CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA
GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTA
AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAG
AGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTT
CGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCG
GAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAAC
GCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACA
AGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGC
GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAA
AGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
TACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
CCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAA

ACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGG
CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATC
GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCT
ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAA
TCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
AGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
GAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCAC
TGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTA
ATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
AGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTA
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACTC
ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGA
TCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG

TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGA
TCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCAT
TAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCC
CAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCAT
TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0045] By "base editing activity" is meant acting to chemically alter a base within a polynucleotide. In one embodiment, a first base is converted to a second base. In one embodiment, the base editing activity is cytidine deaminase activity, e.g., converting target C•G to T•A. In another embodiment, the base editing activity is adenosine or adenine deaminase activity, e.g., converting A•T to G•C.

[0046] The term "base editor system" refers to a system for editing a nucleobase of a target nucleotide sequence. In various embodiments, the base editor (BE) system comprises (1) a polynucleotide programmable nucleotide binding domain and a deaminase domain for deaminating said nucleobase; and (2) a guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable DNA binding domain and a deaminase domain for deaminating said nucleobase; and (2) a guide RNA in conjunction with the polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE).

[0047] In some embodiments, a nucleobase editor system may comprise more than one base editing component. For example, a nucleobase editor system may include more than one deaminase. In some embodiments, a nuclease base editor system may include one or more cytidine deaminase and/or one or more adenosine deaminases. In some embodiments, a single guide polynucleotide may be utilized to target different deaminases to a target nucleic acid sequence. In some embodiments, a single pair of guide polynucleotides may be utilized to target different deaminases to a target nucleic acid sequence.

[0048] The nucleobase component and the polynucleotide programmable nucleotide binding component of a base editor system may be associated with each other covalently or non-covalently. For example, in some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a polynucleotide programmable nucleotide binding domain. In some

embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can target a deaminase domain to a target nucleotide sequence by non-covalently interacting with or associating with the deaminase domain. For example, in some embodiments, the nucleobase editing component, e.g. the deaminase component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a steril alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

[0049] A base editor system may further comprise a guide polynucleotide component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. In some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a guide polynucleotide. For example, in some embodiments, the nucleobase editing component of the base editor system, e.g. the deaminase component, can comprise an additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (e.g., a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the deaminase domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of

binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

[0050] In some embodiments, a base editor system can further comprise an inhibitor of base excision repair (BER) component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. The inhibitor of BER component may comprise a base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be a uracil DNA glycosylase inhibitor (UGI). In some embodiments, the inhibitor of base excision repair can be an inosine base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the polynucleotide programmable nucleotide binding domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain and an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can target an inhibitor of base excision repair to a target nucleotide sequence by non-covalently interacting with or associating with the inhibitor of base excision repair. For example, in some embodiments, the inhibitor of base excision repair component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the guide polynucleotide. For example, in some embodiments, the inhibitor of base excision repair can comprise an additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (e.g., a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the

guide polynucleotide (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the inhibitor of base excision repair. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

[0051] The term "Cas9" or "Cas9 domain" refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease. An exemplary Cas9, is *Streptococcus pyogenes* Cas9, the amino acid sequence of which is provided below:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIY
NQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS

EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain)

[0052] The term "conservative amino acid substitution" or "conservative mutation" refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)). According to such analyses, groups of amino acids can be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and Schirmer, R. H., supra). Non-limiting examples of conservative mutations include amino acid substitutions of amino acids, for example, lysine for arginine and vice versa such that a positive charge can be maintained; glutamic acid for aspartic acid and vice versa such that a negative charge can be maintained; serine for threonine such that a free –OH can be maintained; and glutamine for asparagine such that a free $–NH_2$ can be maintained.

[0053] The term "coding sequence" or "protein coding sequence" as used interchangeably herein refers to a segment of a polynucleotide that codes for a protein. The region or sequence is bounded nearer the 5' end by a start codon and nearer the 3' end with a stop codon. Coding sequences can also be referred to as open reading frames.

[0054] The term "deaminase" or "deaminase domain," as used herein, refers to a protein or enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase or deaminase domain is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine, respectively. In some embodiments, the deaminase or deaminase domain is a cytosine deaminase, catalyzing the hydrolytic deamination of cytosine to uracil. In some embodiments, the deaminase is an adenosine deaminase, which catalyzes the hydrolytic deamination of adenine to hypoxanthine. In some embodiments, the deaminase is an adenosine deaminase, which catalyzes the hydrolytic deamination of adenosine or adenine (A) to inosine (I). In some embodiments, the deaminase or deaminase domain is an adenosine deaminase, catalyzing the hydrolytic deamination of adenosine or deoxyadenosine to inosine or

deoxyinosine, respectively. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenosine in deoxyribonucleic acid (DNA). The adenosine deaminases (*e.g.* engineered adenosine deaminases, evolved adenosine deaminases) provided herein can be from any organism, such as a bacterium. In some embodiments, the adenosine deaminase is from a bacterium, such as *E. coli*, *S. aureus*, *S. typhi*, *S. putrefaciens*, *H. influenzae*, or *C. crescentus*. In some embodiments, the adenosine deaminase is a TadA deaminase. In some embodiments, the deaminase or deaminase domain is a variant of a naturally occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% identical to a naturally occurring deaminase. For example, deaminase domains are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)), and Rees, H.A., et al., "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1, the entire contents of which are hereby incorporated by reference.

[0055] By "detectable label" is meant a composition that when linked to a molecule of interest renders the latter detectable, via spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes, magnetic beads, metallic beads, colloidal particles, fluorescent dyes, electron-dense reagents, enzymes (for example, as commonly used in an ELISA), biotin, digoxigenin, or haptens.

[0056] By "disease" is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. An example of a disease includes Rett Syndrome.

[0057] By "effective amount" is meant the amount of a required to ameliorate the symptoms of a disease relative to an untreated patient. The effective amount of active compound(s) used to

practice the present invention for therapeutic treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an "effective" amount. In one embodiment, an effective amount is the amount of a base editor of the invention sufficient to introduce an alteration in a gene of interest (e.g., Mecp2) in a cell (e.g., a cell in vitro or in vivo). In one embodiment, an effective amount is the amount of a base editor required to achieve a therapeutic effect (e.g., to reduce or control Rett Syndrome or a symptom or condition thereof). Such therapeutic effect need not be sufficient to alter Mecp2 in all cells of a subject, tissue or organ, but only to alter Mecp2 in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an effective amount is sufficient to ameliorate one or more symptoms of Rett Syndrome.

[0058] By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, preferably, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

[0059] "Hybridization" means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleobases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds.

[0060] The term "inhibitor of base repair" or "IBR" refers to a protein that is capable in inhibiting the activity of a nucleic acid repair enzyme, for example a base excision repair enzyme. In some embodiments, the IBR is an inhibitor of inosine base excision repair. Exemplary inhibitors of base repair include inhibitors of APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGGl, hNEILl, T7 Endol, T4PDG, UDG, hSMUGl, and hAAG. In some embodiments, the IBR is an inhibitor of Endo V or hAAG. In some embodiments, the IBR is a catalytically inactive EndoV or a catalytically inactive hAAG. In some embodiments, the base repair inhibitor is an inhibitor of Endo V or hAAG. In some embodiments, the base repair inhibitor is a catalytically inactive EndoV or a catalytically inactive hAAG. In some embodiments, the base repair inhibitor is uracil glycosylase inhibitor (UGI). UGI refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a fragment of a wild-type UGI. In some embodiments, the UGI proteins provided herein include fragments of UGI and

proteins homologous to a UGI or a UGI fragment. In some embodiments, the base repair inhibitor is an inhibitor of inosine base excision repair. In some embodiments, the base repair inhibitor is a "catalytically inactive inosine specific nuclease" or "dead inosine specific nuclease. Without wishing to be bound by any particular theory, catalytically inactive inosine glycosylases (*e.g.*, alkyl adenine glycosylase (AAG)) can bind inosine, but cannot create an abasic site or remove the inosine, thereby sterically blocking the newly formed inosine moiety from DNA damage/repair mechanisms. In some embodiments, the catalytically inactive inosine specific nuclease can be capable of binding an inosine in a nucleic acid but does not cleave the nucleic acid. Non-limiting exemplary catalytically inactive inosine specific nucleases include catalytically inactive alkyl adenosine glycosylase (AAG nuclease), for example, from a human, and catalytically inactive endonuclease V (EndoV nuclease), for example, from *E. coli*. In some embodiments, the catalytically inactive AAG nuclease comprises an E125Q mutation or a corresponding mutation in another AAG nuclease.

[0061] The terms "isolated," "purified," or "biologically pure" refer to material that is free to varying degrees from components which normally accompany it as found in its native state. "Isolate" denotes a degree of separation from original source or surroundings. "Purify" denotes a degree of separation that is higher than isolation. A "purified" or "biologically pure" protein is sufficiently free of other materials such that any impurities do not materially affect the biological properties of the protein or cause other adverse consequences. That is, a nucleic acid or peptide of this invention is purified if it is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Purity and homogeneity are typically determined using analytical chemistry techniques, for example, polyacrylamide gel electrophoresis or high-performance liquid chromatography. The term "purified" can denote that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. For a protein that can be subjected to modifications, for example, phosphorylation or glycosylation, different modifications may give rise to different isolated proteins, which can be separately purified.

[0062] By "isolated polynucleotide" is meant a nucleic acid (e.g., a DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term

includes an RNA molecule that is transcribed from a DNA molecule, as well as a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

[0063] By an "isolated polypeptide" is meant a polypeptide of the invention that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide of the invention. An isolated polypeptide of the invention may be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding such a polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

[0064] The term "linker", as used herein, can refer to a covalent linker (*e.g.*, covalent bond), a non-covalent linker, a chemical group, or a molecule linking two molecules or moieties, *e.g.*, two components of a protein complex or a ribonucleocomplex, or two domains of a fusion protein, such as, for example, a polynucleotide programmable DNA binding domain (*e.g.*, dCas9) and a deaminase domain (*e.g.*, an adenosine deaminase or a cytidine deaminase). A linker can join different components of, or different portions of components of, a base editor system. For example, in some embodiments, a linker can join a guide polynucleotide binding domain of a polynucleotide programmable nucleotide binding domain and a catalytic domain of a deaminase. In some embodiments, a linker can join a CRISPR polypeptide and a deaminase. In some embodiments, a linker can join a Cas9 and a deaminase. In some embodiments, a linker can join a dCas9 and a deaminase. In some embodiments, a linker can join a nCas9 and a deaminase. In some embodiments, a linker can join a guide polynucleotide and a deaminase. In some embodiments, a linker can join a deaminating component and a polynucleotide programmable nucleotide binding component of a base editor system. In some embodiments, a linker can join a RNA-binding portion of a deaminating component and a polynucleotide programmable nucleotide binding component of a base editor system. In some embodiments, a linker can join a RNA-binding portion of a deaminating component and a RNA-binding portion of a polynucleotide programmable nucleotide binding component of a base editor system. A linker can be positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond or non-covalent interaction, thus connecting the two. In some embodiments, the linker can be an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker can be a polynucleotide. In some embodiments, the linker can be a DNA linker. In some embodiments, the linker can be a RNA linker. In some

embodiments, a linker can comprise an aptamer capable of binding to a ligand. In some embodiments, the ligand may be carbohydrate, a peptide, a protein, or a nucleic acid. In some embodiments, the linker may comprise an aptamer may be derived from a riboswitch. The riboswitch from which the aptamer is derived may be selected from a theophylline riboswitch, a thiamine pyrophosphate (TPP) riboswitch, an adenosine cobalamin (AdoCbl) riboswitch, an S-adenosyl methionine (SAM) riboswitch, an SAH riboswitch, a flavin mononucleotide (FMN) riboswitch, a tetrahydrofolate riboswitch, a lysine riboswitch, a glycine riboswitch, a purine riboswitch, a GlmS riboswitch, or a pre-queosine1 (PreQ1) riboswitch. In some embodiments, a linker may comprise an aptamer bound to a polypeptide or a protein domain, such as a polypeptide ligand. In some embodiments, the polypeptide ligand may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif. In some embodiments, the polypeptide ligand may be a portion of a base editor system component. For example, a nucleobase editing component may comprise a deaminase domain and a RNA recognition motif.

[0065] In some embodiments, the linker can be an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker can be about 5-100 amino acids in length, for example, about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 amino acids in length. In some embodiments, the linker can be about 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, or 450-500 amino acids in length. Longer or shorter linkers can be also contemplated.

[0066] In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic-acid editing protein (e.g., cytidine or adenosine deaminase). In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. For example, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-200 amino acids in length, for example, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 35, 45, 50, 55, 60, 60, 65, 70, 70, 75, 80, 85, 90, 90, 95, 100, 101, 102, 103, 104, 105, 110, 120, 130, 140, 150, 160, 175, 180, 190, or 200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES, which may also be referred to as the XTEN linker. In some

embodiments, a linker comprises the amino acid sequence SGGS. In some embodiments, a linker comprises $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, $(G)_n$, $(EAAAK)_n$, $(GGS)_n$, SGSETPGTSESATPES, or $(XP)_n$ motif, or a combination of any of these, where n is independently an integer between 1 and 30, and where X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, a linker comprises a plurality of proline residues and is 5-21, 5-14, 5-9, 5-7 amino acids in length, e.g., PAPAP, PAPAPA, PAPAPAP, PAPAPAPA, $P(AP)_4$, $P(AP)_7$, $P(AP)_{10}$. Such proline-rich linkers are also termed "rigid" linkers.

[0067] In some embodiments, the domains of a base editor are fused via a linker that comprises In some embodiments, the domains of a base editor are fused via a linker that comprises the amino acid sequence of SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, or GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS. In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES, which may also be referred to as the XTEN linker. In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES. In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS. In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS. In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS.

[0068] The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)). In some embodiments, the

presently disclosed base editors can efficiently generate an "intended mutation", such as a point mutation, in a nucleic acid (*e.g.*, a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (*e.g.*, cytidine base editor or adenosine base editor) bound to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation. In general, mutations made or identified in a sequence (e.g., an amino acid sequence as described herein) are numbered in relation to a reference (or wild type) sequence, i.e., a sequence that does not contain the mutations. The skilled practitioner in the art would readily understand how to determine the position of mutations in amino acid and nucleic acid sequences relative to a reference sequence.

[0069] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, e.g., a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, e.g., nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (e.g., a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.,* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, e.g., in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine,

guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (e.g., methylated bases); intercalated bases; modified sugars (2′-e.g.,fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5′-*N*-phosphoramidite linkages).

[0070] The term "nuclear localization sequence," "nuclear localization signal," or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus. Nuclear localization sequences are known in the art and described, for example, in Plank *et al.*, International PCT application, PCT/EP2000/011690, filed November 23, 2000, published as WO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In other embodiments, the NLS is an optimized NLS described, for example, by Koblan et al., Nature Biotech. 2018 doi:10.1038/nbt.4172. In some embodiments, an NLS comprises the amino acid sequence KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK, PKKKRKV, or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC.

[0071] The term "nucleobase", "nitrogenous base", or "base", used interchangeably herein, refers to a nitrogen-containing biological compound that forms a nucleoside, which in turn is a component of a nucleotide. The ability of nucleobases to form base pairs and to stack one upon another leads directly to long-chain helical structures such as ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). Five nucleobases – adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U) – are called primary or canonical. Adenine and guanine are derived from purine, and cytosine, uracil, and thymine are derived from pyrimidine. DNA and RNA can also contain other (non-primary) bases that are modified. Non-limiting exemplary modified nucleobases can include hypoxanthine, xanthine, 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine (m5C), and 5-hydromethylcytosine. Hypoxanthine and xanthine can be created through mutagen presence, both of them through deamination (replacement of the amine group with a carbonyl group). Hypoxanthine can be modified from adenine. Xanthine can be modified from guanine. Uracil can result from deamination of cytosine. A "nucleoside" consists of a nucleobase and a five-carbon sugar (either ribose or deoxyribose). Examples of a

nucleoside include adenosine, guanosine, uridine, cytidine, 5-methyluridine (m5U), deoxyadenosine, deoxyguanosine, thymidine, deoxyuridine, and deoxycytidine. Examples of a nucleoside with a modified nucleobase includes inosine (I), xanthosine (X), 7-methylguanosine (m7G), dihydrouridine (D), 5-methylcytidine (m5C), and pseudouridine (Ψ). A "nucleotide" consists of a nucleobase, a five-carbon sugar (either ribose or deoxyribose), and at least one phosphate group.

[0072] The term "nucleic acid programmable DNA binding protein" or "napDNAbp" may be used interchangeably with "polynucleotide programmable nucleotide binding domain" to refer to a protein that associates with a nucleic acid (e.g., DNA or RNA), such as a guide nucleic acid, that guides the napDNAbp to a specific nucleic acid sequence. For example, a Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Examples of nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, and Cas12i. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, although they may not be specifically listed in this disclosure. See, e.g., Makarova et al. "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" *CRISPR J.* 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan et al., "Functionally diverse type V CRISPR-Cas systems" *Science*. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271, the entire contents of each are hereby incorporated by reference.

[0073] The terms "nucleobase editing domain" or "nucleobase editing protein", as used herein, refers to a protein or enzyme that can catalyze a nucleobase modification in RNA or DNA, such as cytosine (or cytidine) to uracil (or uridine) or thymine (or thymidine), and adenine (or adenosine) to hypoxanthine (or inosine) deaminations, as well as non-templated nucleotide additions and insertions. In some embodiments, the nucleobase editing domain is a deaminase domain (*e.g.*, a cytidine deaminase, a cytosine deaminase, an adenine deaminase, or an adenosine deaminase). In some embodiments, the nucleobase editing domain can be a naturally occurring nucleobase editing domain. In some embodiments, the nucleobase editing domain can be an engineered or evolved nucleobase editing domain from the naturally occurring nucleobase editing domain. The nucleobase editing domain can be from any organism, such as a bacterium, human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. For example, nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO

2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see, Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

[0074] As used herein, "obtaining" as in "obtaining an agent" includes synthesizing, purchasing, or otherwise acquiring the agent.

[0075] "Patient" or "subject" as used herein refers to a mammalian subject diagnosed with or suspected of having or developing a disease or a disorder. In some embodiments, the term "patient" refers to a mammalian subject with a higher than average likelihood of developing a disease or a disorder. Exemplary patients can be humans, non-human primates, cats, dogs, pigs, cattle, cats, horses, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other mammalians that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female.

[0076] "Patient in need thereof" or "subject in need thereof" is referred to herein as a patient diagnosed with or suspected of having a disease or disorder, for instance, but not restricted to Rett Syndrome (RTT).

[0077] The terms "pathogenic mutation", "pathogenic variant", "disease casing mutation", "disease causing variant", "deleterious mutation", or "predisposing mutation" refers to a genetic alteration or mutation that increases an individual's susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation comprises at least one wild-type amino acid substituted by at least one pathogenic amino acid in a protein encoded by a gene.

[0078] The term "non-conservative mutations" involve amino acid substitutions between different groups, for example, lysine for tryptophan, or phenylalanine for serine, etc. In this case, it is preferable for the non-conservative amino acid substitution to not interfere with, or inhibit the biological activity of, the functional variant. The non-conservative amino acid substitution can enhance the biological activity of the functional variant, such that the biological activity of the functional variant is increased as compared to the wild-type protein.

[0079] The terms "protein", "peptide", "polypeptide", and their grammatical equivalents are used interchangeably herein, and refer to a polymer of amino acid residues linked together by

peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide can refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide can be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modifications, etc. A protein, peptide, or polypeptide can also be a single molecule or can be a multi-molecular complex. A protein, peptide, or polypeptide can be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide can be naturally occurring, recombinant, or synthetic, or any combination thereof. The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein can be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an amino-terminal fusion protein or a carboxy-terminal fusion protein, respectively. A protein can comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain, or a catalytic domain of a nucleic acid editing protein. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.*, RNA or DNA. Any of the proteins provided herein can be produced by any method known in the art. For example, the proteins provided herein can be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[0080] Polypeptides and proteins disclosed herein (including functional portions and functional variants thereof) can comprise synthetic amino acids in place of one or more naturally-occurring amino acids. Such synthetic amino acids are known in the art, and include, for example, aminocyclohexane carboxylic acid, norleucine, α-amino n-decanoic acid, homoserine, S-acetylaminomethyl-cysteine, trans-3- and trans-4-hydroxyproline, 4-aminophenylalanine, 4-nitrophenylalanine, 4-chlorophenylalanine, 4-carboxyphenylalanine, β-

phenylserine β-hydroxyphenylalanine, phenylglycine, α-naphthylalanine, cyclohexylalanine, cyclohexylglycine, indoline-2-carboxylic acid, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid, aminomalonic acid, aminomalonic acid monoamide, N'-benzyl-N'-methyl-lysine, N',N'-dibenzyl-lysine, 6-hydroxylysine, ornithine, α-aminocyclopentane carboxylic acid, α-aminocyclohexane carboxylic acid, α-aminocycloheptane carboxylic acid, α-(2-amino-2-norbornane)-carboxylic acid, α,γ-diaminobutyric acid, α,β-diaminopropionic acid, homophenylalanine, and α-tert-butylglycine. The polypeptides and proteins can be associated with post-translational modifications of one or more amino acids of the polypeptide constructs. Non-limiting examples of post-translational modifications include phosphorylation, acylation including acetylation and formylation, glycosylation (including N-linked and O-linked), amidation, hydroxylation, alkylation including methylation and ethylation, ubiquitylation, addition of pyrrolidone carboxylic acid, formation of disulfide bridges, sulfation, myristoylation, palmitoylation, isoprenylation, farnesylation, geranylation, glypiation, lipoylation and iodination.

[0081] The term "polynucleotide programmable nucleotide binding domain" refers to a protein that associates with a nucleic acid (*e.g.*, DNA or RNA), such as a guide polynucleotide (*e.g.*, guide RNA), that guides the polynucleotide programmable DNA binding domain to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 protein. A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that has complementary to the guide RNA. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, and Cas12i. Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4,

Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, homologues thereof, or modified or engineered versions thereof. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, though they are not specifically listed in this disclosure.

[0082] The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

[0083] By “reduces” is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

[0084] By “reference” is meant a standard or control condition. In one embodiment, the reference is a wild-type or healthy cell.

[0085] A “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, preferably at least about 20 amino acids, more preferably at least about 25 amino acids, and even more preferably about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least about 50 nucleotides, preferably at least about 60 nucleotides, more preferably at least about 75 nucleotides, and even more preferably about 100 nucleotides or about 300 nucleotides or any integer thereabout or therebetween.

[0086] The term "RNA-programmable nuclease," and "RNA-guided nuclease" are used with (e.g., binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is used interchangeably to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (e.g., and directs binding of a Cas9 complex to the target); and (2) a domain that

binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA, and comprises a stem-loop structure. For example, in some embodiments, domain (2) is identical or homologous to a tracrRNA as provided in Jinek et al., Science 337:816-821(2012), the entire contents of which is incorporated herein by reference. Other examples of gRNAs (e.g., those including domain 2) can be found in U.S. Provisional Patent Application, U.S.S.N. 61/874,682, filed September 6, 2013, entitled "Switchable Cas9 Nucleases and Uses Thereof," and U.S. Provisional Patent Application, U.S.S.N. 61/874,746, filed September 6, 2013, entitled "Delivery System For Functional Nucleases," the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an "extended gRNA." For example, an extended gRNA will, e.g., bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csnl) from Streptococcus pyogenes (see, e.g., "Complete genome sequence of an Ml strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C, Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011).

[0087] The term “single nucleotide polymorphism (SNP)” is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.* > 1%). For example, at a specific base position in the human genome, the C nucleotide can appear in most individuals, but in a minority of individuals, the position is occupied by an A. This means that there is a SNP at this specific position, and the two possible nucleotide variations, C or A, are the to be alleles for this position. SNPs underlie differences in susceptibility to disease; a wide range of human diseases. The severity of illness and the way our body responds to treatments are also manifestations of genetic variations. SNPs can fall within coding regions of genes, non-coding regions of genes, or in the intergenic regions (regions between genes). In some embodiments, SNPs within a coding sequence do not necessarily change the amino acid sequence of the protein that is

produced, due to degeneracy of the genetic code. SNPs in the coding region are of two types: synonymous and nonsynonymous SNPs. Synonymous SNPs do not affect the protein sequence, while nonsynonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types: missense and nonsense. SNPs that are not in protein-coding regions can still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA. Gene expression affected by this type of SNP is referred to as an eSNP (expression SNP) and can be upstream or downstream from the gene. A single nucleotide variant (SNV) is a variation in a single nucleotide without any limitations of frequency and can arise in somatic cells. A somatic single nucleotide variation (*e.g.*, caused by cancer) can also be called a single-nucleotide alteration.

[0088] By "specifically binds" is meant a nucleic acid molecule, polypeptide, or complex thereof (e.g., a nucleic acid programmable DNA binding domain and guide nucleic acid), compound, or molecule that recognizes and binds a polypeptide and/or nucleic acid molecule of the invention, but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample.

[0089] Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having “substantial identity” to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having “substantial identity” to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By “hybridize” is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (e.g., a gene described herein), or portions thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507).

[0090] For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g.,

formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred: embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

[0091] For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42 C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

[0092] By "subject" is meant a mammal, including, but not limited to, a human or non-human mammal, such as a bovine, equine, canine, ovine, or feline.

[0093] By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). Preferably, such a sequence is at least 60%, more preferably 80% or 85%, and more preferably 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison.

[0094] Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, COBALT, EMBOSS Needle, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence. COBALT is used, for example, with the following parameters:

a) alignment parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1,
b) CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and
c) Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

EMBOSS Needle is used, for example, with the following parameters:

a) Matrix: BLOSUM62;
b) GAP OPEN: 10;
c) GAP EXTEND: 0.5;
d) OUTPUT FORMAT: pair;
e) END GAP PENALTY: false;
f) END GAP OPEN: 10; and
g) END GAP EXTEND: 0.5.

[0095] The term "target site" refers to a sequence within a nucleic acid molecule that is modified by a nucleobase editor. In one embodiment, the target site is deaminated by a

deaminase or a fusion protein comprising a deaminase (e.g., a cytidine or an adenine deaminase).

[0096] Because RNA-programmable nucleases (e.g., Cas9) use RNA:DNA hybridization to target DNA cleavage sites, these proteins are able to be targeted, in principle, to any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (e.g., to modify a genome) are known in the art (see e.g., Cong, L. et ah, Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et ah, RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et ah, Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et ah, RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et ah, Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. et ah RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[0097] As used herein, the terms "treat," treating," "treatment," and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith or obtaining a desired pharmacologic and/or physiologic effect. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated. In some embodiments, the effect is therapeutic, *i.e.*, without limitation, the effect partially or completely reduces, diminishes, abrogates, abates, alleviates, decreases the intensity of, or cures a disease and/or adverse symptom attributable to the disease. In some embodiments, the effect is preventative, *i.e.,* the effect protects or prevents an occurrence or reoccurrence of a disease or condition. To this end, the presently disclosed methods comprise administering a therapeutically effective amount of a compositions as described herein.

[0098] By "uracil glycosylase inhibitor" is meant an agent that inhibits the uracil-excision repair system. In one embodiment, the agent is a protein or fragment thereof that binds a host uracil-DNA glycosylase and prevents removal of uracil residues from DNA.

[0099] The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable or aspect herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

[0100] Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

[0101] DNA editing has emerged as a viable means to modify disease states by correcting pathogenic mutations at the genetic level. Until recently, all DNA editing platforms have functioned by inducing a DNA double strand break (DSB) at a specified genomic site and relying on endogenous DNA repair pathways to determine the product outcome in a semi-stochastic manner, resulting in complex populations of genetic products. Though precise, user-defined repair outcomes can be achieved through the homology directed repair (HDR) pathway, a number of challenges have prevented high efficiency repair using HDR in therapeutically-relevant cell types. In practice, this pathway is inefficient relative to the competing, error-prone non-homologous end joining pathway. Further, HDR is tightly restricted to the G1 and S phases of the cell cycle, preventing precise repair of DSBs in post-mitotic cells. As a result, it has proven difficult or impossible to alter genomic sequences in a user-defined, programmable manner with high efficiencies in these populations.

NUCLEOBASE EDITOR

[0102] Disclosed herein is a base editor or a nucleobase editor for editing, modifying or altering a target nucleotide sequence of a polynucleotide. Described herein is a nucleobase editor or a base editor comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain. A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target polynucleotide sequence (*i.e.*, via complementary base pairing between bases of the bound guide nucleic acid and bases of the target polynucleotide sequence) and thereby localize the base editor to the target nucleic acid sequence desired to be edited. In some embodiments, the target polynucleotide sequence comprises single-stranded DNA or double-stranded DNA. In some embodiments, the target polynucleotide sequence comprises RNA. In some embodiments, the target polynucleotide sequence comprises a DNA-RNA hybrid.

Polynucleotide Programmable Nucleotide Binding Domain

[0103] The term "polynucleotide programmable nucleotide binding domain" or "nucleic acid programmable DNA binding protein (napDNAbp)" refers to a protein that associates with a nucleic acid (*e.g.*, DNA or RNA), such as a guide polynucleotide (*e.g.*, guide RNA), that guides the polynucleotide programmable nucleotide binding domain to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding

domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 protein. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cpf1 protein.

[0104] CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, and then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self.

Cas9 domains of Nucleobase Editors

[0105] Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an Ml strain of Streptococcus pyogenes." Ferretti et al., Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences can be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9

families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference.

[0106] In some aspects, a nucleic acid programmable DNA binding protein (napDNAbp) is a Cas9 domain. Non-limiting, exemplary Cas9 domains are provided herein. The Cas9 domain may be a nuclease active Cas9 domain, a nuclease inactive Cas9 domain, or a Cas9 nickase. In some embodiments, the Cas9 domain is a nuclease active domain. For example, the Cas9 domain may be a Cas9 domain that cuts both strands of a duplexed nucleic acid (e.g., both strands of a duplexed DNA molecule). In some embodiments, the Cas9 domain comprises any one of the amino acid sequences as set forth herein. In some embodiments the Cas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more mutations compared to any one of the amino acid sequences set forth herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth herein.

[0107] In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an "nCas9" protein (for "nickase" Cas9). A nuclease-inactivated Cas9 protein can interchangeably be referred to as a "dCas9" protein (for nuclease-dead Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek et al, Science. 337:816-821(2012); Qi et al, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28; 152(5): 1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvCl subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek et al, Science. 337:816-821(2012); Qi

et al, Cell. 28;152(5): 1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to wild type Cas9. In some embodiments, the Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9.

[0108] In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, nucleotide and amino acid sequences as follows):

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCA
CTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTAT
CAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTC
AAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTT
TTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGT
GGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTAT
CATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG

ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGA
GGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTG
CACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAA
TGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATT
TATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGC
TATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCA
ATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAAC
AACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT
GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG
ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGA
AGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATT
CCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTG
AAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATT
TATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGT
TTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAA
TGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGAT
AGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGC
TAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATAT
TGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCT
CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT
GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTT
AAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTA
ACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACT
GGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACA
ACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAAT
TAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTA
TCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTA
AGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATA
AGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGT
CAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTT

GATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAAC
GCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAA
TACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATC
ATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACT
TGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCT
GAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCA
AAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGA
AACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATG
CCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTT
TACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGG
TGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAA
TCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTG
AAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCAT
TAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCC
GGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAG
CTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGA
GCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATT
TTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATAC
GTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCC
ACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAG
GTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIY
NQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF

KEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain)

[0109] In some embodiments, wild type Cas9 corresponds to, or comprises the following nucleotide and/or amino acid sequences:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAA
CCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGAT
TAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTG
AAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTT
TTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGT
CGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATAT
CATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG
ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGA
GGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTAT
AATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCG
CCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAA
TGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTC
GACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATC
TACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGC
AATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCA
ATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGC
AACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATAT
TGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGAT
GGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCG
ACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGA
GGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATA

CCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCG
AAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTT
CATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGT
TTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCA
TGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGAT
TCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCC
TAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATAT
AGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCT
CACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCT
AAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTC
AAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGA
ATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCT
AGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAA
ACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAG
AACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACT
TTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGT
TTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACA
ATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGT
CGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAG
TTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTA
AACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAAT
GAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACC
ACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAA
GCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCT
TTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGG
GGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCC
ATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGA
TTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTA
CGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGA
AAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTT
TTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCAT

AATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGC
GCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATT
TAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGT
TGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTC
ATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCA
TACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGAC
GCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGA
TTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain).

[0110] In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_002737.2 (nucleotide sequence as follows); and Uniprot Reference Sequence: Q99ZW2 (amino acid sequence as follows):

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCA
CTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTAT
CAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTC
AAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTT
TTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGT
GGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTAT
CATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGA
GGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTAC
AATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTG
CACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAA
TGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTT
GATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATT
TATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGC
TATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCA
ATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAAC
AACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT
GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG
ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGA
AGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATT
CCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTG
AAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATT
TATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGT
TTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAA
TGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGAT
AGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGC
TAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATAT
TGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCT
CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT

GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTT
AAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAA
ATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATT
GGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAG
ACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAG
AATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCT
CTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGT
TTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACA
ATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGT
AGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAG
TTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCA
AACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCAT
GAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT
AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACC
ATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAA
ACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAG
TCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCT
TCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGG
GGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCC
ATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAA
TTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATA
TGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGG
AAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCT
TTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAAT
CATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGT
GCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATT
TAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGT
GGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTT
ATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAA
TACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGAT
GCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG
GAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain).

[0111] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1) or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

[0112] In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity.

For example, in some embodiments, a dCas9 domain comprises D10A and an H840A mutation or corresponding mutations in another Cas9. In some embodiments, the dCas9 comprises the amino acid sequence of dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAK
SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain).

[0113] In some embodiments, the Cas9 domain comprises a D10A mutation, while the residue at position 840 remains a histidine in the amino acid sequence provided above, or at corresponding positions in any of the amino acid sequences provided herein.

[0114] In other embodiments, dCas9 variants having mutations other than D10A and H840A are provided, which, e.g., result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 are provided

which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical. In some embodiments, variants of dCas9 are provided having amino acid sequences which are shorter, or longer, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

[0115] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 protein, e.g., one of the Cas9 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only one or more fragments thereof. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

[0116] A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that has complementary to the guide RNA. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Examples of nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (*e.g.*, dCas9 and nCas9), CasX, CasY, Cpf1, CAS12b/C2c1, and Cas12c/C2c3.

[0117] A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9) or catalytically inactive Cas9. Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al., Science.* 337:816-821(2012); Qi *et al.,* "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)).

[0118] In some embodiments, the Cas9 domain is a Cas9 nickase. The Cas9 nickase may be a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule

(*e.g.*, a duplexed DNA molecule). In some embodiments, the Cas9 nickase cleaves the target strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is base paired to (complementary to) a gRNA (*e.g.*, an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position 840. In some embodiments, the Cas9 nickase cleaves the non-target, non-base-edited strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is not base paired to a gRNA (*e.g.*, an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises an H840A mutation and has an aspartic acid residue at position 10, or a corresponding mutation. In some embodiments, the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 nickases provided herein. Additional suitable Cas9 nickases will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure.

[0119] In some embodiments, the Cas9 domain is a nuclease-inactive Cas9 domain (dCas9). For example, the dCas9 domain may bind to a duplexed nucleic acid molecule (*e.g.*, via a gRNA molecule) without cleaving either strand of the duplexed nucleic acid molecule. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a H840X mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid change. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10A mutation and a H840A mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein. As one example, a nuclease-inactive Cas9 domain comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR

KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD (see, *e.g.*, Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell.* 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference).

[0120] It should be appreciated that additional Cas9 proteins (*e.g.*, a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9), including variants and homologs thereof, are within the scope of this disclosure. Exemplary Cas9 proteins include, without limitation, those provided below. In some embodiments, the Cas9 protein is a nuclease dead Cas9 (dCas9). In some embodiments, the Cas9 protein is a Cas9 nickase (nCas9). In some embodiments, the Cas9 protein is a nuclease active Cas9.

[0121] Exemplary catalytically inactive Cas9 (dCas9):
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY

HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

[0122] Exemplary catalytically Cas9 nickase (nCas9):
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET

NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

[0123] Exemplary catalytically active Cas9:
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD.

[0124] In some embodiments, Cas9 refers to a Cas9 from archaea (e.g. nanoarchaea), which constitute a domain and kingdom of single-celled prokaryotic microbes. In some embodiments,

a nucleic acid programmable DNA binding protein refers to CasX or CasY, which have been described in, for example, Burstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21, the entire contents of which is hereby incorporated by reference. Using genome-resolved metagenomics, a number of CRISPR-Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in little- studied nanoarchaea as part of an active CRISPR-Cas system. In bacteria, two previously unknown systems were discovered, CRISPR-CasX and CRISPR-CasY, which are among the most compact systems yet discovered. In some embodiments, in a base editor system described herein Cas9 is replaced by CasX, or a variant of CasX. In some embodiments, in a base editor system described herein Cas9 is replaced by CasY, or a variant of CasY. It should be appreciated that other RNA-guided DNA binding proteins may be used as a nucleic acid programmable DNA binding protein (napDNAbp), and are within the scope of this disclosure.

[0125] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a CasX or CasY protein. In some embodiments, the napDNAbp is a CasX protein. In some embodiments, the napDNAbp is a CasY protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring CasX or CasY protein. In some embodiments, the napDNAbp is a naturally-occurring CasX or CasY protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any CasX or CasY protein described herein. It should be appreciated that CasX and CasY from other bacterial species may also be used in accordance with the present disclosure.

[0126] CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIH CRISPR-associated Casx protein OS = *Sulfolobus islandicus* (strain HVE10/4) GN = SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAER
RGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKEC
EEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGD
YVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISD
AVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLT
G SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0127] >tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = *Sulfolobus islandicus* (strain REY15A) GN=SiRe_0771 PE=4 SV=1
MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAER
RGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKEC
EEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEG
DYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTIS
DAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYL
TGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

Deltaproteobacteria CasX
MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVIS
NNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPE
MDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLK
PVKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDA
CMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVD
fAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTI
NEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAK
RQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLT
DWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIG
SDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGG
GKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKK
IGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSG
GPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFY
HAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKT
LAQYTSKTCSNCGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYNRYK
RQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGH
EVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0128] CasY (ncbi.nlm.nih.gov/protein/APG80656.1)
>APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]
MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDY
VGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRG
GSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKD

QCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFD
TVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLR
ENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDING
KLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVP
DDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKK
KSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKI
YKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAE
NEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHE
EYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFS
QSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDL
APAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPV
KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDE
KKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTA
LEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNR
IHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKL
AVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPP
IFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEE
KKVEDYFERFRKLKNIKVLGQMKKI

[0129] It should be appreciated that polynucleotide programmable nucleotide binding domains can also include nucleic acid programmable proteins that bind RNA. For example, the polynucleotide programmable nucleotide binding domain can be associated with a nucleic acid that guides the polynucleotide programmable nucleotide binding domain to an RNA. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, though they are not specifically listed in this disclosure.

Cas proteins that can be used herein include class 1 and class 2. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 or Csx12), Cas10, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, and Cas12i, CARF, DinG, homologues thereof, or modified versions thereof. An unmodified CRISPR enzyme can have DNA cleavage activity, such as Cas9, which has two functional endonuclease domains: RuvC and HNH. A CRISPR enzyme can direct cleavage of one or both

strands at a target sequence, such as within a target sequence and/or within a complement of a target sequence. For example, a CRISPR enzyme can direct cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

[0130] A vector that encodes a CRISPR enzyme that is mutated to with respect, to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence can be used. Cas9 can refer to a polypeptide with at least or at least about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence homology to a wild type exemplary Cas9 polypeptide (*e.g.*, Cas9 from *S. pyogenes*). Cas9 can refer to a polypeptide with at most or at most about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence homology to a wild type exemplary Cas9 polypeptide (*e.g.*, from *S. pyogenes*). Cas9 can refer to the wild type or a modified form of the Cas9 protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof.

[0131] In some embodiments, the methods described herein can utilize an engineered Cas protein. A guide RNA (gRNA) is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Thus, a skilled artisan can change the genomic target of the Cas protein specificity is partially determined by how specific the gRNA targeting sequence is for the genomic target compared to the rest of the genome.

[0132] The Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9 undergoes a second conformational change upon target binding that positions the nuclease domains to cleave opposite strands of the target DNA. The end result of Cas9-mediated DNA cleavage is a double-strand break (DSB) within the target DNA (~3-4 nucleotides upstream of the PAM sequence). The resulting DSB is then repaired by one of two general repair pathways: (1) the efficient but error-prone non-homologous end joining (NHEJ) pathway; or (2) the less efficient but high-fidelity homology directed repair (HDR) pathway.

[0133] The "efficiency" of non-homologous end joining (NHEJ) and/or homology directed repair (HDR) can be calculated by any convenient method. For example, in some cases, efficiency can be expressed in terms of percentage of successful HDR. For example, a surveyor nuclease assay can be used can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage. For example, a surveyor nuclease enzyme can be used that directly cleaves DNA containing a newly integrated restriction sequence as the

result of successful HDR. More cleaved substrate indicates a greater percent HDR (a greater efficiency of HDR). As an illustrative example, a fraction (percentage) of HDR can be calculated using the following equation [(cleavage products)/(substrate plus cleavage products)] (*e.g.*, (b+c)/(a+b+c), where "a" is the band intensity of DNA substrate and "b" and "c" are the cleavage products).

[0134] In some cases, efficiency can be expressed in terms of percentage of successful NHEJ. For example, a T7 endonuclease I assay can be used to generate cleavage products, and the ratio of products to substrate can be used to calculate the percentage NHEJ. T7 endonuclease I cleaves mismatched heteroduplex DNA which arises from hybridization of wild-type and mutant DNA strands (NHEJ generates small random insertions or deletions (indels) at the site of the original break). More cleavage indicates a greater percent NHEJ (a greater efficiency of NHEJ). As an illustrative example, a fraction (percentage) of NHEJ can be calculated using the following equation: $(1-(1-(b+c)/(a+b+c))^{1/2}) \times 100$, where "a" is the band intensity of DNA substrate and "b" and "c" are the cleavage products (Ran *et. al.*, Cell. 2013 Sep. 12; 154(6):1380-9; and Ran *et al.*, Nat Protoc. 2013 Nov.; 8(11): 2281–2308).

[0135] The NHEJ repair pathway is the most active repair mechanism, and it frequently causes small nucleotide insertions or deletions (indels) at the DSB site. The randomness of NHEJ-mediated DSB repair has important practical implications, because a population of cells expressing Cas9 and a gRNA or a guide polynucleotide can result in a diverse array of mutations. In most cases, NHEJ gives rise to small indels in the target DNA that result in amino acid deletions, insertions, or frameshift mutations leading to premature stop codons within the open reading frame (ORF) of the targeted gene. The ideal end result is a loss-of-function mutation within the targeted gene.

[0136] While NHEJ-mediated DSB repair often disrupts the open reading frame of the gene, homology directed repair (HDR) can be used to generate specific nucleotide changes ranging from a single nucleotide change to large insertions like the addition of a fluorophore or tag.

[0137] In order to utilize HDR for gene editing, a DNA repair template containing the desired sequence can be delivered into the cell type of interest with the gRNA(s) and Cas9 or Cas9 nickase. The repair template can contain the desired edit as well as additional homologous sequence immediately upstream and downstream of the target (termed left & right homology arms). The length of each homology arm can be dependent on the size of the change being introduced, with larger insertions requiring longer homology arms. The repair template can be a single-stranded oligonucleotide, double-stranded oligonucleotide, or a double-stranded DNA plasmid. The efficiency of HDR is generally low (<10% of modified alleles) even in cells that

express Cas9, gRNA and an exogenous repair template. The efficiency of HDR can be enhanced by synchronizing the cells, since HDR takes place during the S and G2 phases of the cell cycle. Chemically or genetically inhibiting genes involved in NHEJ can also increase HDR frequency.

[0138] In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence can have additional sites throughout the genome where partial homology exists. These sites are called off-targets and need to be considered when designing a gRNA. In addition to optimizing gRNA design, CRISPR specificity can also be increased through modifications to Cas9. Cas9 generates double-strand breaks (DSBs) through the combined activity of two nuclease domains, RuvC and HNH. Cas9 nickase, a D10A mutant of SpCas9, retains one nuclease domain and generates a DNA nick rather than a DSB. The nickase system can also be combined with HDR-mediated gene editing for specific gene edits.

[0139] In some cases, Cas9 is a variant Cas9 protein. A variant Cas9 polypeptide has an amino acid sequence that is different by one amino acid (*e.g.*, has a deletion, insertion, substitution, fusion) when compared to the amino acid sequence of a wild type Cas9 protein. In some instances, the variant Cas9 polypeptide has an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the nuclease activity of the Cas9 polypeptide. For example, in some instances, the variant Cas9 polypeptide has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9 protein. In some cases, the variant Cas9 protein has no substantial nuclease activity. When a subject Cas9 protein is a variant Cas9 protein that has no substantial nuclease activity, it can be referred to as "dCas9."

[0140] In some cases, a variant Cas9 protein has reduced nuclease activity. For example, a variant Cas9 protein 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 endonuclease activity of a wild-type Cas9 protein, *e.g.*, a wild-type Cas9 protein.

[0141] In some cases, a variant Cas9 protein can cleave the complementary strand of a guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the RuvC domain. As a non-limiting example, in some embodiments, a variant Cas9 protein has a D10A (aspartate to alanine at amino acid position 10) and can therefore cleave the complementary strand of a double stranded guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence (thus resulting in a single strand break (SSB) instead of a

double strand break (DSB) when the variant Cas9 protein cleaves a double stranded target nucleic acid) (see, for example, Jinek et al., Science. 2012 Aug. 17; 337(6096):816-21).

[0142] In some cases, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB when the variant Cas9 protein cleaves a double stranded guide target sequence). Such a Cas9 protein has a reduced ability to cleave a guide target sequence (*e.g.*, a single stranded guide target sequence) but retains the ability to bind a guide target sequence (*e.g.*, a single stranded guide target sequence).

[0143] In some cases, a variant Cas9 protein has a reduced ability to cleave both the complementary and the non-complementary strands of a double stranded target DNA. As a non-limiting example, in some cases, the variant Cas9 protein harbors both the D10A and the H840A mutations such that the polypeptide has a reduced ability to cleave both the complementary and the non-complementary strands of a double stranded target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

[0144] As another non-limiting example, in some cases, the variant Cas9 protein harbors W476A and W1126A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

[0145] As another non-limiting example, in some cases, the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA).

[0146] As another non-limiting example, in some cases, the variant Cas9 protein harbors H840A, W476A, and W1126A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded

target DNA). As another non-limiting example, in some cases, the variant Cas9 protein harbors H840A, D10A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, the variant Cas9 has restored catalytic His residue at position 840 in the Cas9 HNH domain (A840H).

[0147] As another non-limiting example, in some cases, the variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some cases, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some cases, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such cases, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some cases, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

[0148] In some embodiments, a variant Cas9 protein that has reduced catalytic activity (*e.g.*, when a Cas9 protein has a D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987 mutation, *e.g.*, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A), the variant Cas9 protein can still bind to target DNA in a site-specific manner (because it is still guided to a target DNA sequence by a guide RNA) as long as it retains the ability to interact with the guide RNA.

[0149] Alternatives to *S. pyogenes* Cas9 can include RNA-guided endonucleases from the Cpf1 family that display cleavage activity in mammalian cells. CRISPR from *Prevotella* and

Francisella 1 (CRISPR/Cpf1) is a DNA-editing technology analogous to the CRISPR/Cas9 system. Cpf1 is an RNA-guided endonuclease of a class II CRISPR/Cas system. This acquired immune mechanism is found in *Prevotella* and *Francisella* bacteria. Cpf1 genes are associated with the CRISPR locus, coding for an endonuclease that use a guide RNA to find and cleave viral DNA. Cpf1 is a smaller and simpler endonuclease than Cas9, overcoming some of the CRISPR/Cas9 system limitations. Unlike Cas9 nucleases, the result of Cpf1-mediated DNA cleavage is a double-strand break with a short 3′ overhang. Cpf1's staggered cleavage pattern can open up the possibility of directional gene transfer, analogous to traditional restriction enzyme cloning, which can increase the efficiency of gene editing. Like the Cas9 variants and orthologues described above, Cpf1 can also expand the number of sites that can be targeted by CRISPR to AT-rich regions or AT-rich genomes that lack the NGG PAM sites favored by SpCas9. The Cpf1 locus contains a mixed alpha/beta domain, a RuvC-I followed by a helical region, a RuvC-II and a zinc finger-like domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9. Furthermore, Cpf1 does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alpha-helical recognition lobe of Cas9. Cpf1 CRISPR-Cas domain architecture shows that Cpf1 is functionally unique, being classified as Class 2, type V CRISPR system. The Cpf1 loci encode Cas1, Cas2 and Cas4 proteins more similar to types I and III than from type II systems. Functional Cpf1 does not require the trans-activating CRISPR RNA (tracrRNA), therefore, only CRISPR (crRNA) is required. This benefits genome editing because Cpf1 is not only smaller than Cas9, but also it has a smaller sgRNA molecule (proximately half as many nucleotides as Cas9). The Cpf1-crRNA complex cleaves target DNA or RNA by identification of a protospacer adjacent motif 5'-YTN-3' in contrast to the G-rich PAM targeted by Cas9. After identification of PAM, Cpf1 introduces a sticky-end-like DNA double- stranded break of 4 or 5 nucleotides overhang.

[0150] Some aspects of the disclosure provide fusion proteins comprising domains that act as nucleic acid programmable DNA binding proteins, which may be used to guide a protein, such as a base editor, to a specific nucleic acid (e.g., DNA or RNA) sequence. In particular embodiments, a fusion protein comprises a nucleic acid programmable DNA binding protein domain and a deaminase domain. DNA binding proteins include, without limitation, Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, and Cas12i. One example of a programmable polynucleotide-binding protein that has different PAM specificity than Cas9 is Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (Cpf1). Similar to Cas9, Cpf1 is also a

class 2 CRISPR effector. It has been shown that Cpf1 mediates robust DNA interference with features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN, or YTN). Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from Acidaminococcus and Lachnospiraceae are shown to have efficient genome-editing activity in human cells. Cpf1 proteins are known in the art and have been described previously, for example Yamano et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962; the entire contents of which is hereby incorporated by reference.

[0151] Also useful in the present compositions and methods are nuclease-inactive Cpf1 (dCpf1) variants that may be used as a guide nucleotide sequence-programmable polynucleotide-binding protein domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alfa-helical recognition lobe of Cas9. It was shown in Zetsche et al., *Cell,* 163, 759-771, 2015 (which is incorporated herein by reference) that, the RuvC-like domain of Cpf1 is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cpf1 nuclease activity. For example, mutations corresponding to D917A, E1006A, or D1255A in *Francisella novicida* Cpf1 inactivate Cpf1 nuclease activity. In some embodiments, the dCpf1 of the present disclosure comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/E1006A/D1255A. It is to be understood that any mutations, e.g., substitution mutations, deletions, or insertions that inactivate the RuvC domain of Cpf1, may be used in accordance with the present disclosure.

[0152] In some embodiments, the nucleic acid programmable nucleotide binding protein of any of the fusion proteins provided herein may be a Cpf1 protein. In some embodiments, the Cpf1 protein is a Cpf1 nickase (nCpf1). In some embodiments, the Cpf1 protein is a nuclease inactive Cpf1 (dCpf1). In some embodiments, the Cpf1, the nCpf1, or the dCpf1 comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a Cpf1 sequence disclosed herein. In some embodiments, the dCpf1comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a Cpf1 sequence disclosed herein, and comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or

D917A/E1006A/D1255A. It should be appreciated that Cpf1 from other bacterial species may also be used in accordance with the present disclosure.

[0153] Wild type *Francisella novicida* Cpf1 (D917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0154] *Francisella novicida* Cpf1 D917A (A917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI

NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0155] *Francisella novicida* Cpf1 E1006A (D917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI

THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0156] *Francisella novicida* Cpf1 D1255A (D917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0157] *Francisella novicida* Cpf1 D917A/E1006A (A917, A1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0158] *Francisella novicida* Cpf1 D917A/D1255A (A917, E1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ

QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0159] *Francisella novicida* Cpf1 E1006A/D1255A (D917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE

KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0160] *Francisella novicida* Cpf1 D917A/E1006A/D1255A (A917, A1006, and A1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0161] In some embodiments, one of the Cas9 domains present in the fusion protein may be replaced with a guide nucleotide sequence-programmable DNA-binding protein domain that has no requirements for a PAM sequence.

[0162] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a single effector of a microbial CRISPR-Cas system. Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cpf1, Cas12b/C2c1, and Cas12c/C2c3. Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multisubunit effector complexes, while Class 2 systems have a single protein effector. For example, Cas9 and Cpf1 are Class 2 effectors. In addition to Cas9 and Cpf1, three distinct Class 2 CRISPR-Cas systems (Cas12b/C2c1, and Cas12c/C2c3) have been described by Shmakov et al., "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov. 5; 60(3): 385-397, the entire contents of which is hereby incorporated by reference. Effectors of two of the systems, Cas12b/C2c1, and Cas12c/C2c3, contain RuvC-like endonuclease domains related to Cpf1. A third system, contains an effector with two predicated HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by Cas12b/C2c1. Cas12b/C2c1 depends on both CRISPR RNA and tracrRNA for DNA cleavage.

[0163] The crystal structure of *Alicyclobaccillus acidoterrastris* Cas12b/C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See e.g., Liu et al., "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan. 19; 65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See e.g., Yang et al., "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec. 15; 167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with Cas12b/C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between Cas12b/C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

[0164] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12b/C2c1, or a Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a Cas12b/C2c1 protein. In some embodiments, the napDNAbp is a Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or

at ease 99.5% identical to a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any one of the napDNAbp sequences provided herein. It should be appreciated that Cas12b/C2c1 or Cas12c/C2c3 from other bacterial species may also be used in accordance with the present disclosure.

Cas12b/C2c1 (uniprot.org/uniprot/T0D7A2#2)
sp|T0D7A2|/C2C1_ALIAG CRISPR-associated endo- nuclease C2c1 OS =
= *Alicyclobacillus acido- terrestris* ((strain ATCC 49025 / DSM 3922/ CIP 106132 /
NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1
MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDG
EQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKG
DAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSA
DRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERM
MSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESK
EQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAK
LAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGG
NLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPI
ALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARG
ERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLR
TSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAI
REERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREA
FENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRG
YAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKED
RLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENN
QLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQE
HNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAA
QNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYY
ERERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFW
SMV NQRIEGYLVKQIRSRVPLQDSACENTGDI

BhCas12b (Bacillus hisashii) NCBI Reference Sequence: WP_095142515

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQ
EAIYEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEEL
VPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKK
KWEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDM
FIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLR
DTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSV
YEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEER
SGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQI
FLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMT
VNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSID
LGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRK
AREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIR
ELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTR
KFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYD
VRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGL
QVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGD
LYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQT
VYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDS
FDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSS
KQSMKRPAATKKAGQAKKKK

[0165] In some embodiments, the Cas12b is BvCas12B, which is a variant of BhCas12b and comprises the following changes relative to BhCas12B: S893R, K846R, and E837G.
BvCas12b (Bacillus sp. V3-13) NCBI Reference Sequence: WP_101661451.1
MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAY
QAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVD
PNSQSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGL
LPLFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKE
KTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEK
WSKLPESASPEELWKVVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAA
YNGLQKKLSRTKEQATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKII
WPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFN
RKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYK
PKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKL

FYSINDTELFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLET
KKTPDERKKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKESLVELH
HRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRTPGEANRIE
TDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIIL
FENLNRYLFNLDRSRRENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAK
TGAPGIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGELFVTLSKRY
KKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGEDKLYIPKSQTETI
KKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQK
KYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL.

[0166] In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises a N579A mutation, or a corresponding mutation in any of the amino acid sequences provided herein.

[0167] In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT or a NNNRRT PAM sequence. In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation, or one or more corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation, or corresponding mutations in any of the amino acid sequences provided herein.

[0168] In some embodiments, the variant Cas protein can be SpCas9, SpCas9-VRQR, SpCas9-VRER, xCas9 (sp), SaCas9, SaCas9-KKH, SpCas9-MQKSER, SpCas9-LRKIQK, or SpCas9-LRVSQL.

[0169] Exemplary SaCas9 sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN
VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKE
AKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC

TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL
KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR
LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKN
SKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPL
EDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSY
FRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ
KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDD
YPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL
KKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRP
PRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

Residue N579 above, which is underlined and in bold, may be mutated (*e.g.*, to a A579) to yield a SaCas9 nickase.

[0170] Exemplary SaCas9n sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN
VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKE
AKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL
KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR
LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKN
SKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPL
EDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSY
FRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ
KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDD
YPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL

KKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRP
PRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0171] Residue A579 above, which can be mutated from N579 to yield a SaCas9 nickase, is underlined and in bold.

[0172] Exemplary SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRR
HRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN
VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKE
AKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHC
TYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL
KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY
QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR
LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKN
SKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPL
EDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETF
KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSY
FRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKL
DKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
RKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQ
KLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDD
YPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL
KKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP
HIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

Residue A579 above, which can be mutated from N579 to yield a SaCas9 nickase, is underlined and in bold. Residues K781, K967, and H1014 above, which can be mutated from E781, N967, and R1014 to yield a SaKKH Cas9 are underlined and in italics.

[0173] A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains. For example, a polynucleotide programmable nucleotide binding domain can comprise one or more nuclease domains. In some embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain can comprise an endonuclease or an exonuclease. Herein the term "exonuclease" refers to a protein or polypeptide capable of digesting a nucleic acid (*e.g.*, RNA or DNA) from free ends, and the term "endonuclease" refers to a protein or polypeptide capable of catalyzing (*e.g.* cleaving) internal regions in a nucleic acid (*e.g.*, DNA or RNA). In some embodiments, an endonuclease

can cleave a single strand of a double-stranded nucleic acid. In some embodiments, an endonuclease can cleave both strands of a double-stranded nucleic acid molecule. In some embodiments a polynucleotide programmable nucleotide binding domain can be a deoxyribonuclease. In some embodiments a polynucleotide programmable nucleotide binding domain can be a ribonuclease.

[0174] In some embodiments, a nuclease domain of a polynucleotide programmable nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide. In some cases, the polynucleotide programmable nucleotide binding domain can comprise a nickase domain. Herein the term "nickase" refers to a polynucleotide programmable nucleotide binding domain comprising a nuclease domain that is capable of cleaving only one strand of the two strands in a duplexed nucleic acid molecule (*e.g.* DNA). In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.* natural) form of a polynucleotide programmable nucleotide binding domain by introducing one or more mutations into the active polynucleotide programmable nucleotide binding domain. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can include a D10A mutation and a histidine at position 840. In such cases, the residue H840 retains catalytic activity and can thereby cleave a single strand of the nucleic acid duplex. In another example, a Cas9-derived nickase domain can comprise an H840A mutation, while the amino acid residue at position 10 remains a D. In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.* natural) form of a polynucleotide programmable nucleotide binding domain by removing all or a portion of a nuclease domain that is not required for the nickase activity. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can comprise a deletion of all or a portion of the RuvC domain or the HNH domain.

[0175] A base editor comprising a polynucleotide programmable nucleotide binding domain comprising a nickase domain is thus able to generate a single-strand DNA break (nick) at a specific polynucleotide target sequence (*e.g.* determined by the complementary sequence of a bound guide nucleic acid). In some embodiments, the strand of a nucleic acid duplex target polynucleotide sequence that is cleaved by a base editor comprising a nickase domain (*e.g.* Cas9-derived nickase domain) is the strand that is not edited by the base editor (*i.e.*, the strand that is cleaved by the base editor is opposite to a strand comprising a base to be edited). In other embodiments, a base editor comprising a nickase domain (*e.g.* Cas9-derived nickase domain) can cleave the strand of a DNA molecule which is being targeted for editing. In such cases, the non-targeted strand is not cleaved.

[0176] Also provided herein are base editors comprising a polynucleotide programmable nucleotide binding domain which is catalytically dead (*i.e.*, incapable of cleaving a target polynucleotide sequence). Herein the terms "catalytically dead" and "nuclease dead" are used interchangeably to refer to a polynucleotide programmable nucleotide binding domain which has one or more mutations and/or deletions resulting in its inability to cleave a strand of a nucleic acid. In some embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain base editor can lack nuclease activity as a result of specific point mutations in one or more nuclease domains. For example, in the case of a base editor comprising a Cas9 domain, the Cas9 can comprise both a D10A mutation and an H840A mutation. Such mutations inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain can comprise one or more deletions of all or a portion of a catalytic domain (*e.g.* RuvC1 and/or HNH domains). In further embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises a point mutation (*e.g.* D10A or H840A) as well as a deletion of all or a portion of a nuclease domain.

[0177] Also contemplated herein are mutations capable of generating a catalytically dead polynucleotide programmable nucleotide binding domain from a previously functional version of the polynucleotide programmable nucleotide binding domain. For example, in the case of catalytically dead Cas9 ("dCas9"), variants having mutations other than D10A and H840A are provided, which result in nuclease inactivated Cas9. Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain).

[0178] Additional suitable nuclease-inactive dCas9 domains can be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference). In some embodiments, the dCas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the dCas9 domains provided herein. In some embodiments, the Cas9 domain comprises an amino acid

sequences that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more or more mutations compared to any one of the amino acid sequences set forth herein. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth herein.

[0179] Non-limiting examples of a polynucleotide programmable nucleotide binding domain which can be incorporated into a base editor include a CRISPR protein-derived domain, a restriction nuclease, a meganuclease, TAL nuclease (TALEN), and a zinc finger nuclease (ZFN). In some cases, a base editor comprises a polynucleotide programmable nucleotide binding domain comprising a natural or modified protein or portion thereof which via a bound guide nucleic acid is capable of binding to a nucleic acid sequence during CRISPR (*i.e.*, Clustered Regularly Interspaced Short Palindromic Repeats)-mediated modification of a nucleic acid. Such a protein is referred to herein as a "CRISPR protein". Accordingly, disclosed herein is a base editor comprising a polynucleotide programmable nucleotide binding domain comprising all or a portion of a CRISPR protein (*i.e.* a base editor comprising as a domain all or a portion of a CRISPR protein, also referred to as a "CRISPR protein-derived domain" of the base editor). A CRISPR protein-derived domain incorporated into a base editor can be modified compared to a wild-type or natural version of the CRISPR protein. For example, as described below a CRISPR protein-derived domain can comprise one or more mutations, insertions, deletions, rearrangements and/or recombinations relative to a wild-type or natural version of the CRISPR protein.

[0180] In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is an endonuclease (*e.g.*, deoxyribonuclease or ribonuclease) capable of binding a target polynucleotide when in conjunction with a bound guide nucleic acid. In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is a nickase capable of binding a target polynucleotide when in conjunction with a bound guide nucleic acid. In some embodiments, a CRISPR protein-derived domain incorporated into a base editor is a catalytically dead domain capable of binding a target polynucleotide when in conjunction with a bound guide nucleic acid. In some embodiments, a target polynucleotide bound by a CRISPR

protein derived domain of a base editor is DNA. In some embodiments, a target polynucleotide bound by a CRISPR protein-derived domain of a base editor is RNA.

[0181] In some embodiments, a CRISPR protein-derived domain of a base editor can include all or a portion of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

[0182] In some embodiments, a Cas9-derived domain of a base editor is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 domain comprises a N579X mutation. In some embodiments, the SaCas9 domain comprises a N579A mutation. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a NNGRRT PAM sequence. In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation.

[0183] A base editor can comprise a domain derived from all or a portion of a Cas9 that is a high fidelity Cas9. In some embodiments, high fidelity Cas9 domains of a base editor are engineered Cas9 domains comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and the sugar-phosphate backbone of a DNA, relative to a corresponding wild-type Cas9 domain. High fidelity Cas9 domains that have decreased electrostatic interactions with the sugar-phosphate backbone of DNA can have less off-target effects. In some embodiments, the Cas9 domain (*e.g.*, a wild type Cas9 domain) comprises one or more mutations that decrease the association between the Cas9 domain and the sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and the sugar-phosphate backbone of DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%,

at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or more.

[0184] In some embodiments, the modified Cas9 is a high fidelity Cas9 enzyme. In some embodiments, the high fidelity Cas9 enzyme is SpCas9 (K855A), eSpCas9(1.1), SpCas9-HF1, or hyper accurate Cas9 variant (HypaCas9). The modified Cas9 eSpCas9(1.1) contains alanine substitutions that weaken the interactions between the HNH/RuvC groove and the non-target DNA strand, preventing strand separation and cutting at off-target sites. Similarly, SpCas9-HF1 lowers off-target editing through alanine substitutions that disrupt Cas9's interactions with the DNA phosphate backbone. HypaCas9 contains mutations (SpCas9 N692A/M694A/Q695A/H698A) in the REC3 domain that increase Cas9 proofreading and target discrimination. All three high fidelity enzymes generate less off-target editing than wildtype Cas9. An exemplary high fidelity Cas9 is provided below.

High Fidelity Cas9 domain mutations relative to Cas9 are shown in bold and underlines

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH

YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL SQLGGD

[0185]

Guide Polynucleotides

[0186] As used herein, the term "guide polynucleotide(s)" refer to a polynucleotide which can be specific for a target sequence and can form a complex with a polynucleotide programmable nucleotide binding domain protein (e.g., Cas9 or Cpf1). In an embodiment, the guide polynucleotide is a guide RNA. As used herein, the term "guide RNA (gRNA)" and its grammatical equivalents can refer to an RNA which can be specific for a target DNA and can form a complex with Cas protein. An RNA/Cas complex can assist in "guiding" Cas protein to a target DNA. Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA" or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M. *et al.*, Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti, J.J. *et al.*, Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E. *et al.*, Nature 471:602-607(2011); and "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M.*et al*, Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and S. *thermophilus*. Additional suitable Cas9 nucleases and sequences can be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase.

[0187] In some embodiments, the guide polynucleotide is at least one single guide RNA ("sgRNA" or "gNRA"). In some embodiments, the guide polynucleotide is at least one tracrRNA. In some embodiments, the guide polynucleotide does not require PAM sequence to guide the polynucleotide-programmable DNA-binding domain (e.g., Cas9 or Cpf1) to the target nucleotide sequence.

[0188] The polynucleotide programmable nucleotide binding domain (*e.g.*, a CRISPR-derived domain) of the base editors disclosed herein can recognize a target polynucleotide sequence by associating with a guide polynucleotide. A guide polynucleotide (*e.g.*, gRNA) is typically single-stranded and can be programmed to site-specifically bind (*i.e.*, via complementary base pairing) to a target sequence of a polynucleotide, thereby directing a base editor that is in conjunction with the guide nucleic acid to the target sequence. A guide polynucleotide can be DNA. A guide polynucleotide can be RNA. In some cases, the guide polynucleotide comprises natural nucleotides (*e.g.*, adenosine). In some cases, the guide polynucleotide comprises non-natural (or unnatural) nucleotides (*e.g.*, peptide nucleic acid or nucleotide analogs). In some cases, the targeting region of a guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20 nucleotides in length.

[0189] In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via for example complementary base pairing (*e.g.* a dual guide polynucleotide). For example, a guide polynucleotide can comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). For example, a guide polynucleotide can comprise one or more trans-activating CRISPR RNA (tracrRNA).

[0190] In type II CRISPR systems, targeting of a nucleic acid by a CRISPR protein (*e.g.* Cas9) typically requires complementary base pairing between a first RNA molecule (crRNA) comprising a sequence that recognizes the target sequence and a second RNA molecule (trRNA) comprising repeat sequences which forms a scaffold region that stabilizes the guide RNA-CRISPR protein complex. Such dual guide RNA systems can be employed as a guide polynucleotide to direct the base editors disclosed herein to a target polynucleotide sequence.

[0191] In some embodiments, the base editor provided herein utilizes a single guide polynucleotide (*e.g.*, gRNA). In some embodiments, the base editor provided herein utilizes a dual guide polynucleotide (*e.g.*, dual gRNAs). In some embodiments, the base editor provided herein utilizes one or more guide polynucleotide (*e.g.*, multiple gRNA). In some embodiments, a single guide polynucleotide is utilized for different base editors described herein. For

example, a single guide polynucleotide can be utilized for a cytidine base editor and an adenosine base editor.

[0192] In other embodiments, a guide polynucleotide can comprise both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide polynucleotide can be a single guide RNA (sgRNA or gRNA). Herein the term guide polynucleotide sequence contemplates any single, dual or multi-molecule nucleic acid capable of interacting with and directing a base editor to a target polynucleotide sequence.

[0193] Typically, a guide polynucleotide (*e.g.*, crRNA/trRNA complex or a gRNA) comprises a "polynucleotide-targeting segment" that includes a sequence capable of recognizing and binding to a target polynucleotide sequence, and a "protein-binding segment" that stabilizes the guide polynucleotide within a polynucleotide programmable nucleotide binding domain component of a base editor. In some embodiments, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to a DNA polynucleotide, thereby facilitating the editing of a base in DNA. In other cases, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to an RNA polynucleotide, thereby facilitating the editing of a base in RNA. Herein a "segment" refers to a section or region of a molecule, *e.g.*, a contiguous stretch of nucleotides in the guide polynucleotide. A segment can also refer to a region/section of a complex such that a segment can comprise regions of more than one molecule. For example, where a guide polynucleotide comprises multiple nucleic acid molecules, the protein-binding segment of can include all or a portion of multiple separate molecules that are for instance hybridized along a region of complementarity. In some embodiments, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and can include regions of RNA molecules that are of any total length and can include regions with complementarity to other molecules.

[0194] A guide RNA or a guide polynucleotide can comprise two or more RNAs, *e.g.*, CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA). A guide RNA or a guide polynucleotide can sometimes comprise a single-chain RNA, or single guide RNA (sgRNA) formed by fusion of a portion (*e.g.*, a functional portion) of crRNA and tracrRNA. A guide

RNA or a guide polynucleotide can also be a dual RNA comprising a crRNA and a tracrRNA. Furthermore, a crRNA can hybridize with a target DNA.

[0195] As discussed above, a guide RNA or a guide polynucleotide can be an expression product. For example, a DNA that encodes a guide RNA can be a vector comprising a sequence coding for the guide RNA. A guide RNA or a guide polynucleotide can be transferred into a cell by transfecting the cell with an isolated guide RNA or plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A guide RNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery.

[0196] A guide RNA or a guide polynucleotide can be isolated. For example, a guide RNA can be transfected in the form of an isolated RNA into a cell or organism. A guide RNA can be prepared by *in vitro* transcription using any *in vitro* transcription system known in the art. A guide RNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a guide RNA.

[0197] A guide RNA or a guide polynucleotide can comprise three regions: a first region at the 5' end that can be complementary to a target site in a chromosomal sequence, a second internal region that can form a stem loop structure, and a third 3' region that can be single-stranded. A first region of each guide RNA can also be different such that each guide RNA guides a fusion protein to a specific target site. Further, second and third regions of each guide RNA can be identical in all guide RNAs.

[0198] A first region of a guide RNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the guide RNA can base pair with the target site. In some cases, a first region of a guide RNA can comprise from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or from about 10 nucleotides to about 25 nucleotides; or from 10 nucleotides to about 25 nucleotides; or from about 10 nucleotides to 25 nucleotides) or more. For example, a region of base pairing between a first region of a guide RNA and a target site in a chromosomal sequence can be or can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more nucleotides in length. Sometimes, a first region of a guide RNA can be or can be about 19, 20, or 21 nucleotides in length.

[0199] A guide RNA or a guide polynucleotide can also comprise a second region that forms a secondary structure. For example, a secondary structure formed by a guide RNA can comprise a stem (or hairpin) and a loop. A length of a loop and a stem can vary. For example, a loop can range from or from about 3 to 10 nucleotides in length, and a stem can range from or from about 6 to 20 base pairs in length. A stem can comprise one or more bulges of 1 to 10 or about 10

nucleotides. The overall length of a second region can range from or from about 16 to 60 nucleotides in length. For example, a loop can be or can be about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

[0200] A guide RNA or a guide polynucleotide can also comprise a third region at the 3' end that can be essentially single-stranded. For example, a third region is sometimes not complementarity to any chromosomal sequence in a cell of interest and is sometimes not complementarity to the rest of a guide RNA. Further, the length of a third region can vary. A third region can be more than or more than about 4 nucleotides in length. For example, the length of a third region can range from or from about 5 to 60 nucleotides in length.

[0201] A guide RNA or a guide polynucleotide can target any exon or intron of a gene target. In some cases, a guide can target exon 1 or 2 of a gene, in other cases; a guide can target exon 3 or 4 of a gene. A composition can comprise multiple guide RNAs that all target the same exon or in some cases, multiple guide RNAs that can target different exons. An exon and an intron of a gene can be targeted.

[0202] A guide RNA or a guide polynucleotide can target a nucleic acid sequence of or of about 20 nucleotides. A target nucleic acid can be less than or less than about 20 nucleotides. A target nucleic acid can be at least or at least about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, or anywhere between 1-100 nucleotides in length. A target nucleic acid can be at most or at most about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, or anywhere between 1-100 nucleotides in length. A target nucleic acid sequence can be or can be about 20 bases immediately 5' of the first nucleotide of the PAM. A guide RNA can target a nucleic acid sequence. A target nucleic acid can be at least or at least about 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, or 1-100 nucleotides.

[0203] A guide polynucleotide, for example, a guide RNA, can refer to a nucleic acid that can hybridize to another nucleic acid, for example, the target nucleic acid or protospacer in a genome of a cell. A guide polynucleotide can be RNA. A guide polynucleotide can be DNA. The guide polynucleotide can be programmed or designed to bind to a sequence of nucleic acid site-specifically. A guide polynucleotide can comprise a polynucleotide chain and can be called a single guide polynucleotide. A guide polynucleotide can comprise two polynucleotide chains and can be called a double guide polynucleotide. A guide RNA can be introduced into a cell or embryo as an RNA molecule. For example, a RNA molecule can be transcribed *in vitro* and/or can be chemically synthesized. An RNA can be transcribed from a synthetic DNA molecule, *e.g.*, a gBlocks® gene fragment. A guide RNA can then be introduced into a cell or embryo as an RNA molecule. A guide RNA can also be introduced into a cell or embryo in the form of a

non-RNA nucleic acid molecule, *e.g.*, DNA molecule. For example, a DNA encoding a guide RNA can be operably linked to promoter control sequence for expression of the guide RNA in a cell or embryo of interest. A RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Plasmid vectors that can be used to express guide RNA include, but are not limited to, px330 vectors and px333 vectors. In some cases, a plasmid vector (*e.g.*, px333 vector) can comprise at least two guide RNA-encoding DNA sequences.

[0204] Methods for selecting, designing, and validating guide polynucleotides, e.g. guide RNAs and targeting sequences are described herein and known to those skilled in the art. For example, to minimize the impact of potential substrate promiscuity of a deaminase domain in the nucleobase editor system (e.g., an AID domain), the number of residues that could unintentionally be targeted for deamination (e.g., off-target C residues that could potentially reside on ssDNA within the target nucleic acid locus) may be minimized. In addition, software tools can be used to optimize the gRNAs corresponding to a target nucleic acid sequence, e.g., to minimize total off-target activity across the genome. For example, for each possible targeting domain choice using S. pyogenes Cas9, all off-target sequences (preceding selected PAMs, e.g. NAG or NGG) may be identified across the genome that contain up to certain number (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. First regions of gRNAs complementary to a target site can be identified, and all first regions (e.g. crRNAs) can be ranked according to its total predicted off-target score; the top-ranked targeting domains represent those that are likely to have the greatest on-target and the least off-target activity. Candidate targeting gRNAs can be functionally evaluated by using methods known in the art and/or as set forth herein.

[0205] As a non-limiting example, target DNA hybridizing sequences in crRNAs of a guide RNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design may be carried out using custom gRNA design software based on the public tool cas-offinder as described in Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475 (2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally-determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for a target

nucleic acid sequence, e.g. a target gene may be obtained and repeat elements may be screened using publically available tools, for example, the RepeatMasker program. RepeatMasker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

[0206] Following identification, first regions of guide RNAs, e.g. crRNAs, may be ranked into tiers based on their distance to the target site, their orthogonality and presence of 5' nucleotides for close matches with relevant PAM sequences (for example, a 5′ G based on identification of close matches in the human genome containing a relevant PAM e.g., NGG PAM for S. pyogenes, NNGRRT or NNGRRV PAM for S. aureus). As used herein, orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer targeting domains that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality may be selected to minimize off-target DNA cleavage.

[0207] In some embodiments, a reporter system may be used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system may comprise a reporter gene based assay where base editing activity leads to expression of the reporter gene. For example, a reporter system may include a reporter gene comprising a deactivated start codon, e.g., a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art. Non-limiting examples of reporter genes include gene encoding green fluorescence protein (GFP), red fluorescence protein (RFP), luciferase, secreted alkaline phosphatase (SEAP), or any other gene whose expression are detectable and apparent to those skilled in the art. The reporter system can be used to test many different gRNAs, e.g., in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase will target. sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific base editing protein, e.g. a Cas9 deaminase fusion protein. In some embodiments, such gRNAs can be designed such that the mutated start codon will not be base-paired with the gRNA. The guide polynucleotides can comprise standard ribonucleotides, modified ribonucleotides (e.g., pseudouridine), ribonucleotide isomers, and/or ribonucleotide analogs. In some embodiments, the guide polynucleotide can comprise at least one detectable label. The detectable label can be a

fluorophore (e.g., FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (e.g., biotin, digoxigenin, and the like), quantum dots, or gold particles.

[0208] The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the guide RNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the guide RNA can be synthesized in vitro by operably linking DNA encoding the guide RNA to a promoter control sequence that is recognized by a phage RNA polymerase. Examples of suitable phage promoter sequences include T7, T3, SP6 promoter sequences, or variations thereof. In embodiments in which the guide RNA comprises two separate molecules (e.g.., crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

[0209] In some embodiments, a base editor system may comprise multiple guide polynucleotides, e.g. gRNAs. For example, the gRNAs may target to one or more target loci (e.g., at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a base editor system. Said multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

[0210] A DNA sequence encoding a guide RNA or a guide polynucleotide can also be part of a vector. Further, a vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, GFP or antibiotic resistance genes such as puromycin), origins of replication, and the like. A DNA molecule encoding a guide RNA can also be linear. A DNA molecule encoding a guide RNA or a guide polynucleotide can also be circular.

[0211] In some embodiments, one or more components of a base editor system may be encoded by DNA sequences. Such DNA sequences may be introduced into an expression system, e.g. a cell, together or separately. For example, DNA sequences encoding a polynucleotide programmable nucleotide binding domain and a guide RNA may be introduced into a cell, each DNA sequence can be part of a separate molecule (*e.g.*, one vector containing the polynucleotide programmable nucleotide binding domain coding sequence and a second vector containing the guide RNA coding sequence) or both can be part of a same molecule (*e.g.*, one vector containing coding (and regulatory) sequence for both the polynucleotide programmable nucleotide binding domain and the guide RNA).

[0212] A guide polynucleotide can comprise one or more modifications to provide a nucleic acid with a new or enhanced feature. A guide polynucleotide can comprise a nucleic acid

affinity tag. A guide polynucleotide can comprise synthetic nucleotide, synthetic nucleotide analog, nucleotide derivatives, and/or modified nucleotides.

[0213] In some cases, a gRNA or a guide polynucleotide can comprise modifications. A modification can be made at any location of a gRNA or a guide polynucleotide. More than one modification can be made to a single gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide can undergo quality control after a modification. In some cases, quality control can include PAGE, HPLC, MS, or any combination thereof.

[0214] A modification of a gRNA or a guide polynucleotide can be a substitution, insertion, deletion, chemical modification, physical modification, stabilization, purification, or any combination thereof.

[0215] A gRNA or a guide polynucleotide can also be modified by 5'adenylate, 5'guanosine-triphosphate cap, 5'N7-Methylguanosine-triphosphate cap, 5'triphosphate cap, 3'phosphate, 3'thiophosphate, 5'phosphate, 5'thiophosphate, Cis-Syn thymidine dimer, trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9,3'-3' modifications, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3'DABCYL, black hole quencher 1, black hole quencer 2, DABCYL SE, dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, carboxyl linker, thiol linkers, 2'-deoxyribonucleoside analog purine, 2'-deoxyribonucleoside analog pyrimidine, ribonucleoside analog, 2'-O-methyl ribonucleoside analog, sugar modified analogs, wobble/universal bases, fluorescent dye label, 2'-fluoro RNA, 2'-O-methyl RNA, methylphosphonate, phosphodiester DNA, phosphodiester RNA, phosphothioate DNA, phosphorothioate RNA, UNA, pseudouridine-5'-triphosphate, 5'-methylcytidine-5'-triphosphate, or any combination thereof.

[0216] In some cases, a modification is permanent. In other cases, a modification is transient. In some cases, multiple modifications are made to a gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide modification can alter physiochemical properties of a nucleotide, such as their conformation, polarity, hydrophobicity, chemical reactivity, base-pairing interactions, or any combination thereof.

[0217] A modification can also be a phosphorothioate substitute. In some cases, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases and; a modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a gRNA or a guide polynucleotide. A modification can also enhance biological activity. In some

cases, a phosphorothioate enhanced RNA gRNA can inhibit RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA gRNAs to be used in applications where exposure to nucleases is of high probability *in vivo* or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'- or ''-end of a gRNA which can inhibit exonuclease degradation. In some cases, phosphorothioate bonds can be added throughout an entire gRNA to reduce attack by endonucleases.

Protospacer Adjacent Motif

[0218] The term "protospacer adjacent motif (PAM)")" or PAM-like motif refers to a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. In some embodiments, the PAM can be a 5' PAM (*i.e.*, located upstream of the 5' end of the protospacer). In other embodiments, the PAM can be a 3' PAM (*i.e.*, located downstream of the 5' end of the protospacer).

[0219] The protospacer adjacent motif (PAM) or PAM-like motif refers to a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. In some embodiments, the PAM can be a 5' PAM (*i.e.*, located upstream of the 5' end of the protospacer). In other embodiments, the PAM can be a 3' PAM (*i.e.*, located downstream of the 5' end of the protospacer). The PAM sequence is essential for target binding, but the exact sequence depends on a type of Cas protein.

[0220] A base editor provided herein can comprise a CRISPR protein-derived domain that is capable of binding a nucleotide sequence that contains a canonical or non-canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion of CRISPR proteins that have different PAM specificities. For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. A PAM can be CRISPR protein-specific and can be different between different base editors comprising different CRISPR protein-derived domains. A PAM can be 5' or 3' of a target sequence. A PAM can be upstream or downstream of a target sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. Often, a PAM is between 2-6 nucleotides in length.

[0221] In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some embodiments, the SpCas9

comprises a D9X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is amino acid except for D. In some embodiments, the SpCas9 comprises a D9A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having an NGG, a NGA, or a NGCG PAM sequence. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135E, R1335Q, and T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135E, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a G1218X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9. In some embodiments, the SpCas9 domain comprises a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.

[0222] In some embodiments, the Cas9 domains of any of the fusion proteins provided herein comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a Cas9 polypeptide described herein. In some embodiments, the Cas9 domains of any of the fusion proteins provided herein comprises the amino acid sequence of any Cas9 polypeptide described herein. In some embodiments, the Cas9

domains of any of the fusion proteins provided herein consists of the amino acid sequence of any Cas9 polypeptide described herein.

[0223] The sequences of exemplary SpCas9 proteins capable of binding a PAM sequence follow:

[0224] Exemplary SpCas9

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

[0225] Exemplary SpCas9n

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF

GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

[0226] Exemplary SpEQR Cas9

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK

GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFESPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

Residues E1135, Q1335, and R1337 above, which can be mutated from D1135, R1335, and T1337 to yield a SpEQR Cas9, are underlined and in bold.

[0227] Exemplary SpVQR Cas9

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE

AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD

Residues V1135, Q1335, and R1337 above, which can be mutated from D1135, R1335, and T1337 to yield a SpVQR Cas9, are underlined and in bold.

[0228] Exemplary SpVRER Cas9

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD.

Exemplary SpVRQR Cas9

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF

GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGD.

Residues V1135, R1218, Q1335, and R1337 above, which can be mutated from D1135, G1218, R1335, and T1337 to yield a SpVRQR Cas9, are underlined and in bold.

[0229] In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some embodiments, the recombinant Cas9 domain is a SpyMacCas9 domain. In some embodiments, the SpyMacCas9 domain is a nuclease active SpyMacCas9, a nuclease inactive SpyMacCas9 (SpyMacCas9d), or a SpyMacCas9 nickase (SpyMacCas9n). In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpyMacCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a NAA PAM sequence.

Exemplary SpyMacCas9

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNF
EEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYH
DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRR
YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS
GQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFV
YGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPL
KKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVG
KNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDY
LQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQL
GATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

High fidelity Cas9 domains

[0230] Some aspects of the disclosure provide high fidelity Cas9 domains. In some embodiments, high fidelity Cas9 domains are engineered Cas9 domains comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and a sugar-phosphate backbone of a DNA, as compared to a corresponding wild-type Cas9 domain. Without wishing to be bound by any particular theory, high fidelity Cas9 domains that have decreased electrostatic interactions with a sugar-phosphate backbone of DNA may have less off-target effects. In some embodiments, a Cas9 domain (*e.g.*, a wild type Cas9 domain) comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate

backbone of a DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70%.

[0231] In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497X, a R661X, a Q695X, and/or a Q926X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497A, a R661A, a Q695A, and/or a Q926A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the Cas9 domain comprises a D10A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. Cas9 domains with high fidelity are known in the art and would be apparent to the skilled artisan. For example, Cas9 domains with high fidelity have been described in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); the entire contents of each are incorporated herein by reference.

High Fidelity Cas9 domain mutations relative to Cas9 are shown in bold and underline

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLL
NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDEN
DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE

SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD.

[0232] In some cases, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA or RNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some cases, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some cases, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such cases, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some cases, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

[0233] In some embodiments, a CRISPR protein-derived domain of a base editor can comprise all or a portion of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM sequence. Such sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., et al., "Engineered CRISPR-Cas9 nucleases with altered PAM

specificities" Nature 523, 481-485 (2015); and Kleinstiver, B. P., et al., "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference.

[0234] In some examples, a PAM recognized by a CRISPR protein-derived domain of a base editor disclosed herein can be provided to a cell on a separate oligonucleotide to an insert (*e.g.* an AAV insert) encoding the base editor. In such cases, providing PAM on a separate oligonucleotide can allow cleavage of a target sequence that otherwise would not be able to be cleaved, because no adjacent PAM is present on the same polynucleotide as the target sequence.

[0235] In an embodiment, *S. pyogenes* Cas9 (SpCas9) can be used as a CRISPR endonuclease for genome engineering. However, others can be used. In some cases, a different endonuclease can be used to target certain genomic targets. In some cases, synthetic SpCas9-derived variants with non-NGG PAM sequences can be used. Additionally, other Cas9 orthologues from various species have been identified and these "non-SpCas9s" can bind a variety of PAM sequences that can also be useful for the present disclosure. For example, the relatively large size of SpCas9 (approximately 4kb coding sequence) can lead to plasmids carrying the SpCas9 cDNA that cannot be efficiently expressed in a cell. Conversely, the coding sequence for *Staphylococcus aureus* Cas9 (SaCas9) is approximately 1 kilobase shorter than SpCas9, possibly allowing it to be efficiently expressed in a cell. Similar to SpCas9, the SaCas9 endonuclease is capable of modifying target genes in mammalian cells *in vitro* and in mice *in vivo*. In some cases, a Cas protein can target a different PAM sequence. In some cases, a target gene can be adjacent to a Cas9 PAM, 5'-NGG, for example. In other cases, other Cas9 orthologs can have different PAM requirements. For example, other PAMs such as those of *S. thermophilus* (5'-NNAGAA for CRISPR1 and 5'-NGGNG for CRISPR3) and *Neisseria meningiditis* (5'-NNNNGATT) can also be found adjacent to a target gene.

[0236] In some embodiments, for a *S. pyogenes* system, a target gene sequence can precede (*i.e.*, be 5' to) a 5'-NGG PAM, and a 20-nt guide RNA sequence can base pair with an opposite strand to mediate a Cas9 cleavage adjacent to a PAM. In some cases, an adjacent cut can be or can be about 3 base pairs upstream of a PAM. In some cases, an adjacent cut can be or can be about 10 base pairs upstream of a PAM. In some cases, an adjacent cut can be or can be about 0-20 base pairs upstream of a PAM. For example, an adjacent cut can be next to, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 base pairs upstream of a PAM. An adjacent cut can also be downstream of a PAM by 1 to 30 base pairs.

Fusion proteins comprising a nuclear localization sequence (NLS)

In some embodiments, the fusion proteins provided herein further comprise one or more (e.g., 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In one embodiment, a bipartite NLS is used. In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.,* by nuclear transport). In some embodiments, any of the fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus of the deaminase. In some embodiments, the NLS is fused to the C-terminus of the deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises the amino acid sequence PKKKRKVEGADKRTADGSEFES PKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV, or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC. In some embodiments, the NLS is present in a linker or the NLS is flanked by linkers, for example, the linkers described herein. In some embodiments, the N-terminus or C-terminus NLS is a bipartite NLS. A bipartite NLS comprises two basic amino acid clusters, which are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while monopartite NLSs are not). The NLS of nucleoplasmin, KR[PAATKKAGQA]KKKK, is the prototype of the ubiquitous bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids. The sequence of an exemplary bipartite NLS follows: PKKKRKVEGADKRTADGSEFES PKKKRKV.

[0237] In some embodiments, the fusion proteins of the invention do not comprise a linker sequence. In some embodiments, linker sequences between one or more of the domains or proteins are present.

[0238] It should be appreciated that the fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein may comprise inhibitors, cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

[0239] A vector that encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSs) can be used. For example, there can be or be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs used. A CRISPR enzyme can comprise the NLSs at or near the ammo-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs at or near the carboxy-terminus, or any combination of these (*e.g.*, one or more NLS at the ammo-terminus and one or more NLS at the carboxy terminus). When more than one NLS is present, each can be selected independently of others, such that a single NLS can be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies.

[0240] CRISPR enzymes used in the methods can comprise about 6 NLSs. An NLS is considered near the N- or C-terminus when the nearest amino acid to the NLS is within about 50 amino acids along a polypeptide chain from the N- or C-terminus, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, or 50 amino acids.

[0241] The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

Cas9 Domains with Reduced Exclusivity

[0242] Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenosine (A), thymidine (T), or cytosine (C), and the G is guanosine. This may limit the ability

to edit desired bases within a genome. In some embodiments, the base editing fusion proteins provided herein may need to be placed at a precise location, for example a region comprising a target base that is upstream of the PAM. See e.g., Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.,* NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); Nishimasu, H., *et al.*, "Engineered CRISPR-Cas9 nuclease with expanded targeting space" *Science*. 2018 Sep 21;361(6408):1259-1262, Chatterjee, P., *et al.*, Minimal PAM specificity of a highly similar SpCas9 ortholog" *Sci Adv*. 2018 Oct 24;4(10):eaau0766. doi: 10.1126/sciadv.aau0766, the entire contents of each are hereby incorporated by reference. Several PAM variants are described in Table 1 below.

Table 1. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
SpyMacCas9	NAA
Cpf1	5' (TTTV)

Nucleobase Editing Domain

[0243] Described herein are base editors comprising a fusion protein that includes a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, deaminase domain). The base editor can be programmed to edit one or more bases in a target polynucleotide sequence by interacting with a guide polynucleotide capable of recognizing the target sequence. Once the target sequence has been recognized, the base editor is anchored on the polynucleotide where editing is to occur and the deaminase domain component of the base editor can then edit a target base.

[0244] In some embodiments, the nucleobase editing domain is a deaminase domain. In some cases, a deaminase domain can be a cytosine deaminase or a cytidine deaminase. In some embodiments, the terms "cytosine deaminase" and "cytidine deaminase" can be used interchangeably. In some cases, a deaminase domain can be an adenine deaminase or an adenosine deaminase. In some embodiments, the terms "adenine deaminase" and "adenosine deaminase" can be used interchangeably. Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

C to T Editing

[0245] In some embodiments, a base editor disclosed herein comprises a fusion protein comprising cytidine deaminase capable of deaminating a target cytidine (C) base of a polynucleotide to produce uridine (U), which has the base pairing properties of thymine. In some embodiments, for example where the polynucleotide is double-stranded (*e.g.* DNA), the uridine base can then be substituted with a thymidine base (*e.g.* by cellular repair machinery) to give rise to a C:G to a T:A transition. In other embodiments, deamination of a C to U in a nucleic acid by a base editor cannot be accompanied by substitution of the U to a T.

[0246] The deamination of a target C in a polynucleotide to give rise to a U is a non-limiting example of a type of base editing that can be executed by a base editor described herein. In another example, a base editor comprising a cytidine deaminase domain can mediate conversion of a cytosine (C) base to a guanine (G) base. For example, a U of a polynucleotide produced by deamination of a cytidine by a cytidine deaminase domain of a base editor can be excised from the polynucleotide by a base excision repair mechanism (*e.g.*, by a uracil DNA glycosylase (UDG) domain), producing an abasic site. The nucleobase opposite the abasic site can then be substituted (*e.g.* by base repair machinery) with another base, such as a C, by for example a translesion polymerase. Although it is typical for a nucleobase opposite an abasic site to be replaced with a C, other substitutions (*e.g.* A, G or T) can also occur.

[0247] Accordingly, in some embodiments a base editor described herein comprises a deamination domain (*e.g.*, cytidine deaminase domain) capable of deaminating a target C to a U in a polynucleotide. Further, as described below, the base editor can comprise additional domains which facilitate conversion of the U resulting from deamination to, in some embodiments, a T or a G. For example, a base editor comprising a cytidine deaminase domain can further comprise a uracil glycosylase inhibitor (UGI) domain to mediate substitution of a U by a T, completing a C-to-T base editing event. In another example, a base editor can incorporate a translesion polymerase to improve the efficiency of C-to-G base editing, since a translesion polymerase can facilitate incorporation of a C opposite an abasic site (*i.e.*, resulting in incorporation of a G at the abasic site, completing the C-to-G base editing event).

[0248] A base editor comprising a cytidine deaminase as a domain can deaminate a target C in any polynucleotide, including DNA, RNA and DNA-RNA hybrids. Typically, a cytidine deaminase catalyzes a C nucleobase that is positioned in the context of a single-stranded portion of a polynucleotide. In some embodiments, the entire polynucleotide comprising a target C can be single-stranded. For example, a cytidine deaminase incorporated into the base editor can deaminate a target C in a single-stranded RNA polynucleotide. In other embodiments, a base editor comprising a cytidine deaminase domain can act on a double-stranded polynucleotide, but the target C can be positioned in a portion of the polynucleotide which at the time of the deamination reaction is in a single-stranded state. For example, in embodiments where the NAGPB domain comprises a Cas9 domain, several nucleotides can be left unpaired during formation of the Cas9-gRNA-target DNA complex, resulting in formation of a Cas9 "R-loop complex". These unpaired nucleotides can form a bubble of single-stranded DNA that can serve as a substrate for a single-strand specific nucleotide deaminase enzyme (*e.g.*, cytidine deaminase).

[0249] In some embodiments, a cytidine deaminase of a base editor can comprise all or a portion of an apolipoprotein B mRNA editing complex (APOBEC) family deaminase. Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) is a family of evolutionarily conserved cytidine deaminases. Members of this family are C-to-U editing enzymes. The N-terminal domain of APOBEC like proteins is the catalytic domain, while the C-terminal domain is a pseudocatalytic domain. More specifically, the catalytic domain is a zinc dependent cytidine deaminase domain and is important for cytidine deamination. APOBEC family members include APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D ("APOBEC3E" now refers to this), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and Activation-induced (cytidine) deaminase. A number of modified cytidine deaminases are commercially available, including but not limited to SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3, which are available from Addgene (plasmids 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177). In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC1 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC2 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of is an APOBEC3 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC3A deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3B deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3C deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3D deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3E deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3F deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3G deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3H deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC4 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of activation-induced deaminase (AID).

[0250] In some embodiments a deaminase incorporated into a base editor comprises all or a portion of cytidine deaminase 1 (CDA1). It will be appreciated that a base editor can comprise a deaminase from any suitable organism (*e.g.*, a human or a rat). In some embodiments, a

deaminase domain of a base editor is from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase domain of the base editor is derived from rat (*e.g.*, rat APOBEC1). In some embodiments, the deaminase domain of the base editor is human APOBEC1. In some embodiments, the deaminase domain of the base editor is pmCDA1.

[0251] The base sequence and amino acid sequence of PmCDA1 and the base sequence and amino acid sequence of CDS of human AID are shown herein below.

[0252] >tr|A5H718|A5H718_PETMA Cytosine deaminase OS=Petromyzon marinus OX=7757 PE=2 SV=1

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSG
TERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLK
IWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKT
LKRAEKRRSELSIMIQVKILHTTKSPAV

>EF094822.1 Petromyzon marinus isolate PmCDA.21 cytosine deaminase mRNA, complete cds

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGGA
CATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGA
GAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCA
TAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTG
AATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAAT
ACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGC
TGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGC
AAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGT
TGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATT
GAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATG
ATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

>tr|Q6QJ80|Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens OX=9606 GN=AICDA PE=2 SV=1

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
AEPEGLRRLHRAGVQIAIMTFKAPV

>NG_011588.1:5001-15681 Homo sapiens activation induced cytidine deaminase (AICDA), RefSeqGene (LRG_17) on chromosome 12

AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCTG
GACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAA
ATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGAT

CAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAAT
TCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTA
CTGATTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTT
TAGAGACTCAAATCATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAA
TTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTT
AGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTAT
GATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGG
CAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTAGG
TGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTTA
TTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAAT
TTAACTCCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACA
GGGTCTTAGCCCAGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAA
GTGCTGGAATTATAGACATGAGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAAT
GTTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAA
ACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAA
AGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAATTCAACAGGACAGAA
GGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCA
GGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAG
CACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTACTCAG
GACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACAA
AGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTA
GGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCT
TCTTATCTATACTTCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACAC
ACACACACACACACACACACACACACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGA
TGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGT
AAGAGATGATGCTTAATCTGTTTAACACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAA
GCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGG
TTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTT
TTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTG
GAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCCTTGG
CCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATATG
CATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGT
TTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAAC
AGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTT
GAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGG
CATGGTGGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGG
TCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCA
AAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACT
TAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAAT
ATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAACC
CCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATATCCATGCAA
TGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAAG

AGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTAC
AGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAA
GATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGG
GAAAAAAAGAAAGCAAATCAGTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTG
GTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAA
AATATTCGTAGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCTCAATGTAAGAA
AAAATAATGTGTAAGAAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACT
TCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGC
ATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTACAGCT
TGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCAGAGTATTT
CCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGG
AGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGC
AGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACT
AGAGGAATAAACACAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACA
CTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTAT
TCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTT
TGGTCTTGTTGCCCATGCTGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTC
AAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATG
TTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGG
ATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAAGTATTGC
TCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATACATTAAAAAATA
GGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGGT
CAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTAAAAATACAAACATTACCTGG
GCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCA
GATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCG
ACAAAGTGAGACCGTAACAAAAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAA
AGTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCT
TCAGCATGGGAATGGCATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTAT
TTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGG
GCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCA
ACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTTCCAGGCTG
CTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTTGAGGAATA
TTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAAT
ATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTT
AATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTT
GGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCA
GGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCC
AAGTAATGACTTCCTTAAAAAGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATT
TTCTCTCCCTCCTCTCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTC
CGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCT
TTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAATGGT
CAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGGCATTTGTGTCTCT

ATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACATGGCCCAA
AATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGAA
CAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAA
GCATTTTTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGT
CTTAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCC
AAATGATTTCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCA
GAAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTT
CCAAGAGACTTCTTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCC
TAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTG
GCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATCCCAATC
TCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGCACTTTCATTTAC
TTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTATTTCCCCT
TTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCACA
ATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACA
AATTAAATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCC
ACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTG
TGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATC
ATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTC
TGTCTCTCCAGAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCC
TGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCC
GACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACC
GCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAA
AGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGGGATTGCGATGCGGAATGAAT
GAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAAA
GATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAAAATGGGGCCAGG
GTTGCTTCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTT
TTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCC
TGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCT
TCCTCGCTTTTTAAATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTC
TTATTGTTCAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTC
TGCTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGTTTCA
CATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTCTCCTTTTTTT
TTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTTCCCAATTTACTTTCTT
CCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCCAACCTCGTTTT
GAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGACG
CATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG
ACAGTGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTT
ACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTC
TGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACT
GCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGT
AGATCCTAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTG
AGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAA

CTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAACCAT
CTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCATGTTTTTATGTTTGT
ACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCTACTTTAAT
AAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAGC
AAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAACTTCACATATC
ATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGG
TCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCA
AAGGTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCT
CATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGG
GCAACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTG
GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGTTTGAGA
CCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGGCGTGGTAGCAGG
CACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCA
GTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGAAAAAAAAA
AAAAAAAGAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAAT
TGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGT
CCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAG
CAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTA
TTGTACATAAGTTTGTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTG
CGTCTAATGATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAA
ATTTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTACGAA
GCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTTCAATCAGTTA
AATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAAATAATATAAAAGTGATTTAT
ATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[0253] Other exemplary deaminases that can be fused to Cas9 according to aspects of this disclosure are provided below. It should be understood that, in some embodiments, the active domain of the respective sequence can be used, e.g., the domain without a localizing signal (nuclear localization sequence, without nuclear export signal, cytoplasmic localizing signal).

Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYF
CEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLS
RQLRRILLPLYEVDDLRDAFRTLGL

(underline: nuclear localization sequence; double underline: nuclear export signal)

Mouse AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYF
CEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLT

RQLRRILLPLYEVDDLRDAFRMLGF

(underline: nuclear localization sequence; double underline: nuclear export signal)

Canine AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFC
EDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSR
QLRRILLPLYEVDDLRDAFRTLGL

(underline: nuclear localization sequence; double underline: nuclear export signal)

Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHV
ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC
DKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLS
RQLRRILLPLYEVDDLRDAFRTLGL

(underline: nuclear localization sequence; double underline: nuclear export signal)

Rat AID

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLL
MKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALP
AGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLR
DAFRTLGL

(underline: nuclear localization sequence; double underline: nuclear export signal)

Mouse APOBEC-3

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFR
PWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSE
EEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKI*
*RSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL
WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDL
VNDFGNLQLGPPMS

(italic: nucleic acid editing domain)

Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSL HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATH HNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFR PWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSE EEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKI RSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDL VNDFGNLQLGPPMS

(italic: nucleic acid editing domain)

Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY*H PEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVARLY YFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKH YTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQH RGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQEMAKFIS NNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPF QPWDGLDEHSQALSGRLRAI

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTI FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW NNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVL LNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSC*AQE MAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDH QGCPFQPWDGLEEHSQALSGRLRAILQNQGN

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLY PEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIF VARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRK

NLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWV
LLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSC*AQK
MAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVD
RQGRPFQPWDGLDEHSQALSGRLRAI
(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Human APOBEC-3G:
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQV
YSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTI
FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPW
NNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWV
LLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC*AQ
EMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVD
HQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN
(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

Human APOBEC-3F:
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQV
YSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTIS
AARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFD
DNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVS
WKRGVFRNQVDPETH*CHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEF
LARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYND
DEPFKPWKGLKYNFLFLDSKLQEILE
(italic: nucleic acid editing domain)

Human APOBEC-3B:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQ
VYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTI
SAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKF
DENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMD
QHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGE
VRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVY
RQGCPFQPWDGLEEHSQALSGRLRAILQNQGN
(italic: nucleic acid editing domain)

Rat APOBEC-3B:
MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNF
LCYEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYM
SWSPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMD
LPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQFNNS
HRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELS
QVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIH
VDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL

Bovine APOBEC-3B:
DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFK
QQFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLN
PSQSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNA
GISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI

Chimpanzee APOBEC-3B:
MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRG
QMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNV
TLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPW
YKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDNGTWV
LMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSW
GCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYC
WDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLP
LCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQP
PCSSRIRETEGWASVSKEGRDLG

Human APOBEC-3C:
MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRN
QVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLT
IFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLK
TNFRLLKRRLRESLQ
(italic: nucleic acid editing domain)

Gorilla APOBEC-3C3C
MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRN
QVDSETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTI

FTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLK YNFRFLKRRLQEILE

Human APOBEC-3A:
MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQ AKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENT HVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQP WDGLDEHSQALSGRLRAILQNQGN
(italic: nucleic acid editing domain)

Rhesus macaque APOBEC-3A:
MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGF LCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVRVF LQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGR PFQPWDGLDEHSQALSGRLRAILQNQGN
(italic: nucleic acid editing domain)

Bovine APOBEC-3A3A:
MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAE LYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFG CHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQA ILKTQQN
(italic: nucleic acid editing domain)

Human APOBEC-3H:
MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICF INEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQ QKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRL ERIKIPGVRAQGRYMDILCDAEV
(italic: nucleic acid editing domain)

Rhesus macaque APOBEC-3H:
MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIR FINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRP NYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRR LERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR

Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGP
VLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPC*LPCVVKVT
KFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVC
NEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLC
FTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSP*
*CPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFV
SCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

(italic: nucleic acid editing domain)

Human APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARL
FWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWM
MLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR

Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTS
NHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLY
HHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLY
VLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK

Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL
ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

Human APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNV
EYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRY
NVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKL
RIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

Mouse APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNV
EYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKY
NVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCK
LRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

Rat APOBEC-2:
MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNV
EYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKY
NVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCK
LRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

Bovine APOBEC-2:
MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNV
EYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRY
MVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCR
LRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

Petromyzon marinus CDA1 (pmCDAl)
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNK
PQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRG
NGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ
LNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV

Human APOBEC3G D316R D317R
MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDP
KVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKF NYDEFQHCWSKFVYSQ
RELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVER
MHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTC
FTSWSPCFSCAQEMAKFISK KHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISF T
YSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

Human APOBEC3G chain A
MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHG
FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCI

FTARIYDDQGRCQEGLRTLAEAGAKISF TYSEFKHCWDTFVDHQGCPFQPWDGLD EHSQDLSGRLRAILQ

Human APOBEC3G chain A D120R D121R
MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHG FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCI FTARIYRRQGRCQEGLRTLAEAGAKISFMTYSEFKHCWDTFVDHQGCPFQPWDGLDE HSQDLSGRLRAILQ

[0254] Some aspects of the present disclosure are based on the recognition that modulating the deaminase domain catalytic activity of any of the fusion proteins described herein, for example by making point mutations in the deaminase domain, affect the processivity of the fusion proteins (*e.g.*, base editors). For example, mutations that reduce, but do not eliminate, the catalytic activity of a deaminase domain within a base editing fusion protein can make it less likely that the deaminase domain will catalyze the deamination of a residue adjacent to a target residue, thereby narrowing the deamination window. The ability to narrow the deamination window can prevent unwanted deamination of residues adjacent to specific target residues, which can decrease or prevent off-target effects.

[0255] For example, in some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of H121X, H122X, R126X, R126X, R118X, W90X, W90X, and R132X of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase, wherein X is any amino acid. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of H121R, H122R, R126A, R126E, R118A, W90A, W90Y, and R132E of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase.

[0256] In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise one or more mutations selected from the group consisting of D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase, wherein X is any amino acid. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising one or more mutations selected from the group consisting of D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase.

[0257] In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise a H121R and a H122R mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R118A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90A mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y and a R126E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R126E and a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y and a R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W90Y, R126E, and R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase.

[0258] In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a D316R and a D317R mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, any of the fusion proteins provided herein comprise an APOBEC deaminase comprising a R320A mutation of hAPOBEC3G, or one or more corresponding mutations in

another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R320E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R313A mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285A mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y and a R320E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a R320E and a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y and a R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase. In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise an APOBEC deaminase comprising a W285Y, R320E, and R326E mutation of hAPOBEC3G, or one or more corresponding mutations in another APOBEC deaminase.

[0259] A number of modified cytidine deaminases are commercially available, including but not limited to SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3, which are available from Addgene (plasmids 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177).

[0260] Other exemplary deaminases that can be fused to Cas9 according to aspects of this disclosure are provided below. It should be understood that, in some embodiments, the active domain of the respective sequence can be used, e.g., the domain without a localizing signal (nuclear localization sequence, without nuclear export signal, cytoplasmic localizing signal).

[0261] Details of C to T nucleobase editing proteins are described in International PCT Application No. PCT/US2016/058344 (WO 2017/070632) and Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference.

A to G Editing

[0262] In some embodiments, a base editor described herein can comprise a deaminase domain which includes an adenosine deaminase. Such an adenosine deaminase domain of a base editor can facilitate the editing of an adenine (A) nucleobase to a guanine (G) nucleobase by deaminating the A to form inosine (I), which exhibits base pairing properties of G. Adenosine deaminase is capable of deaminating (*i.e.*, removing an amine group) adenine of a deoxyadenosine residue in deoxyribonucleic acid (DNA).

[0263] In some embodiments, the nucleobase editors provided herein can be made by fusing together one or more protein domains, thereby generating a fusion protein. In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity (*e.g.*, efficiency, selectivity, and specificity) of the fusion proteins. For example, the fusion proteins provided herein can comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, the fusion proteins provided herein can have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). Without wishing to be bound by any particular theory, the presence of the catalytic residue (*e.g.*, H840) maintains the activity of the Cas9 to cleave the non-edited (*e.g.*, non-deaminated) strand containing a T opposite the targeted A. Mutation of the catalytic residue (*e.g.*, D10 to A10) of Cas9 prevents cleavage of the edited strand containing the targeted A residue. Such Cas9 variants are able to generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a T to C change on the non-edited strand. In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine residue (*e.g.*, inosine), which can improve the activity or efficiency of the base editor.

[0264] A base editor comprising an adenosine deaminase can act on any polynucleotide, including DNA, RNA and DNA-RNA hybrids. In certain embodiments, a base editor comprising an adenosine deaminase can deaminate a target A of a polynucleotide comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR, *e.g.*, ADAR1 or ADAR2). In another embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from *Escherichia coli* (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, E155V, L84F, H123Y, I157F, or a corresponding mutation in another adenosine deaminase.

[0265] The adenosine deaminase can be derived from any suitable organism. In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli, Staphylococcus aureus, Salmonella typhi, Shewanella putrefaciens, Haemophilus influenzae, Caulobacter crescentus,* or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (*e.g.*, mutations in ecTadA). The corresponding residue in any homologous protein can be identified by *e.g.*, sequence alignment and determination of homologous residues. The mutations in any naturally-occurring adenosine deaminase (*e.g.*, having homology to ecTadA) that corresponds to any of the mutations described herein (*e.g.*, any of the mutations identified in ecTadA) can be generated accordingly.

TadA

[0266] In particular embodiments, the TadA is any one of the TadA described in PCT/US2017/045381 (WO 2018/027078), which is incorporated herein by reference in its entirety.

[0267] In one embodiment, a fusion protein of the invention comprises a wild-type TadA linked to TadA7.10, which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA7.10 domain (e.g., provided as a monomer). In other embodiments, the ABE7.10 editor comprises TadA7.10 and TadA(wt), which are capable of forming heterodimers. The relevant sequences follow:
SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHA EIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD, which is termed "the TadA reference sequence" or wild type TadA (TadA(wt)).
TadA7.10:
SEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD

[0268] In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (e.g., any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein..

[0269] In some embodiments the TadA deaminase is a full-length *E. coli* TadA deaminase. For example, in certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWN
RPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVF
GARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEI
KAQKKAQSSTD

[0270] It will be appreciated, however, that additional adenosine deaminases useful in the present application would be apparent to the skilled artisan and are within the scope of this disclosure. For example, the adenosine deaminase may be a homolog of adenosine deaminase acting on tRNA (AD AT). Exemplary AD AT homologs include, without limitation:

Staphylococcus aureus TadA:
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAE
HIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCS GS
LMNLLQQS NFNHRAIVDKG VLKE AC S TLLTTFFKNLRANKKS TN

Bacillus subtilis TadA:
MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML
VIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGC S
GTLMN LLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

Salmonella typhimurium (*S. typhimurium*) TadA:
MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWN
RPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVF
GARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIK
ALKKADRAEGAGPAV

Shewanella putrefaciens (*S. putrefaciens*) TadA:
MDE YWMQVAMQM AEKAEAAGE VPVGA VLVKDGQQIATGYNLS IS QHDPT
AHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGA
AGTVVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

Haemophilus influenzae F3031 (*H. influenzae*) TadA:
MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDP
TAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHS RIKRLVFG
ASDYKTGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKS
LSDK

Caulobacter crescentus (*C. crescentus*) TadA:
MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDP
TAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADD
PKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:
MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDP
SAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGG
AAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALF
IDERKVPPEP.

[0271] In some embodiments, the adenosine deaminase comprises a D108X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108G, D108N, D108V, D108A, or D108Y mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase. It should be appreciated, however, that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein.

[0272] In some embodiments, the adenosine deaminase comprises an A106X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0273] In some embodiments, the adenosine deaminase comprises a E155X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a E155D, E155G, or E155V mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0274] In some embodiments, the adenosine deaminase comprises a D147X mutation, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D147Y, mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0275] It should be appreciated that any of the mutations provided herein (e.g., TadA reference sequence) may be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (e.g., bacterial adenosine deaminases). It would be apparent to the skilled artisan how to identify sequences that are

homologous to the mutated residues relative to the TadA reference sequence. Thus, any of the mutations identified relative to the TadA reference sequence may be made in other adenosine deaminases that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein may be made individually or in any combination relative to the TadA reference sequence or another adenosine deaminase. For example, an adenosine deaminase may contain a D108N, a A106V, a E155V, and/or a D147Y mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase. In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") in TadA reference sequence, or corresponding mutations in another adenosine deaminase: D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V; A106V and D147Y; E155V and D147Y; D108N, A106V, and E55V; D108N, A106V, and D147Y; D108N, E55V, and D147Y; A106V, E55V, and D 147Y; and D108N, A106V, E55V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein may be made in an adenosine deaminase (e.g., wild type TadA or ecTadA).

[0276] In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, Kl lOX, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or E85G, M94L, 1951, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, Kl 101, Ml18K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0277] In some embodiments, the adenosine deaminase comprises one or more of H8X, D108X, and/or N127X mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid. In some embodiments, the adenosine deaminase comprises one or more of a H8Y, D108N, and/or N127S mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0278] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X, R152X, and Q154X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, M61X, M70X, D108X, N127X, Q154X, E155X, and Q163X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, N127X, E155X, and T166X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, D108X, mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, R126X, L68X, D108X, N127X, D147X, and E155X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, A109X, N127X, and E155X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

[0279] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, D108N, N127S, D147Y, R152C, and Q154H relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R, E155G and Q163H relative to the TadA reference

sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R126W, L68Q, D108N, N127S, D147Y, and E155V relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, A109T, N127S, and E155G relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

[0280] In some embodiments, the adenosine deaminase comprises one or more of the or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, R24W, D108N, N127S, D147Y, and E155V mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and S127S mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y and E155V

mutation relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase.

[0281] In some embodiments, the adenosine deaminase comprises one or more of a, S2X, H8X, I49X, L84X, H123X, N127X, I156X and/or K160X mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of S2A, H8Y, I49F, L84F, H123Y, N127S, I156F and/or K160S mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0282] In some embodiments, the adenosine deaminase comprises an L84X mutation adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an L84F mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0283] In some embodiments, the adenosine deaminase comprises an H123X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H123Y mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0284] In some embodiments, the adenosine deaminase comprises an I157X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an I157F mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0285] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84X, A106X, D108X, H123X, D147X, E155X, and I156X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2X, I49X, A106X, D108X, D147X, and

E155X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, A106X, D108X, N127X, and K160X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

[0286] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84F, A106V, D108N, H123Y, D147Y, E155V, and I156F relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2A, I49F, A106V, D108N, D147Y, and E155V relative to the TadA reference sequence.

[0287] In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, A106T, D108N, N127S, and K160S relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

[0288] In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R07K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q and/or A143R mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein corresponding to TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0289] In some embodiments, the adenosine deaminase comprises an E25X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type

adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M, E25D, E25A, E25R, E25V, E25S, or E25Y mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0290] In some embodiments, the adenosine deaminase comprises an R26X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0291] In some embodiments, the adenosine deaminase comprises an R107X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R107P, R07K, R107A, R107N, R107W, R107H, or R107S mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0292] In some embodiments, the adenosine deaminase comprises an A142X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A142N, A142D, A142G, mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0293] In some embodiments, the adenosine deaminase comprises an A143X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q and/or A143R mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0294] In some embodiments, the adenosine deaminase comprises one or more of a H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, and/or K161X mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N,

and/or K161T mutation relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0295] In some embodiments, the adenosine deaminase comprises an H36X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H36L mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0296] In some embodiments, the adenosine deaminase comprises an N37X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an N37T, or N37S mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0297] In some embodiments, the adenosine deaminase comprises an P48X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an P48T, or P48L mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0298] In some embodiments, the adenosine deaminase comprises an R51X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R51H, or R51L mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0299] In some embodiments, the adenosine deaminase comprises an S146X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an S146R, or S146C mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0300] In some embodiments, the adenosine deaminase comprises an K157X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a K157N mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0301] In some embodiments, the adenosine deaminase comprises an P48X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a P48S, P48T, or P48A mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0302] In some embodiments, the adenosine deaminase comprises an A142X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A142N mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0303] In some embodiments, the adenosine deaminase comprises an W23X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a W23R, or W23L mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0304] In some embodiments, the adenosine deaminase comprises an R152X mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a R152P, or R152H mutation relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

[0305] In one embodiment, the adenosine deaminase may comprise the mutations H36L, R51L, L84F, A106V, D108N, H123Y, S 146C, D147Y, E155V, I156F, and K157N. In some embodiments, the adenosine deaminase comprises the following combination of mutations relative to TadA reference sequence, where each mutation of a combination is separated by a "_" and each combination of mutations is between parentheses:

(A106V_D108N), (R107C_D108N),

(H8Y_D108N_S127S_D147Y_Q154H), (H8Y_R24W_D108N_N127S_D147Y_E155V), (D108N_D147Y_E155V), (H8Y_D108N_S127S), (H8Y_D108N_N127S_D147Y_Q154H), (A106V_D108N_D147Y_E155V), (D108Q_D147Y_E155V) (D108M_D147Y_E155V), (D108L_D147Y_E155V), (D108K_D147Y_E155V), (D108I_D147Y_E155V), (D108F_D147Y_E155V), (A106V_D108N_D147Y), (A106V_D108M_D147Y_E155V), (E59A_A106V_D108N_D147Y_E155V), (E59A cat dead_A106V_D108N_D147Y_E155V), (L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y), (L84F_A106V_D108N_H123Y_D147Y_E155V_I156F), (D103A_D104N), (G22P_D103A_D104N), (G22P_D103A_D104N_S138A), (D103A_D104N_S138A), (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F), (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F), (L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F), (R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F), (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (A106V_D108N_A142N_D147Y_E155V), (R26G_A106V_D108N_A142N_D147Y_E155V), (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V), (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V), (E25D_R26G_A106V_D108N_A142N_D147Y_E155V), (A106V_R107K_D108N_A142N_D147Y_E155V), (A106V_D108N_A142N_A143G_D147Y_E155V), (A106V_D108N_A142N_A143L_D147Y_E155V), (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),

(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N),
(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E), (R74Q
L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),

(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_ I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F _K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P _E155V_I156F _K157N),

(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F _K161T),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F _K157N).

[0306] In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity of the fusion proteins. For example, any of the fusion proteins provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the fusion proteins provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9).

[0307] In some embodiments, the adenosine deaminase comprises a D108X mutation inthe TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108G, D108N, D108V, D108A, or D108Y mutation, or a corresponding mutation in another adenosine deaminase.

[0308] In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, relative to the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E155D, E155G, or E155V mutation. In some embodiments, the adenosine deaminase comprises a D147Y.

[0309] It should be appreciated that any of the mutations provided herein (*e.g.*, based on the the TadA reference sequence amino acid sequence) can be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.*, bacterial adenosine deaminases). Any of the mutations identified relative to the TadA reference sequencecan be made in other adenosine deaminases that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein can be made individually or in any combination relative to the TadA reference sequence or another adenosine deaminase.

[0310] For example, an adenosine deaminase can contain a D108N, a A106V, a E155V, and/or a D147Y , or a corresponding mutation in another adenosine deaminase. In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") relative to the TadA reference sequence, or corresponding mutations in another adenosine deaminase: D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V; A106V and D147Y; E155V and D147Y; D108N, A106V, and E55V; D108N, A106V, and D147Y; D108N, E55V, and D147Y; A106V, E55V, and D 147Y; and D108N, A106V, E55V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein can be made in an adenosine deaminase (*e.g.*, the TadA reference sequence or ecTadA).

[0311] In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or E85G, M94L, 1951, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, K110I, Ml18K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of a H8X, D108X, and/or N127X relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid. In some embodiments, the adenosine deaminase comprises one or more of a H8Y, D108N, and/or N127S relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0312] In some embodiments, the adenosine deaminase comprises one or more of H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X, and/or T166X relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H or Q154R, E155G or

E155V or E155D, K161Q, Q163H, and/or T166P relative to the TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

[0313] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X, R152X, and Q154X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, M61X, M70X, D108X, N127X, Q154X, E155X, and Q163X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, N127X, E155X, and T166X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

[0314] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, D108X, mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, R126X, L68X, D108X, N127X, D147X, and E155X, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, A109X, N127X, and E155X relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

[0315] In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, D108N, N127S, D147Y, R152C, and Q154H relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises

one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R, E155G and Q163H relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R126W, L68Q, D108N, N127S, D147Y, and E155V relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, A109T, N127S, and E155G relative to the TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

[0316] Any of the mutations provided herein and any additional mutations (*e.g.*, based on the the TadA reference sequenceamino acid sequence) can be introduced into any other adenosine deaminases. Any of the mutations provided herein can be made individually or in any combination relative to the TadA reference sequenceor another adenosine deaminase.

[0317] Details of A to G nucleobase editing proteins are described in International PCT Application No. PCT/2017/045381 (WO 2018/027078) and Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017), the entire contents of which are hereby incorporated by reference.

Cytidine deaminase

[0318] In one embodiment, a fusion protein of the invention comprises a cytidine deaminase. In some embodiments, the cytidine deaminases provided herein are capable of deaminating cytosine or 5-methylcytosine to uracil or thymine. In some embodiments, the cytosine deaminases provided herein are capable of deaminating cytosine in DNA. The cytidine deaminase may be derived from any suitable organism. In some embodiments, the cytidine deaminase is a naturally-occurring cytidine deaminase that includes one or more mutations corresponding to any of the mutations provided herein. One of skill in the art will be able to

identify the corresponding residue in any homologous protein, e.g., by sequence alignment and determination of homologous residues. Accordingly, one of skill in the art would be able to generate mutations in any naturally-occurring cytidine deaminase that corresponds to any of the mutations described herein. In some embodiments, the cytidine deaminase is from a prokaryote. In some embodiments, the cytidine deaminase is from a bacterium. In some embodiments, the cytidine deaminase is from a mammal (e.g., human).

[0319] In some embodiments, the cytidine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the cytidine deaminase amino acid sequences set forth herein. It should be appreciated that cytidine deaminases provided herein may include one or more mutations (e.g., any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the cytidine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to a reference sequence, or any of the cytidine deaminases provided herein. In some embodiments, the cytidine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

Additional Domains

[0320] A base editor described herein can include any domain which helps to facilitate the nucleobase editing, modification or altering of a nucleobase of a polynucleotide. In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9), a nucleobase editing domain (*e.g.*, deaminase domain), and one or more additional domains. In some cases, the additional domain can facilitate enzymatic or catalytic functions of the base editor, binding functions of the base editor, or be inhibitors of cellular machinery (*e.g.*, enzymes) that could interfere with the desired base editing result. In some embodiments, a base editor can comprise a nuclease, a nickase, a recombinase, a deaminase, a

methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain.

[0321] In some embodiments, a base editor can comprise a uracil glycosylase inhibitor (UGI) domain. A UGI domain can for example improve the efficiency of base editors comprising a cytidine deaminase domain by inhibiting the conversion of a U formed by deamination of a C back to the C nucleobase. In some cases, cellular DNA repair response to the presence of U:G heteroduplex DNA can be responsible for a decrease in nucleobase editing efficiency in cells. In such cases, uracil DNA glycosylase (UDG) can catalyze removal of U from DNA in cells, which can initiate base excision repair (BER), mostly resulting in reversion of the U:G pair to a C:G pair. In such cases, BER can be inhibited in base editors comprising one or more domains that bind the single strand, block the edited base, inhibit UGI, inhibit BER, protect the edited base, and /or promote repairing of the non-edited strand. Thus, this disclosure contemplates a base editor fusion protein comprising a UGI domain.

[0322] In some embodiments, a base editor comprises as a domain all or a portion of a double-strand break (DSB) binding protein. For example, a DSB binding protein can include a Gam protein of bacteriophage Mu that can bind to the ends of DSBs and can protect them from degradation. See Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire content of which is hereby incorporated by reference.

[0323] In some embodiments, a base editor can comprise as a domain all or a portion of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion of a eukaryotic NAP. In some embodiments, a NAP or portion thereof incorporated into a base editor is a DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor has translesion polymerase activity. In some cases, a NAP or portion thereof incorporated into a base editor is a translesion DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor is a Rev7, Rev1 complex, polymerase iota, polymerase kappa, or polymerase eta. In some embodiments, a NAP or portion thereof incorporated into a base editor is a eukaryotic polymerase alpha, beta, gamma, delta, epsilon, gamma, eta, iota, kappa, lambda, mu, or nu component. In some embodiments, a NAP or portion thereof incorporated into a base editor comprises an amino acid sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to a nucleic acid polymerase (*e.g.*, a translesion DNA polymerase).

BASE EDITOR SYSTEM

[0324] Use of the base editor system provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide (*e.g.*, a double-stranded DNA or RNA, a single-stranded DNA or RNA) of a subject with a base editor system comprising a nucleobase editor (*e.g.*, an adenosine base editor or a cytidine base editor) and a guide polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence comprises a targeted nucleobase pair; (b) inducing strand separation of said target region; (c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase. It should be appreciated that in some embodiments, step (b) is omitted. In some embodiments, said targeted nucleobase pair is a plurality of nucleobase pairs in one or more genes. In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more genes, wherein at least one gene is located in a different locus.

[0325] In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9 domain. In some embodiments, the first base is adenine, and the second base is not a G, C, A, or T. In some embodiments, the second base is inosine.

[0326] Base editing system as provided herein provides a new approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a cytidine deaminase, and an inhibitor of base excision repair to induce programmable, single nucleotide (C→T or A→G) changes in DNA without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

[0327] Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a deaminase domain) for editing the nucleobase; and (2) a guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system comprises a cytosine base editor (CBE). In some embodiments, the base editor system comprises an adenosine base

editor (ABE). In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the nucleobase editing domain is a deaminase domain. In some cases, a deaminase domain can be a cytosine deaminase or a cytidine deaminase. In some embodiments, the terms "cytosine deaminase" and "cytidine deaminase" can be used interchangeably. In some cases, a deaminase domain can be an adenine deaminase or an adenosine deaminase. In some embodiments, the terms "adenine deaminase" and "adenosine deaminase" can be used interchangeably. Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

[0328] In some embodiments, the base editor inhibits base excision repair of the edited strand. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the base editor comprises UGI activity. In some embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edit of base pair is upstream of a PAM site. In some embodiments, the intended edit of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edit of base-pair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site.

[0329] In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker or a spacer. In some embodiments, the linker or spacer is 1-25 amino acids in length. In some embodiments, the linker or spacer is 5-20 amino acids in length. In some embodiments, the linker or spacer is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

[0330] In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1- 10 nucleotides. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edit of base pair is within the target window. In some embodiments, the target window comprises the intended edit of base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target window is a deamination window.

[0331] In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, non-limiting exemplary CBE is BE1 (APOBEC1-XTEN-dCas9), BE2 (APOBEC1-XTEN-dCas9-UGI), BE3 (APOBEC1-XTEN-dCas9(A840H)-UGI), BE3-Gam, saBE3, saBE4-Gam, BE4, BE4-Gam, saBE4, or saB4E-Gam. BE4 extends the APOBEC1-Cas9n(D10A) linker to 32 amino acids and the Cas9n-UGI linker to 9 amino acids, and appends a second copy of UGI to the C terminus of the construct with another 9-amino acid linker into a single base editor construct. The base editors saBE3 and saBE4 have the *S. pyogenes* Cas9n(D10A) replaced with the smaller *S. aureus* Cas9n(D10A). BE3-Gam, saBE3-Gam, BE4-Gam, and saBE4-Gam have 174 residues of Gam protein fused to the N-terminus of BE3, saBE3, BE4, and saBE4 via the 16 amino acid XTEN linker.

[0332] In some embodiments, the base editor is an adenosine base editor (ABE). In some embodiments, the adenosine base editor can deaminate adenine in DNA. In some embodiments, ABE is generated by replacing APOBEC1 component of BE3 with natural or engineered *E. coli* TadA, human ADAR2, mouse ADA, or human ADAT2. In some embodiments, ABE comprises evolved TadA variant. In some embodiments, the ABE is ABE 1.2 (TadA*-XTEN-nCas9-NLS). In some embodiments, TadA* comprises A106V and D108N mutations.

[0333] In some embodiments, the ABE is a second-generation ABE. In some embodiments, the ABE is ABE2.1, which comprises additional mutations D147Y and E155V in TadA* (TadA*2.1). In some embodiments, the ABE is ABE2.2, ABE2.1 fused to catalytically inactivated version of human alkyl adenine DNA glycosylase (AAG with E125Q mutation). In some embodiments, the ABE is ABE2.3, ABE2.1 fused to catalytically inactivated version of *E. coli* Endo V (inactivated with D35A mutation). In some embodiments, the ABE is ABE2.6 which has a linker twice as long (32 amino acids, $(SGGS)_2$-XTEN-$(SGGS)_2$) as the linker in ABE2.1. In some embodiments, the ABE is ABE2.7, which is ABE2.1 tethered with an additional wild-type TadA monomer. In some embodiments, the ABE is ABE2.8, which is ABE2.1 tethered with an additional TadA*2.1 monomer. In some embodiments, the ABE is

ABE2.9, which is a direct fusion of evolved TadA (TadA*2.1) to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.10, which is a direct fusion of wild type TadA to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.11, which is ABE2.9 with an inactivating E59A mutation at the N-terminus of TadA* monomer. In some embodiments, the ABE is ABE2.12, which is ABE2.9 with an inactivating E59A mutation in the internal TadA* monomer.

[0334] In some embodiments, the ABE is a third generation ABE. In some embodiments, the ABE is ABE3.1, which is ABE2.3 with three additional TadA mutations (L84F, H123Y, and I157F).

[0335] In some embodiments, the ABE is a fourth generation ABE. In some embodiments, the ABE is ABE4.3, which is ABE3.1 with an additional TadA mutation A142N (TadA*4.3).

[0336] In some embodiments, the ABE is a fifth generation ABE. In some embodiments, the ABE is ABE5.1, which is generated by importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1. In some embodiments, the ABE is ABE5.3, which has a heterodimeric construct containing wild-type E. coli TadA fused to an internal evolved TadA*. In some embodiments, the ABE is ABE5.2, ABE5.4, ABE5.5, ABE5.6, ABE5.7, ABE5.8, ABE5.9, ABE5.10, ABE5.11, ABE5.12, ABE5.13, or ABE5.14, as shown in below Table 2. In some embodiments, the ABE is a sixth generation ABE. In some embodiments, the ABE is ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5, or ABE6.6, as shown in below Table 2. In some embodiments, the ABE is a seventh generation ABE. In some embodiments, the ABE is ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9, or ABE7.10, as shown in below Table 2.

Table 2. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	I	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	I	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.3	I	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	I	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	I	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0337] In some embodiments, the base editor is a fusion protein comprising a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9-derived domain) fused to a nucleobase editing domain (*e.g.*, all or a portion of a deaminase domain). In some embodiments, the base editor further comprises a domain comprising all or a portion of a uracil glycosylase inhibitor (UGI). In some embodiments, the base editor comprises a domain comprising all or a portion of a uracil binding protein (UBP), such as a uracil DNA glycosylase (UDG). In some embodiments, the base editor comprises a domain comprising all or a portion of a nucleic acid polymerase. In some embodiments, a nucleic acid polymerase or portion thereof incorporated into a base editor is a translesion DNA polymerase.

[0338] In some embodiments, a domain of the base editor can comprise multiple domains. For example, the base editor comprising a polynucleotide programmable nucleotide binding domain derived from Cas9 can comprise an REC lobe and an NUC lobe corresponding to the REC lobe and NUC lobe of a wild-type or natural Cas9. In another example, the base editor can comprise one or more of a RuvCI domain, BH domain, REC1 domain, REC2 domain, RuvCII domain, L1 domain, HNH domain, L2 domain, RuvCIII domain, WED domain, TOPO domain or CTD domain. In some embodiments, one or more domains of the base editor comprise a mutation (*e.g.*, substitution, insertion, deletion) relative to a wild type version of a polypeptide comprising the domain. For example, an HNH domain of a polynucleotide programmable DNA binding domain can comprise an H840A substitution. In another example, a RuvCI domain of a polynucleotide programmable DNA binding domain can comprise a D10A substitution.

[0339] Different domains (*e.g.* adjacent domains) of the base editor disclosed herein can be connected to each other with or without the use of one or more linker domains (*e.g.* an XTEN linker domain). In some cases, a linker domain can be a bond (*e.g.*, covalent bond), chemical group, or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a first domain (*e.g.*, Cas9-derived domain) and a second domain (*e.g.*, a cytidine deaminase domain or adenosine deaminase domain). In some embodiments, a linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-hetero atom bond, etc.). In

certain embodiments, a linker is a carbon nitrogen bond of an amide linkage. In certain embodiments, a linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, a linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, etc.). In certain embodiments, a linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, a linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.). In some embodiments, a linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, a linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, a linker comprises a polyethylene glycol moiety (PEG). In certain embodiments, a linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. A linker can include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile can be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic acid editing protein. In some embodiments, a linker joins a dCas9 and a second domain (*e.g.*, cytidine deaminase, UGI, etc.).

[0340] Typically, a linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, a linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, a linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, a linker is 2-100 amino acids in length, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. In some embodiments, the linker is about 3 to about 104 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100) amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker domain comprises the amino acid sequence SGSETPGTSESATPES, which can also be referred to as the XTEN linker. Any method for linking the fusion protein domains can be employed (*e.g.*, ranging from very flexible linkers of the form (SGGS)n, (GGGS)n, (GGGGS)n, and (G)n, to more rigid linkers of the form (EAAAK)n, (GGS)n, SGSETPGTSESATPES (see, e.g.,

Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents are incorporated herein by reference), or $(XP)_n$ motif, in order to achieve the optimal length for activity for the nucleobase editor. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises a $(GGS)_n$ motif, wherein n is 1, 3, or 7. In some embodiments, the Cas9 domain of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES. In some embodiments, a linker comprises a plurality of proline residues and is 5-21, 5-14, 5-9, 5-7 amino acids in length, e.g., PAPAP, PAPAPA, PAPAPAP, PAPAPAPA, $P(AP)_4$, $P(AP)_7$, $P(AP)_{10}$ (see, e.g., Tan J, Zhang F, Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat Commun. 2019 Jan 25;10(1):439; the entire contents are incorporated herein by reference). Such proline-rich linkers are also termed "rigid" linkers.

[0341] A fusion protein of the invention comprises a nucleic acid editing domain. In some embodiments, the nucleic acid editing domain can catalyze a C to U base change. In some embodiments, the nucleic acid editing domain is a deaminase domain. In some embodiments, the deaminase is a cytidine deaminase or an adenosine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBECl deaminase. In some embodiments, the deaminase is an APOBEC2 deaminase. In some embodiments, the deaminase is an APOBEC3 deaminase. In some embodiments, the deaminase is an APOBEC3 A deaminase. In some embodiments, the deaminase is an APOBEC3B deaminase. In some embodiments, the deaminase is an APOBEC3C deaminase. In some embodiments, the deaminase is an APOBEC3D deaminase. In some embodiments, the deaminase is an APOBEC3E deaminase. In some embodiments, the deaminase is an APOBEC3F deaminase. In some embodiments, the deaminase is an APOBEC3G deaminase. In some embodiments, the deaminase is an APOBEC3H deaminase. In some embodiments, the deaminase is an APOBEC4 deaminase. In some embodiments, the deaminase is an activation-induced deaminase (AID). In some embodiments, the deaminase is a vertebrate deaminase. In some embodiments, the deaminase is an invertebrate deaminase. In some embodiments, the deaminase is a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse deaminase. In some embodiments, the deaminase is a human deaminase. In some embodiments, the deaminase is a rat deaminase, e.g., rAPOBECl. In some embodiments, the deaminase is a *Petromyzon marinus* cytidine deaminase 1 (pmCDAl). In some embodiments, the deaminase is a human APOBEC3G. In some embodiments, the

deaminase is a fragment of the human APOBEC3G. In some embodiments, the deaminase is a human APOBEC3G variant comprising a D316R D317R mutation. In some embodiments, the deaminase is a fragment of the human APOBEC3G and comprising mutations corresponding to the D316R D317R mutations. In some embodiments, the nucleic acid editing domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%), or at least 99.5% identical to the deaminase domain of any deaminase described herein.

Cas9 complexes with guide RNAs

[0342] Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA (e.g., a guide that targets an Mecp2 allele bearing RTT targetable mutations).

[0343] In some embodiments, the guide nucleic acid (e.g., guide RNA) is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is a sequence in the genome of a bacteria, yeast, fungi, insect, plant, or animal. In some embodiments, the target sequence is a sequence in the genome of a human. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to a non-canonical PAM sequence (e.g., a sequence listed in Table 1 or 5'-NAA-3'). In some embodiments, the guide nucleic acid (e.g., guide RNA) is complementary to a sequence in Mecp2 allele bearing RTT targetable mutations.

[0344] Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence. In some embodiments, the 3' end of the target

sequence is immediately adjacent to an NGA, NAA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence.

[0345] It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, e.g., in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, e.g., by sequence alignment and determination of homologous residues.

[0346] It will be apparent to those of skill in the art that in order to target any of the fusion proteins disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

[0347] The domains of the base editor disclosed herein can be arranged in any order. Non-limiting examples of a base editor comprising a fusion protein comprising *e.g.*, a polynucleotide-programmable nucleotide-binding domain and a deaminase domain can be arranged as following:

NH_2-[nucleobase editing domain]-Linker1-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, adenosine deaminase]-Linker1-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, adenosine deaminase]-Linker1-[*e.g.*, Cas9 derived domain]-Linker2-[UGI]-COOH;

NH_2-[*e.g.*, TadA7.10]-Linker1-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, adenosine deaminase]-Linker1-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, TadA7.10]-Linker1-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, TadA7.10]-Linker1-[*e.g.*, Cas9 derived domain]-Linker2-[UGI]-COOH

NH_2-[*e.g.*, adenosine deaminase]-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, Cas9 derived domain]-[*e.g.*, adenosine deaminase]-COOH;

NH_2-[*e.g.*, adenosine deaminase]-[*e.g.*, Cas9 derived domain]-[inosine BER inhibitor]-COOH;

NH_2-[*e.g.*, adenosine deaminase]-[inosine BER inhibitor]-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[inosine BER inhibitor]-[*e.g.*, adenosine deaminase]-[*e.g.*, Cas9 derived domain]-COOH;

NH_2-[*e.g.*, Cas9 derived domain]-[*e.g.*, adenosine deaminase]-[inosine BER inhibitor]-COOH;

NH_2-[*e.g.*, Cas9 derived domain]-[inosine BER inhibitor]-[*e.g.*, adenosine deaminase]-COOH; or

NH_2-[inosine BER inhibitor]-[*e.g.*, Cas9 derived domain]-[*e.g.*, adenosine deaminase]-COOH.

[0348] Additionally, in some cases, a Gam protein can be fused to an N terminus of a base editor. In some cases, a Gam protein can be fused to a C terminus of a base editor. The Gam protein of bacteriophage Mu can bind to the ends of double strand breaks (DSBs) and protect them from degradation. In some embodiments, using Gam to bind the free ends of DSB can reduce indel formation during the process of base editing. In some embodiments, 174-residue Gam protein is fused to the N terminus of the base editors. See. Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017). In some cases, a mutation or mutations can change the length of a base editor domain relative to a wild type domain. For example, a deletion of at least one amino acid in at least one domain can reduce the length of the base editor. In another case, a mutation or mutations do not change the length of a domain relative to a wild type domain. For example, substitution(s) in any domain does/do not change the length of the base editor Non-limiting examples of such base editors, where the length of all the domains is the same as the wild type domains, can include:

NH_2-[APOBEC1]-Linker1-[Cas9(D10A)]-Linker2-[UGI]-COOH;

NH_2-[CDA1]-Linker1-[Cas9(D10A)]-Linker2-[UGI]-COOH;

NH_2-[AID]-Linker1-[Cas9(D10A)]-Linker2-[UGI]-COOH;

NH_2-[APOBEC1]-Linker1-[Cas9(D10A)]-Linker2-[SSB]-COOH;

NH_2-[UGI]-Linker1-[ABOBEC1]-Linker2-[Cas9(D10A)]-COOH;
NH_2-[APOBEC1]-Linker1-[Cas9(D10A)]-Linker2-[UGI]-Linker3-[UGI]-COOH;
NH_2-[Cas9(D10A)]-Linker1-[CDA1]-Linker2-[UGI]-COOH;
NH_2-[Gam]-Linker1-[APOBEC1]-Linker2-[Cas9(D10A)]-Linker3-[UGI]-COOH;
NH_2-[Gam]-Linker1-[APOBEC1]-Linker2-[Cas9(D10A)]-Linker3-[UGI]-Linker4-[UGI]-COOH;
NH2-[APOBEC1]-Linker1-[dCas9(D10A, H840A)]-Linker2-[UGI]-COOH; or
NH2-[APOBEC1]-Linker1-[dCas9(D10A, H840A)]-COOH.

[0349] In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (*e.g.*, a "deamination window"). In some cases, a target can be within a 4 base region. In some cases, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

[0350] A defined target region can be a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

[0351] The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a polynucleotide programmable nucleotide binding domain. In some embodiments, an NLS of the base editor is localized C-terminal to a polynucleotide programmable nucleotide binding domain.

[0352] Other exemplary features that can be present in a base editor as disclosed herein are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are

useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

[0353] Non-limiting examples of protein domains which can be included in the fusion protein include a deaminase domain (*e.g.*, cytidine deaminase and/or adenosine deaminase), a uracil glycosylase inhibitor (UGI) domain, epitope tags, reporter gene sequences, and/or protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, and nucleic acid binding activity. Additional domains can be a heterologous functional domain. Such heterologous functional domains can confer a function activity, such as DNA methylation, DNA damage, DNA repair, modification of a target polypeptide associated with target DNA (*e.g.*, a histone, a DNA-binding protein, etc.), leading to, for example, histone methylation, histone acetylation, histone ubiquitination, and the like.

[0354] Other functions conferred can include methyltransferase activity, demethylase activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, remodeling activity, protease activity, oxidoreductase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, synthase activity, synthetase activity, and demyristoylation activity, or any combination thereof.

[0355] Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan

fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

Base Editor Efficiency

[0356] CRISPR-Cas9 nucleases have been widely used to mediate targeted genome editing. In most genome editing applications, Cas9 forms a complex with a guide polynucleotide (*e.g.*, single guide RNA (sgRNA)) and induces a double-stranded DNA break (DSB) at the target site specified by the sgRNA sequence. Cells primarily respond to this DSB through the non-homologous end-joining (NHEJ) repair pathway, which results in stochastic insertions or deletions (indels) that can cause frameshift mutations that disrupt the gene. In the presence of a donor DNA template with a high degree of homology to the sequences flanking the DSB, gene correction can be achieved through an alternative pathway known as homology directed repair (HDR). Unfortunately, under most non-perturbative conditions HDR is inefficient, dependent on cell state and cell type, and dominated by a larger frequency of indels. As most of the known genetic variations associated with human disease are point mutations, methods that can more efficiently and cleanly make precise point mutations are needed. Base editing system as provided herein provides a new way to edit genome editing without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

[0357] The base editors provided herein are capable of modifying a specific nucleotide base without generating a significant proportion of indels. The term "indel(s)", as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.*, mutate or deaminate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the target nucleotide sequence. In certain embodiments, any of the base editors provided herein are capable of generating a greater proportion of intended modifications (*e.g.*, point mutations or deaminations) versus indels.

[0358] In some embodiments, any of base editor system provided herein results in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less

than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence.

[0359] Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as a point mutation, in a nucleic acid (*e.g.* a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations.

[0360] In some embodiments, any of the base editors provided herein are capable of generating at least 0.01% of intended mutations (*i.e.* at least 0.01% base editing efficiency). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of intended mutations.

[0361] In some embodiments, the base editors provided herein are capable of generating a ratio of intended point mutations to indels that is greater than 1:1. In some embodiments, the base editors provided herein are capable of generating a ratio of intended point mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 8.5:1, at least 9:1, at least 10:1, at least 11:1, at least 12:1, at least 13:1, at least 14:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 200:1, at least 300:1, at least 400:1, at least 500:1, at least 600:1, at least 700:1, at least 800:1, at least 900:1, or at least 1000:1, or more.

[0362] The number of intended mutations and indels can be determined using any suitable method, for example, as described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632); Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017); the entire contents of which are hereby incorporated by reference.

[0363] In some embodiments, to calculate indel frequencies, sequencing reads are scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels can occur. If no exact matches are located, the read is excluded from analysis. If the length of this indel window exactly matches the reference sequence the read is classified as not containing an indel. If the indel window is two or more bases longer or shorter than the reference sequence, then the sequencing read is classified as an insertion or deletion, respectively. In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor.

[0364] The number of indels formed at a target nucleotide region can depend on the amount of time a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, the number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing the target nucleotide sequence (*e.g.*, a nucleic acid within the genome of a cell) to a base editor. It should be appreciated that the characteristics of the base editors as described herein can be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Multiplex Editing

[0365] In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more gene, wherein at least one gene is located in a different locus. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more base editor system. In some embodiments, the multiplex editing can comprise one or more base editor systems with a single guide polynucleotide. In some embodiments, the multiplex editing can comprise one or more base editor system with a plurality of guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more guide polynucleotide with a single base editor system. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide

sequence. In some embodiments, the multiplex editing can comprise a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editor provided herein. It should also be appreciated that the multiplex editing using any of the base editors as described herein can comprise a sequential editing of a plurality of nucleobase pairs.

[0366] The methods provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide of a subject (*e.g.*, a double-stranded DNA sequence) with a base editor system comprising a nucleobase editor (*e.g.*, an adenosine base editor or a cytidine base editor) and a guide polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence comprises a targeted nucleobase pair; (b) inducing strand separation of said target region; (c) editing a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase.

[0367] In some embodiments, said plurality of nucleobase pairs are in one more genes. In some embodiments, said plurality of nucleobase pairs are in the same gene. In some embodiments, at least one gene in said one more genes is located in a different locus.

[0368] In some embodiments, said editing is editing of said plurality of nucleobase pairs in at least one protein coding region. In some embodiments, said editing is editing of said plurality of nucleobase pairs in at least one protein non-coding region. In some embodiments, said editing is editing of said plurality of nucleobase pairs in at least one protein coding region and at least one protein non-coding region.

[0369] In some embodiments, said editing is in conjunction with one or more guide polynucleotides. In some embodiments, said base editor system can comprise one or more base editor system. In some embodiments, said base editor system can comprise one or more base editor systems in conjunction with a single guide polynucleotide. In some embodiments, said base editor system can comprise one or more base editor system in conjunction with a plurality of guide polynucleotides. In some embodiments, said editing is in conjunction with one or more guide polynucleotide with a single base editor system. In some embodiments, said editing is in conjunction with at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, said editing is in

conjunction with at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, said editing is in conjunction with a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editors provided herein. It should also be appreciated that said editing can comprise a sequential editing of a plurality of nucleobase pairs.

METHODS OF USING BASE EDITORS

[0370] The correction of point mutations in disease-associated genes and alleles provides new strategies for gene correction with applications in therapeutics and basic research.

[0371] The present disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a base editor system provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, *e.g.*, a disease caused by a genetic mutation, an effective amount of a nucleobase editor (*e.g.*, an adenosine deaminase base editor or a cytidine deaminase base editor) that corrects the point mutation in the disease associated gene. The present disclosure provides methods for the treatment of RTT that are associated or caused by a point mutation that can be corrected by deaminase mediated gene editing. Some such diseases are described herein, and additional suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure. It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering can be different, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species can affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues.

[0372] Provided herein are methods of using the base editor or base editor system for editing a nucleobase in a target nucleotide sequence associated with a disease or disorder. In some embodiments, the activity of the base editor (*e.g.*, comprising an adenosine deaminase and a Cas9 domain) results in a correction of the point mutation. In some embodiments, the target

DNA sequence comprises a G→A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence comprises a T→C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

[0373] In some embodiments, the target DNA sequence encodes a protein, and the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon. In some embodiments, the deamination of the mutant A results in a change of the amino acid encoded by the mutant codon. In some embodiments, the deamination of the mutant A results in the codon encoding the wild-type amino acid. In some embodiments, the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon. In some embodiments, the deamination of the mutant C results in the codon encoding the wild-type amino acid. In some embodiments, the subject has or has been diagnosed with a disease or disorder.

[0374] In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine of a deoxyadenosine residue of DNA. Other aspects of the disclosure provide fusion proteins that comprise an adenosine deaminase (*e.g.*, an adenosine deaminase that deaminates deoxyadenosine in DNA as described herein) and a domain (*e.g.*, a Cas9 or a Cpf 1 protein) capable of binding to a specific nucleotide sequence. For example, the adenosine can be converted to an inosine residue, which typically base pairs with a cytosine residue. Such fusion proteins are useful *inter alia* for targeted editing of nucleic acid sequences. Such fusion proteins can be used for targeted editing of DNA *in vitro*, *e.g.*, for the generation of mutant cells or animals; for the introduction of targeted mutations, *e.g.*, for the correction of genetic defects in cells *ex vivo*, *e.g.*, in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations *in vivo*, *e.g.*, the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a G to A, or a T to C to mutation can be treated using the nucleobase editors provided herein. The present disclosure provides deaminases, fusion proteins, nucleic acids, vectors, cells, compositions, methods, kits, systems, etc. that utilize the deaminases and nucleobase editors.

Use of Nucleobase Editors to Target Nucleotides in the Mecp2 gene

[0375] The suitability of nucleobase editors that target a nucleotide in *the Mecp2 gene* is evaluated as described herein. In one embodiment, a single cell of interest is transfected,

transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a nucleobase editor described herein together with a small amount of a vector encoding a reporter (e.g., GFP). These cells can be immortalized human cell lines, such as 293T, K562 or U20S. Alternatively, primary human cells may be used. Cells may also be obtained from a subject or individual, such as from tissue biopsy, surgery, blood, plasma, serum, or other biological fluid. Such cells may be relevant to the eventual cell target,

[0376] Delivery may be performed using a viral vector as further described below. In one embodiment, transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation. Following transfection, expression of GFP can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different nucleobase editors to determine which combinations of editors give the greatest activity.

[0377] The activity of the nucleobase editor is assessed as described herein, i.e., by sequencing the target gene to detect alterations in the target sequence. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, miniprepped and sequenced with a single primer. Sequencing may also be performed using next generation sequencing techniques. When using next generation sequencing, amplicons may be 300-500 bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, e.g., for use in high throughput sequencing (for example on an Illumina MiSeq).

[0378] The fusion proteins that induce the greatest levels of target specific alterations in initial tests can be selected for further evaluation.

[0379] In particular embodiments, the nucleobase editors are used to target polynucleotides of interest. In one embodiment, a nucleobase editor of the invention is delivered to cells (e.g., a neuron) in conjunction with a guide RNA that is used to target a nucleic acid sequence, e.g., a *Mecp2* polynucleotide harboring RTT-associated mutations, thereby altering the target gene, i.e., *Mecp2*.

[0380] In some embodiments, a base editor is targeted by a guide RNA to introduce one or more edits to the sequence of a gene of interest. In some embodiments, the one or more alterations introduced into the *Mecp2* gene are as presented in Table 6 *infra*.

Generating an Intended Mutation

[0381] In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via gene editing. In some embodiments, the function of a dysfunctional gene is restored by introducing an intended mutation. The nucleobase editing proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*, *e.g.*, by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, *e.g.*, the fusion proteins comprising a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9) and a nucleobase editing domain (*e.g.*, an adenosine deaminase domain or a cytidine deaminase domain) can be used to correct any single point A to G or C to T mutation. In the first case, deamination of the mutant A to I corrects the mutation, and in the latter case, deamination of the A that is base-paired with the mutant T, followed by a round of replication, corrects the mutation.

[0382] In some embodiments, the present disclosure provides base editors that can efficiently generating an intended mutation, such as a point mutation, in a nucleic acid (*e.g.*, a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (*e.g.*, cytidine base editor or adenosine base editor) bound to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation. In some embodiments, the intended mutation is a mutation associated with a disease or disorder. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation associated with a disease or disorder. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation associated with a disease or disorder. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation within the coding region or non-coding region of a gene. In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation within the coding region or non-coding region of a gene. In some embodiments, the intended mutation is a point mutation that generates a stop codon, for example, a premature stop codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon.

[0383] In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations (*e.g.*, intended point mutations : unintended point mutations) that is greater than 1 : 1. In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations (*e.g.*, intended point mutations : unintended point mutations) that is at least 1.5: 1, at least 2: 1, at least 2.5: 1, at least 3: 1, at least 3.5: 1, at least 4: 1, at least 4.5: 1, at least 5: 1, at

least 5.5: 1, at least 6: 1, at least 6.5: 1, at least 7: 1, at least 7.5: 1, at least 8: 1, at least 10: 1, at least 12: 1, at least 15: 1, at least 20: 1, at least 25: 1, at least 30: 1, at least 40: 1, at least 50: 1, at least 100: 1, at least 150: 1, at least 200: 1, at least 250: 1, at least 500: 1, or at least 1000: 1, or more

[0384] Details of base editor efficiency are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

[0385] In some embodiments, said editing of said plurality of nucleobase pairs in one or more genes result in formation of at least one intended mutation. In some embodiments, said formation of said at least one intended mutation results in a precise correction of a disease-causing mutation. It should be appreciated that the characteristics of the multiplex editing of the base editors as described herein can be applied to any of combination of the methods of using the base editor provided herein.

Precise Correction of Pathogenic Mutations

[0386] In some embodiments, the intended mutation is a precise correction of a pathogenic mutation or a disease-causing mutation. The pathogenic mutation can be a pathogenic single nucleotide polymorphism (SNP) or be caused by a SNP. For example, the pathogenic mutation can be an amino acid change in a protein encoded by a gene. In another example, the pathogenic mutation can be a pathogenic SNP in a gene. The precise correction can be to revert the pathogenic mutation back to its wild-type state. In some embodiments, the pathogenic mutation is a G→A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base with an A-to-G base editor (ABE) results in a sequence that is not associated with a disease or disorder. In some embodiments, the pathogenic mutation is a C→T point mutation. The C→T point mutation can be corrected, for example, by targeting an A-to-G base editor (ABE) to the opposite strand and editing the complement A of the pathogenic T nucleobase. In some embodiments, the pathogenic mutation is a T→C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base with a

C-to-T base editor (BE or CBE) results in a sequence that is not associated with a disease or disorder. In some embodiments, the pathogenic mutation is an A→G point mutation. The A→G point mutation can be corrected, for example, by targeting a CBE to the opposite strand and editing the complement C of the pathogenic G nucleobase. A base editor can be targeted to a pathogenic SNP, or to the complement of the pathogenic SNP. The nomenclature of the description of pathogenic or disease-causing mutations and other sequence variations are described in den Dunnen, J.T. and Antonarakis, S.E., "Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion." Human Mutation 15:712 (2000), the entire contents of which is hereby incorporated by reference.

[0387] In a particular embodiment, the disease or disorder is Rett Syndrome (RTT). In some embodiments, the pathogenic mutation is in the *Mecp2* gene.

DELIVERY SYSTEM

[0388] A base editor disclosed herein can be encoded on a nucleic acid that is contained in a viral vector. Exemplary viral vectors include retroviral vectors (e.g. Maloney murine leukemia virus, MML-V), adenoviral vectors (e.g. AD100), lentiviral vectors (HIV and FIV-based vectors), herpesvirus vectors (e.g. HSV-2), and adeno-associated viral vectors.

Adeno-Associated Viral Vectors (AAVs)

[0389] Adeno-associated virus ("AAV") vectors can also be used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, *e.g.*, West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[0390] In terms of *in vivo* delivery, AAV can be advantageous over other viral vectors. In some cases, AAV allows low toxicity, which can be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response. In some cases, AAV allows low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0391] AAV is a small, single-stranded DNA dependent virus belonging to the parvovirus family. The 4.7 kb wild-type (wt) AAV genome is made up of two genes that encode four replication proteins and three capsid proteins, respectively, and is flanked on either side by 145-

bp inverted terminal repeats (ITRs). The virion is composed of three capsid proteins, Vp1, Vp2, and Vp3, produced in a 1:1:10 ratio from the same open reading frame but from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. A phospholipase domain, which functions in viral infectivity, has been identified in the unique N terminus of Vp1.

[0392] AAV has a packaging limit of 4.5 or 4.75 Kb. Accordingly, a disclosed base editor as well as a promoter and transcription terminator can be harbored in a single viral vector. Constructs larger than 4.5 or 4.75 Kb can lead to significantly reduced virus production. For example, SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore, embodiments of the present disclosure include utilizing a disclosed base editor which is shorter in length than conventional base editors. In some examples, the base editors are less than 4 kb. Disclosed base editors can be less than 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb. In some cases, the disclosed base editors are 4.5 kb or less in length.

[0393] An AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the type of AAV with regard to the cells to be targeted; *e.g.*, one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)).

[0394] Similar to wt AAV, recombinant AAV (rAAV) utilizes the *cis*-acting 145-bp ITRs to flank vector transgene cassettes, providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. Although there are numerous examples of rAAV success using this system, in vitro and in vivo, the limited packaging capacity has limited the use of AAV-mediated gene delivery when the length of the coding sequence of the gene is equal or greater in size than the wt AAV genome.

[0395] The small packaging capacity of AAV vectors makes the delivery of a number of genes that exceed this size and/or the use of large physiological regulatory elements challenging. These challenges can be addressed, for example, by dividing the protein(s) to be delivered into two or more fragments, wherein the N-terminal fragment is fused to a split intein-N and the C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or

more AAV vectors. As used herein, "intein" refers to a self-splicing protein intron (e.g., peptide) that ligates flanking N-terminal and C-terminal exteins (e.g., fragments to be joined). The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem. 289(21); 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

[0396] A fragment of a fusion protein of the invention can vary in length. In some embodiments, a protein fragment ranges from 2 amino acids to about 1000 amino acids in length. In some embodiments, a protein fragment ranges from about 5 amino acids to about 500 amino acids in length. In some embodiments, a protein fragment ranges from about 20 amino acids to about 200 amino acids in length. In some embodiments, a protein fragment ranges from about 10 amino acids to about 100 amino acids in length. Suitable protein fragments of other lengths will be apparent to a person of skill in the art.

[0397] In some embodiments, a portion or fragment of a nuclease (e.g., Cas9) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

[0398] In one embodiment, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5′ and 3′ ends, or head and tail), where each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette is then achieved upon co-infection of the same cell by both dual AAV vectors followed by: (1) homologous recombination (HR) between 5′ and 3′ genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5′ and 3′ genomes (dual AAV *trans*-splicing vectors); or (3) a combination of these two mechanisms (dual AAV hybrid vectors). The use of dual AAV vectors *in vivo* results in the expression of full-length proteins. The use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of >4.7 kb in size.

[0399] The use of RNA or DNA viral based systems for the delivery of a base editor takes advantage of highly evolved processes for targeting a virus to specific cells in culture or in the

host and trafficking the viral payload to the nucleus or host cell genome. Viral vectors can be administered directly to cells in culture, patients (*in vivo*), or they can be used to treat cells *in vitro*, and the modified cells can optionally be administered to patients (*ex vivo*). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0400] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, *e.g.*, Buchscher et al., J. Virol., 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol., 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0401] Retroviral vectors, especially lentiviral vectors, can require polynucleotide sequences smaller than a given length for efficient integration into a target cell. For example, retroviral vectors of length greater than 9 kb can result in low viral titers compared with those of smaller size. In some aspects, a base editor of the present disclosure is of sufficient size so as to enable efficient packaging and delivery into a target cell via a retroviral vector. In some cases, a base editor is of a size so as to allow efficient packing and delivery even when expressed together with a guide nucleic acid and/or other components of a targetable nuclease system.

[0402] In applications where transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system.

[0403] A base editor described herein can therefore be delivered with viral vectors. One or more components of the base editor system can be encoded on one or more viral vectors. For example, a base editor and guide nucleic acid can be encoded on a single viral vector. In other cases, the base editor and guide nucleic acid are encoded on different viral vectors. In either case, the base editor and guide nucleic acid can each be operably linked to a promoter and terminator.

[0404] The combination of components encoded on a viral vector can be determined by the cargo size constraints of the chosen viral vector.

[0405] Any suitable promoter can be used to drive expression of the base editor and, where appropriate, the guide nucleic acid. For ubiquitous expression, promoters that can be used include CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc. For brain or other CNS cell expression, suitable promoters can include: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc. For liver cell expression, suitable promoters include the Albumin promoter. For lung cell expression, suitable promoters can include SP-B. For endothelial cells, suitable promoters can include ICAM. For hematopoietic cells suitable promoters can include IFNbeta or CD45. For Osteoblasts suitable promoters can include OG-2.

[0406] A promoter used to drive base editor coding nucleic acid molecule expression can include AAV ITR. This can be advantageous for eliminating the need for an additional promoter element, which can take up space in the vector. The additional space freed up can be used to drive the expression of additional elements, such as a guide nucleic acid or a selectable marker. ITR activity is relatively weak, so it can be used to reduce potential toxicity due to over expression of the chosen nuclease.

[0407] In some cases, a base editor of the present disclosure is of small enough size to allow separate promoters to drive expression of the base editor and a compatible guide nucleic acid within the same nucleic acid molecule. For instance, a vector or viral vector can comprise a first promoter operably linked to a nucleic acid encoding the base editor and a second promoter operably linked to the guide nucleic acid.

[0408] The promoter used to drive expression of a guide nucleic acid can include: Pol III promoters such as U6 or H1 Use of Pol II promoter and intronic cassettes to express gRNA Adeno Associated Virus (AAV).

[0409] A base editor described herein with or without one or more guide nucleic acids can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No.

8,454,972 (formulations, doses for adenovirus), U.S. Patent No. 8,404,658 (formulations, doses for AAV) and U.S. Patent No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in U.S. Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Patent No. 5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (*e.g.* a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (*e.g.*, physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific base editing, the expression of the base editor and optional guide nucleic acid can be driven by a cell-type specific promoter.

[0410] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[0411] Lentiviruses can be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media is changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells are transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5 μg of psPAX2 (gag/pol/rev/tat). Transfection can be done in 4 mL OptiMEM with a cationic lipid delivery agent (50 μl Lipofectamine 2000 and 100 ul Plus reagent). After 6 hours, the media is changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

[0412] Lentivirus can be purified as follows. Viral supernatants are harvested after 48 hours. Supernatants are first cleared of debris and filtered through a 0.45 μm low protein binding (PVDF) filter. They are then spun in an ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets

are resuspended in 50 µl of DMEM overnight at 4° C. They are then aliquoted and immediately frozen at -80°C.

[0413] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated. In another embodiment, RetinoStat.RTM., an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is contemplated to be delivered via a subretinal injection. In another embodiment, use of self-inactivating lentiviral vectors is contemplated.

[0414] Any RNA of the systems, for example a guide RNA or a base editor-encoding mRNA, can be delivered in the form of RNA. Base editor-encoding mRNA can be generated using *in vitro* transcription. For example, nuclease mRNA can be synthesized using a PCR cassette containing the following elements: T7 promoter, optional Kozak sequence (GCCACC), nuclease sequence, and 3' UTR such as a 3' UTR from beta globin-polyA tail. The cassette can be used for transcription by T7 polymerase. Guide polynucleotides (*e.g.*, gRNA) can also be transcribed using *in vitro* transcription from a cassette containing a T7 promoter, followed by the sequence "GG", and guide polynucleotide sequence.

[0415] To enhance expression and reduce possible toxicity, the base editor-coding sequence and/or the guide nucleic acid can be modified to include one or more modified nucleoside *e.g.* using pseudo-U or 5-Methyl-C.

[0416] The disclosure in some embodiments comprehends a method of modifying a cell or organism. The cell can be a prokaryotic cell or a eukaryotic cell. The cell can be a mammalian cell. The mammalian cell many be a non-human primate, bovine, porcine, rodent or mouse cell. The modification introduced to the cell by the base editors, compositions and methods of the present disclosure can be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the methods of the present disclosure can be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

[0417] The system can comprise one or more different vectors. In an aspect, the base editor is codon optimized for expression the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell.

[0418] In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (*e.g.* about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with

codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ (visited Jul. 9, 2002), and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (*e.g.* 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding an engineered nuclease correspond to the most frequently used codon for a particular amino acid.

[0419] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and psi.2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA can be packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus can promote replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid in some cases is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

Non-Viral Delivery of Base Editors

[0420] Non-viral delivery approaches for delivery of base editors are also available. One important category of non-viral nucleic acid vectors are nanoparticles, which can be organic or inorganic. Nanoparticles are well known in the art. Any suitable nanoparticle design can be used to deliver genome editing system components or nucleic acids encoding such components. For instance, organic (e.g. lipid and/or polymer) nanoparticles can be suitable for use as delivery vehicles in certain embodiments of this disclosure. Exemplary lipids for use in nanoparticle formulations, and/or gene transfer are shown in Table 3 (below).

Table 3

Lipids Used for Gene Transfer

Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleyloxy)prophyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Dioctadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy-N-[2(sperminecarboxamido-ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Dioctadecylamidoglicylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammoniun bromide	CLIP-6	Cationic
Ethyldimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearyloxy-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylpho sphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic

Lipids Used for Gene Transfer

Lipid	Abbreviation	Feature
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
N1 -Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl-methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 4 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 4

Polymers Used for Gene Transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis (succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine)biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(N-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amidoethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β-aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α-[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

Table 5 summarizes delivery methods for a polynucleotide encoding a fusion protein described herein.

Table 5

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical	(e.g., electroporation, particle gun, Calcium Phosphate transfection	YES	Transient	NO	Nucleic Acids and Proteins
Viral	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modification	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral Delivery Vehicles	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological	YES	Transient	NO	Nucleic Acids

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
	liposomes: Erythrocyte Ghosts and Exosomes				

[0421] In another aspect, the delivery of genome editing system components or nucleic acids encoding such components, for example, a nucleic acid binding protein such as, for example, Cas9 or variants thereof, and a gRNA targeting a genomic nucleic acid sequence of interest, may be accomplished by delivering a ribonucleoprotein (RNP) to cells. The RNP comprises the nucleic acid binding protein, e.g., Cas9, in complex with the targeting gRNA. RNPs may be delivered to cells using known methods, such as electroporation, nucleofection, or cationic lipid-mediated methods, for example, as reported by Zuris, J.A. et al., 2015, *Nat. Biotechnology*, 33(1):73-80. RNPs are advantageous for use in CRISPR base editing systems, particularly for cells that are difficult to transfect, such as primary cells. In addition, RNPs can also alleviate difficulties that may occur with protein expression in cells, especially when eukaryotic promoters, e.g., CMV or EF1A, which may be used in CRISPR plasmids, are not well-expressed. Advantageously, the use of RNPs does not require the delivery of foreign DNA into cells. Moreover, because an RNP comprising a nucleic acid binding protein and gRNA complex is degraded over time, the use of RNPs has the potential to limit off-target effects. In a manner similar to that for plasmid based techniques, RNPs can be used to deliver binding protein (e.g., Cas9 variants) and to direct homology directed repair (HDR).

PHARMACEUTICAL COMPOSITIONS

[0422] Other aspects of the present disclosure relate to pharmaceutical compositions comprising any of the base editors, fusion proteins, or the fusion protein-guide polynucleotide complexes described herein. The term "pharmaceutical composition", as used herein, refers to a composition formulated for pharmaceutical use. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier. In some embodiments, the pharmaceutical composition comprises additional agents (*e.g.*, for specific delivery, increasing half-life, or other therapeutic compounds).

[0423] As used here, the term "pharmaceutically-acceptable carrier" means a pharmaceutically-acceptable material, composition or vehicle, such as a liquid or solid filler, diluent, excipient, manufacturing aid (*e.g.*, lubricant, talc magnesium, calcium or zinc stearate,

or steric acid), or solvent encapsulating material, involved in carrying or transporting the compound from one site (*e.g.*, the delivery site) of the body, to another site (*e.g.*, organ, tissue or portion of the body). A pharmaceutically acceptable carrier is "acceptable" in the sense of being compatible with the other ingredients of the formulation and not injurious to the tissue of the subject (*e.g.*, physiologically compatible, sterile, physiologic pH, etc.).

[0424] Some examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum lcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation. The terms such as "excipient", "carrier", "pharmaceutically acceptable carrier" or the like are used interchangeably herein.

[0425] Pharmaceutical compositions can comprise one or more pH buffering compounds to maintain the pH of the formulation at a predetermined level that reflects physiological pH, such as in the range of about 5.0 to about 8.0. The pH buffering compound used in the aqueous liquid formulation can be an amino acid or mixture of amino acids, such as histidine or a mixture of amino acids such as histidine and glycine. Alternatively, the pH buffering compound is preferably an agent which maintains the pH of the formulation at a predetermined level, such as in the range of about 5.0 to about 8.0, and which does not chelate calcium ions. Illustrative examples of such pH buffering compounds include, but are not limited to, imidazole and acetate ions. The pH buffering compound may be present in any amount suitable to maintain the pH of the formulation at a predetermined level.

[0426] Pharmaceutical compositions can also contain one or more osmotic modulating agents, i.e., a compound that modulates the osmotic properties (e.g., tonicity, osmolality, and/or osmotic pressure) of the formulation to a level that is acceptable to the blood stream and blood cells of recipient individuals. The osmotic modulating agent can be an agent that does not chelate calcium ions. The osmotic modulating agent can be any compound known or available to those skilled in the art that modulates the osmotic properties of the formulation. One skilled in the art may empirically determine the suitability of a given osmotic modulating agent for use in the inventive formulation. Illustrative examples of suitable types of osmotic modulating agents include, but are not limited to: salts, such as sodium chloride and sodium acetate; sugars, such as sucrose, dextrose, and mannitol; amino acids, such as glycine; and mixtures of one or more of these agents and/or types of agents. The osmotic modulating agent(s) may be present in any concentration sufficient to modulate the osmotic properties of the formulation.

[0427] In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, *e.g.*, for gene editing. Suitable routes of administrating the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

[0428] In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site (*e.g.*, tumor site). In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber.

[0429] In other embodiments, the pharmaceutical composition described herein is delivered in a controlled release system. In one embodiment, a pump can be used (see, *e.g.*, Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al, 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used. (See, *e.g.*, Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy et al., 1985, Science

228: 190; During et al., 1989, Ann. Neurol. 25:351; Howard et ah, 1989, J. Neurosurg. 71: 105.) Other controlled release systems are discussed, for example, in Langer, supra.

[0430] In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, *e.g.*, a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic use as solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

[0431] A pharmaceutical composition for systemic administration can be a liquid, *e.g.*, sterile saline, lactated Ringer's or Hank's solution. In addition, the pharmaceutical composition can be in solid forms and re-dissolved or suspended immediately prior to use. Lyophilized forms are also contemplated. The pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration. The particles can be of any suitable structure, such as unilamellar or plurilamellar, so long as compositions are contained therein. Compounds can be entrapped in "stabilized plasmid-lipid particles" (SPLP) containing the fusogenic lipid dioleoylphosphatidylethanolamine (DOPE), low levels (5-10 mol%) of cationic lipid, and stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. et ah, Gene Ther. 1999, 6: 1438-47). Positively charged lipids such as N-[l-(2,3-dioleoyloxi)propyl]-N,N,N-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. *See, e.g.*, U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

[0432] The pharmaceutical composition described herein can be administered or packaged as a unit dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

[0433] Further, the pharmaceutical composition can be provided as a pharmaceutical kit comprising (a) a container containing a compound of the invention in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (*e.g.*, sterile used for reconstitution or dilution of the lyophilized compound of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

[0434] In some embodiments, any of the fusion proteins, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease (*e.g.*, Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

[0435] Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

[0436] Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage

form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease.

[0437] Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

[0438] The compositions, as described above, can be administered in effective amounts. The effective amount will depend upon the mode of administration, the particular condition being treated, and the desired outcome. It may also depend upon the stage of the condition, the age and physical condition of the subject, the nature of concurrent therapy, if any, and like factors well-known to the medical practitioner. For therapeutic applications, it is that amount sufficient to achieve a medically desirable result.

Methods of Treating RTT

[0439] Provided also are methods of treating Rett Syndrome (RTT) and/or the genetic mutations in *Mecp2* that cause RTT that comprise administering to a subject (e.g., a mammal, such as a human) a therapeutically effective amount of a pharmaceutical composition that comprises a polynucleotide encoding a base editor system (e.g., base editor and gRNA) described herein. In some embodiments, the base editor is a fusion protein that comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain or a cytidine deaminase domain. A cell of the subject is transduced with the base editor and one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration (if the cell is transduced with an adenosine deaminase domain) or a C•G to U•A alteration (if the cell is transduced with a cytidine deaminase domain) of a nucleic acid sequence containing mutations in the *Mecp2* gene.

[0440] The methods herein include administering to the subject (including a subject identified as being in need of such treatment, or a subject suspected of being at risk of disease and in need of such treatment) an effective amount of a composition described herein. Identifying a subject

in need of such treatment can be in the judgment of a subject or a health care professional and can be subjective (e.g. opinion) or objective (e.g. measurable by a test or diagnostic method).

[0441] The therapeutic methods, in general, comprise administration of a therapeutically effective amount of a pharmaceutical composition comprising, for example, a vector encoding a base editor and a gRNA that targets the *Mecp2* gene of a subject (e.g., a human patient) in need thereof. Such treatment will be suitably administered to a subject, particularly a human subject, suffering from, having, susceptible to, or at risk for RTT. The compositions herein may be also used in the treatment of any other disorders in which RTT may be implicated.

[0442] In one embodiment, the invention provides a method of monitoring treatment progress is provided. The method includes the step of determining a level of diagnostic marker (Marker) (e.g., SNP associated with RTT) or diagnostic measurement (e.g., screen, assay) in a subject suffering from or susceptible to a disorder or symptoms thereof associated with RTT in which the subject has been administered a therapeutic amount of a composition herein sufficient to treat the disease or symptoms thereof. The level of Marker determined in the method can be compared to known levels of Marker in either healthy normal controls or in other afflicted patients to establish the subject's disease status. In preferred embodiments, a second level of Marker in the subject is determined at a time point later than the determination of the first level, and the two levels are compared to monitor the course of disease or the efficacy of the therapy. In certain preferred embodiments, a pre-treatment level of Marker in the subject is determined prior to beginning treatment according to this invention; this pre-treatment level of Marker can then be compared to the level of Marker in the subject after the treatment commences, to determine the efficacy of the treatment.

[0443] In some embodiments, cells are obtained from the subject and contacted with a pharmaceutical composition as provided herein. In some embodiments, cells removed from a subject and contacted *ex vivo* with a pharmaceutical composition are re-introduced into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

Kits

[0444] Various aspects of this disclosure provide kits comprising a base editor system. In one embodiment, the kit comprises a nucleic acid construct comprising a nucleotide sequence encoding a nucleobase editor fusion protein. The fusion protein comprises a deaminase (e.g., cytidine deaminase or adenine deaminase) and a nucleic acid programmable DNA binding protein (napDNAbp). In some embodiments, the kit comprises at least one guide RNA capable of targeting a nucleic acid molecule of interest, e.g., *Mecp2* RTT-associated mutations. In some embodiments, the kit comprises a nucleic acid construct comprising a nucleotide sequence encoding at least one guide RNA.

[0445] The kit provides, in some embodiments, instructions for using the kit to edit one or more *Mecp2* RTT-associated mutations. The instructions will generally include information about the use of the kit for editing nucleic acid molecules. In other embodiments, the instructions include at least one of the following: precautions; warnings; clinical studies; and/or references. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container. In a further embodiment, a kit can comprise instructions in the form of a label or separate insert (package insert) for suitable operational parameters. In yet another embodiment, the kit can comprise one or more containers with appropriate positive and negative controls or control samples, to be used as standard(s) for detection, calibration, or normalization. The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as (sterile) phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

[0446] In certain embodiments, the kit is useful for the treatment of a subject having Rett Syndrome.

[0447] The practice of the embodiments disclosed herein employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, for example, Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988)

Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R.I. Freshney, ed. (2010)).

[0448] The following numbered additional embodiments encompassing the methods and compositions of the base editor systems and uses are envisioned herein:

1. A method of treating Rett Syndrome (RTT) in a subject in need thereof, comprising administering to the subject a base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide of a cell in the subject, thereby treating RTT.

2. A method of treating Rett Syndrome (RTT) in a subject in need thereof, comprising

(a) introducing into a cell a base editor system comprising

a guide polynucleotides or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain, and

(b) administering the cell to the subject,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide in the cell, thereby treating the disease.

3. The method of embodiment 2, wherein the cell is a neuron.

4. The method of embodiment 2 or 3, wherein the cell is autologous, allogenic, or xenogenic to the subject.

5. A method of editing a *MECP2* polynucleotide, comprising contacting the *MECP2* polynucleotide with a base editor system comprising

a guide polynucleotides;

a polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain,

wherein the guide polynucleotides is capable of targeting the base editor system to effect an A•T to G•C of a single nucleotide polymorphism (SNP) causative of Rett Syndrome (RTT) in the *MECP2* polynucleotide.

6. A method of producing a modified cell for treatment of Rett Syndrome (RTT), comprising introducing into a cell a base editor system comprising

a guide polynucleotides or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide in the cell.

7. The method of embodiment 6, wherein the introduction is *in vivo*.

8. The method of embodiment 6, wherein the introduction is *ex vivo*.

9. The method of any one of embodiments 6-8, wherein the cell is a neuron.

10. The method of any one of embodiments 6-9, wherein the cell is obtained from a subject having RTT.

11. The method of any one of the preceding embodiments, wherein *MECP2* polynucleotide encodes an MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP.

12. The method of embodiment 11, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

13. A method of treating Rett Syndrome (RTT) in a subject in need thereof, comprising administering to the subject a base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide of a cell in the subject,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid, thereby treating RTT.

14. A method of treating Rett Syndrome (RTT) in a subject in need thereof, comprising

(a) contacting a cell with a base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

(b) administering the cell to the subject,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide in the cell,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid, thereby treating RTT.

15. The method of embodiment 14, wherein the cell is a neuron.

16. The method of embodiment 14 or 15, wherein the cell is autologous, allogenic, or xenogenic to the subject.

17. A method of correcting a single nucleotide polymorphism (SNP) causative of RTT in a *MECP2* polynucleotide, comprising contacting the *MECP2* polynucleotide with a base editor system comprising

a guide polynucleotide;

a polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain,

wherein the guide polynucleotides is capable of targeting the base editor system to effect an A•T to G•C of the SNP,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop

codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid, thereby correcting the SNP.

18. A method of producing a modified cell for treatment of Rett Syndrom (RTT), comprising introducing into a cell a base editor system comprising

a guide polynucleotides or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) in a *MECP2* polynucleotide in the cell,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

19. The method of embodiment 18, wherein the introduction is *in vivo*.

20. The method of embodiment 18, wherein the introduction is *ex vivo*.

21. The method of any one of embodiments 18-20, wherein the cell is a neuron.

22. The method of any one of embodiments 18-21, wherein the cell is obtained from a subject having RTT.

23. The method of any one of embodiments 12-22, wherein the wild type amino acid is an arginine.

24. The method of embodiment 23, wherein the A•T to G•C alteration substitutes a stop codon at position 255 of the MECP2 protein with an arginine.

25. The method of embodiment 24, wherein the SNP is at position 763 of the *MECP2* polynucleotide.

26. The method of any one of the preceding embodiments, wherein the polynucleotide programmable DNA binding domain is a Cas9 domain.

27. The method of embodiment 26, wherein the Cas9 domain is a nuclease inactive Cas9 domain.

28. The method of embodiment 26, wherein the Cas9 domain is a Cas9 nickase domain.

29. The method of any one of embodiments 26-28, wherein the Cas9 domain comprises a SpCas9 domain

30. The method of embodiment 29, wherein the SpCas9 domain comprises a D10A and/or a H840A amino acid substitution or corresponding amino acid substitutions thereof.

31. The method of embodiment 29 or 30, wherein the SpCas9 domain has specificity for a NGG PAM.

32. The method of any one of embodiments 29-31, wherein the SpCas9 domain has specificity for a NGA PAM, a NGT PAM, or a NGC PAM.

33. The method of any one of embodiments 29-32, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337I, T1337V, T1337F, and T1337M or corresponding amino acid substitutions thereof.

34. The method of any one of embodiments 29-32, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

35. The method of any one of embodiments 29-32, wherein the SpCas9 domain comprises amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, T1337, and A1322R, and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

36. The method of any one of embodiments 29-32, wherein the SpCas9 domain comprises amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R, or corresponding amino acid substitutions thereof.

37. The method of any one of embodiments 29-32, wherein the SpCas9 domain has specificity for a NG PAM, a NNG PAM, a GAA PAM, a GAT PAM, or a CAA PAM.

38. The method of embodiment 36, wherein the Cas9 domain comprises amino acid substitutions E480K, E543K, and E1219V or corresponding amino acid substitutions thereof.

39. The method of any one of embodiments 26-28, wherein the Cas9 domain comprises a SaCas9 domain.

40. The method of embodiment 39, wherein the SaCas9 domain has specificity for a NNNRRT PAM.

41. The method of embodiment 40, wherein the SaCas9 domain has specificity for a NNGRRT PAM.

42. The method of any one of embodiments 39-41, wherein the SaCas9 domain comprises an amino acid substitution N579A or a corresponding amino acid substitution thereof.

43. The method of any one of embodiments 39-42, wherein the SaCas9 domain comprises amino acid substitutions E782K, N968K, and R1015H, or corresponding amino acid substitutions thereof.

44. The method of any one of embodiments 26-28, wherein the Cas9 domain comprises a St1Cas9 domain.

45. The method of embodiment 44, wherein the St1Cas9 domain has specificity for a NNACCA PAM.

46. The method of any one of the preceding embodiments, wherein the adenosine deaminase domain is a modified adenosine deaminase domain that does not occur in nature.

47. The method of embodiment 46, wherein the adenosine deaminase domain comprises a TadA domain.

48. The method of embodiment 47, wherein the TadA domain comprises the amino acid sequence of TadA 7.10.

49. The method of any one of the preceding embodiments, wherein the base editor system further comprises a zinc finger domain.

50. The method of embodiment 97, wherein the zinc finger domain comprises recognition helix sequences RNEHLEV, QSTTLKR, and RTEHLAR or recognition helix sequences RGEHLRQ, QSGTLKR, and RNDKLVP.

51. The method of embodiment 49 or 50, wherein the zinc finger domain is zf1ra or zf1rb.

52. The method of any one of the preceding embodiments, wherein the base editor system further comprises a nuclear localization signal (NLS).

53. The method of any one of the preceding embodiments, wherein the base editor system further comprises one or more linkers.

54. The method of embodiment 53, wherein two or more of the polynucleotide programmable DNA binding domain, the adenosine deaminase domain, the zinc finger domain, and the NLS are connected via a linker.

55. The method of embodiment 54, wherein the linker is a peptide linker, thereby forming a base editing fusion protein.

56. The method of embodiment 55, wherein the peptide linker comprises an amino acid sequence selected from the group consisting of SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS, SGGSSGGSSGSETPGTSESATPES, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESS GGSSGGS, PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEG SAP GTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)n, (GGGS)n, (GGGGS)n, (G)n, (EAAAK)n, (GGS)n, SGSETPGTSESATPES, and (XP)n.

57. The method of embodiment 55 or 56, wherein the base editing fusion protein comprises the amino acid sequence selected from the group consisting of MPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEG WNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHS RIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR MRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSH EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS TDSGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIIHKNSR IFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPY QLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENS KQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQ TQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVK NEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDI EQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSS IFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQK ANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTG

KTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDS
MDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTLYA
SRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALI
IAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTL
KSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYT
QDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINDKGKEVPCN
PFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVS
PWRADVYFNKTTGKYEILGLKYADLQFDKGTGTYKISQEKYNDIKKKEGVDSD
SEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGE
ALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFPKK
KRKVEGADKRTADGSEFESPKKKRKV,
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
IELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY

YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKV,
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDA
KTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD
SGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPC
VMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSKRNYI
LGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLL
KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPE
ELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK
EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDI
QEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVP
KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQ
KMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHIL
NLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNN
LDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKK
VMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLIND
TLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIM
EQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRN
KVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ
AEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTI
ASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNS
NANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRNEHLEVHTRTHTGEKPFQC
RICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLKTHLRGSSAQ, or
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDA
KTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

SGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPC
VMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL
LCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSKRNYI
LGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLL
KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPE
ELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK
EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDI
QEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVP
KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQ
KMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN
NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHIL
NLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNN
LDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKK
VMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLIND
TLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIM
EQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRN
KVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ
AEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTI
ASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVSSGNS
NANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRGEHLRQHTRTHTGEKPFQC
RICMRNFSQSGTLKRHLRTHTGEKPFQCRICMRNFSRNDKLVPHLKTHLRGSSAQ.

58. The method of any one of the preceding embodiments, wherein the guide polynucleotide comprises two individual polynucleotides, wherein the two individual polynucleotides are two DNAs, two RNAs or a DNA and an RNA.

59. The method of any one of embodiments 1-58, wherein the guide polynucleotides comprise a crRNA and a tracrRNA, wherein the crRNA comprises a nucleic acid sequence complementary to a target sequence in the *MECP2* polynucleotide.

60. The method of embodiment 59, wherein the target sequence comprises sequence selected from the group consisting of CCATGTCCAGCCTTCAGGCA, TCCATGTCCAGCCTTCAGGC, TTCCATGTCCAGCCTTCAGG, GCTTCCATGTCCAGCCTTCA, CTTCCATGTCCAGCCTTCAGG,

AGCTTCCATGTCCAGCCTTCAGG, AGCTTCCATGTCCAGCCTTC, CTTAAGCTTCCATGTCCAGC, AGCAAAAGGCTTTTCCCTGG, GAGCAAAAGGCTTTTCCCTG, AGAGCAAAAGGCTTTTCCCT, TAGAGCAAAAGGCTTTTCCC, TAGAGCAAAAGGCTTTTCCCT, TTTAGAGCAAAAGGCTTTTCCCT, CCATGAAGTCAAAATCATTA, ACCATGAAGTCAAAATCATT, TACCATGAAGTCAAAATCAT, TTACCATGAAGTCAAAATCAT, AGTTACCATGAAGTCAAAATCAT, CTTTTCACTTCCTGCCGGGG, TCAGCCCCGTTTCTTGGGAA, GCTTTCAGCCCCGTTTCTTG, CGGCTTTCAGCCCCGTTTCTT, CCCGGCTTTCAGCCCCGTTTCTT, GCACTTCTTGATGGGGAGTA, TCTTGCACTTCTTGATGGGG, CTTGCACTTCTTGATGGGGAG, and GTCTTGCACTTCTTGATGGGGAG.

61. The method of embodiment 58 or 59, wherein the base editor system comprises a single guide RNA (sgRNA).

62. The method of embodiment 60, wherein the sgRNA comprises CUUUUCACUUCCUGCCGGGG.

63. A modified cell comprising a base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of Rett Syndrome (RTT) in a *MECP2* polynucleotide in the cell.

64. A modified cell comprising a base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of Rett Syndrome (RTT) in a *MECP2* polynucleotide in the cell,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

65. The modified cell of embodiment 63, wherein *MECP2* polynucleotide encodes an MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP.

66. The modified cell of embodiment 65, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

67. The modified cell of any one of embodiments 63-66, wherein the cell is a neuron.

68. The modified cell of any one of embodiments 63-67, wherein the cell is obtained from a subject having RTT.

69. The modified cell of any one of embodiments 66-68, wherein the wild type amino acid is an arginine.

70. The modified cell of embodiment 69, wherein the the A•T to G•C alteration substitutes a stop codon at position 255 of the MECP2 protein with an arginine

71. The modified cell of embodiment 70, wherein the SNP is at position 763 of the *MECP2* polynucleotide.

72. The modified cell of any one of embodiments 63-71, wherein the polynucleotide programmable DNA binding domain is a Cas9 domain.

73. The modified cell of embodiment 72, wherein the Cas9 domain is a nuclease inactive Cas9 domain.

74. The modified cell of embodiment 72, wherein the Cas9 domain is a Cas9 nickase domain.

75. The modified cell of any one of embodiments 72-74, wherein the Cas9 domain comprises a SpCas9 domain.

76. The modified cell of embodiment 75, wherein the SpCas9 domain comprises a D10A and/or a H840A amino acid substitution or corresponding amino acid substitutions thereof.

77. The modified cell of embodiment 75 or 76, wherein the SpCas9 domain has specificity for a NGG PAM.

78. The modified cell of any one of embodiments 75-77, wherein the SpCas9 domain has specificity for a NGA PAM, a NGT PAM, or a NGC PAM.

79. The modified cell of any one of embodiments 75-78, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337I, T1337V, T1337F, and T1337M or corresponding amino acid substitutions thereof.

80. The modified cell of any one of embodiments 75-78, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

81. The modified cell of any one of embodiments 75-78, wherein the SpCas9 domain comprises amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, T1337, and A1322R, and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

82. The modified cell of any one of embodiments 75-78, wherein the SpCas9 domain comprises amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R, or corresponding amino acid substitutions thereof.

83. The modified cell of any one of embodiments 75or 76, wherein the SpCas9 domain has specificity for a NG PAM, a NNG PAM, a GAA PAM, a GAT PAM, or a CAA PAM.

84. The modified cell of embodiment 83, wherein the SpCas9 domain comprises amino acid substitutions E480K, E543K, and E1219V or corresponding amino acid substitutions thereof.

85. The modified cell of any one of embodiments 72-74, wherein the Cas9 domain comprises a SaCas9 domain.

86. The modified cell of embodiment 85, wherein the SaCas9 domain has specificity for a NNNRRT PAM.

87. The modified cell of embodiment 86, wherein the SaCas9 domain has specificity for a NNGRRT PAM.

88. The modified cell of any one of embodiments 85-87, wherein the SaCas9 domain comprises an amino acid substitution N579A or a corresponding amino acid substitution thereof.

89. The modified cell of any one of embodiments 85-88, wherein the SaCas9 domain comprises amino acid substitutions E782K, N968K, and R1015H, or corresponding amino acid substitutions thereof.

90. The modified cell of any one of embodiments 72-74, wherein the Cas9 domain comprises a St1Cas9 domain.

91. The modified cell of embodiment 90, wherein the St1Cas9 domain has specificity for a NNACCA PAM.

92. The modified cell of any one of embodiments 63-91, wherein the adenosine deaminase domain is a modified adenosine deaminase domain that does not occur in nature.

93. The modified cell of embodiment 92, wherein the adenosine deaminase domain comprises a TadA domain.

94. The modified cell of embodiment 93, wherein the TadA domain comprises the amino acid sequence of TadA 7.10.

95. The modified cell of embodiments 63-94, wherein the base editor system further comprises a zinc finger domain.

96. The modified cell of embodiment 95, wherein the zinc finger domain comprises recognition helix sequences RNEHLEV, QSTTLKR, and RTEHLAR or recognition helix sequences RGEHLRQ, QSGTLKR, and RNDKLVP.

97. The modified cell of embodiment 95 or 96, wherein the zinc finger domain is zf1ra or zf1rb.

98. The modified cell of any one of embodiments 63-97, wherein the base editor system further comprises a nuclear localization signal (NLS).

99. The modified cell of any one embodiments 63-98, wherein the base editor system further comprises one or more linkers.

100. The modified cell of embodiment 99, wherein two or more of the polynucleotide programmable DNA binding domain, the adenosine deaminase domain, the zinc finger domain, and the NLS are connected via a linker.

101. The modified cell of embodiment 54, wherein the linker is a peptide linker, thereby forming a base editing fusion protein.

102. The modified cell of embodiment 100, wherein the peptide linker comprises an amino acid sequence selected from the group consisting of

SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS, SGGSSGGSSGSETPGTSESATPES, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESS GGSSGGS, PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEG SAP GTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)n, (GGGS)n, (GGGGS)n, (G)n, (EAAAK)n, (GGS)n, SGSETPGTSESATPES, and (XP)n.

103. The modified cell of embodiment 101 or 102, wherein the base editing fusion protein comprises the amino acid sequence selected from the group consisting MPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEG WNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHS RIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR MRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSH EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS TDSGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIIHKNSR IFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPY QLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENS KQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQ TQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVK NEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDI EQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSS IFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQK ANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTG KTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDS MDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTLYA SRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALI IAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTL

KSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYT
QDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINDKGKEVPCN
PFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVS
PWRADVYFNKTTGKYEILGLKYADLQFDKGTGTYKISQEKYNDIKKKEGVDSD
SEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGE
ALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFPKK
KRKVEGADKRTADGSEFESPKKKRKV,
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKV,

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIII
ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVS
SGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRNEHLEVHTRT
HTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLK
THLRGSSAQ, or

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA

LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
IELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVS
SGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRGEHLRQHTRT
HTGEKPFQCRICMRNFSQSGTLKRHLRTHTGEKPFQCRICMRNFSRNDKLVPHL
KTHLRGSSAQ.

104. The modified cell of any one embodiments 63-103, wherein the guide polynucleotide comprises two individual polynucleotides, wherein the two individual polynucleotides are two DNAs, two RNAs or a DNA and an RNA.

105. The modified cell of any one of embodiments 63-104, wherein the guide polynucleotides comprise a crRNA and a tracrRNA, wherein the crRNA comprises a nucleic acid sequence complementary to a target sequence in the *MECP2* polynucleotide.

106. The modified cell of embodiment 105, wherein the target sequence comprises sequence selected from the group consisting of CCATGTCCAGCCTTCAGGCA,

TCCATGTCCAGCCTTCAGGC, TTCCATGTCCAGCCTTCAGG, GCTTCCATGTCCAGCCTTCA, CTTCCATGTCCAGCCTTCAGG, AGCTTCCATGTCCAGCCTTCAGG, AGCTTCCATGTCCAGCCTTC, CTTAAGCTTCCATGTCCAGC, AGCAAAAGGCTTTTCCCTGG, GAGCAAAAGGCTTTTCCCTG, AGAGCAAAAGGCTTTTCCCT, TAGAGCAAAAGGCTTTTCCC, TAGAGCAAAAGGCTTTTCCCT, TTTAGAGCAAAAGGCTTTTCCCT, CCATGAAGTCAAAATCATTA, ACCATGAAGTCAAAATCATT, TACCATGAAGTCAAAATCAT, TTACCATGAAGTCAAAATCAT, AGTTACCATGAAGTCAAAATCAT, CTTTTCACTTCCTGCCGGGG, TCAGCCCCGTTTCTTGGGAA, GCTTTCAGCCCCGTTTCTTG, CGGCTTTCAGCCCCGTTTCTT, CCCGGCTTTCAGCCCCGTTTCTT, GCACTTCTTGATGGGGAGTA, TCTTGCACTTCTTGATGGGG, CTTGCACTTCTTGATGGGGAG, and GTCTTGCACTTCTTGATGGGGAG.

107. The modified cell of embodiment 105 or 106, wherein the base editor system comprises a single guide RNA (sgRNA).

108. The modified cell of embodiment 107, wherein the sgRNA comprises CUUUUCACUUCCUGCCGGGG.

109. A base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of Rett Syndrome (RTT) in a *MECP2* polynucleotide.

110. A base editor system comprising

a guide polynucleotide or a nucleic acid encoding the guide polynucleotide;

a polynucleotide programmable DNA binding domain or a nucleic acid encoding the polynucleotide programmable DNA binding domain, and

an adenosine deaminase domain or a nucleic acid encoding the adenosine deaminase domain,

wherein the guide polynucleotide is capable of targeting the base editor system to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) causative of Rett Syndrome (RTT) in a *MECP2* polynucleotide,

wherein the MECP2 polypeptide encodes a MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

111. The base editor system of embodiment 109, wherein *MECP2* polynucleotide encodes an MECP2 protein comprising a pathogenic amino acid comprising a tryptophan at position 106, a cysteine at position 133, a stop codon at position 255, a stop codon at position 270, or a cysteine at position 306 resulted from the SNP.

112. The base editor system of embodiment 111, wherein the A•T to G•C alteration substitutes the pathogenic amino acid with a wild type amino acid.

113. The base editor system of any one of embodiments 110-112, wherein the wild type amino acid is an arginine.

114. The base editor system of embodiment 113, wherein the the A•T to G•C alteration substitutes a stop codon at position 255 of the MECP2 protein with an arginine

115. The base editor system of embodiment 114, wherein the SNP is at position 763 of the *MECP2* polynucleotide.

116. The base editor system of any one of embodiments 109-115, wherein the polynucleotide programmable DNA binding domain is a Cas9 domain.

117. The base editor system of embodiment 116, wherein the Cas9 domain is a nuclease inactive Cas9 domain.

118. The base editor system of embodiment 117, wherein the Cas9 domain is a Cas9 nickase domain.

119. The method of any one of embodiments 116-118, wherein the Cas9 domain comprises a SpCas9 domain

120. The base editor system of embodiment 119, wherein the SpCas9 domain comprises a D10A and/or a H840A amino acid substitution or corresponding amino acid substitutions thereof.

121. The base editor system of embodiment 119 or 120, wherein the SpCas9 domain has specificity for a NGG PAM.

122. The base editor system of any one of embodiments 119-121, wherein the SpCas9 domain has specificity for a NGA PAM, a NGT PAM, or a NGC PAM.

123. The base editor system of any one of embodiments 119-121, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337I, T1337V, T1337F, and T1337M or corresponding amino acid substitutions thereof.

124. The base editor system of any one of embodiments 119-121, wherein the SpCas9 domain comprises amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

125. The base editor system of any one of embodiments 119-121, wherein the SpCas9 domain comprises amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, T1337, and A1322R, and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M or corresponding amino acid substitutions thereof.

126. The base editor system of any one of embodiments 119-121, wherein the SpCas9 domain comprises amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R, or corresponding amino acid substitutions thereof.

127. The base editor system of any one of embodiments 119 or 120, wherein the SpCas9 domain has specificity for a NG PAM, a NNG PAM, a GAA PAM, a GAT PAM, or a CAA PAM.

128. The base editor system of embodiment 127, wherein the Cas9 domain comprises amino acid substitutions E480K, E543K, and E1219V or corresponding amino acid substitutions thereof.

129. The base editor system of any one of embodiments 116-118, wherein the Cas9 domain comprises a SaCas9 domain.

130. The base editor system of embodiment 129, wherein the SaCas9 domain has specificity for a NNNRRT PAM.

131. The base editor system of embodiment 130, wherein the Sa Cas9 domain has specificity for a NNGRRT PAM.

132. The base editor system of any one of embodiments 129-131, wherein the SaCas9 domain comprises an amino acid substitution N579A or a corresponding amino acid substitution thereof.

133. The base editor system of any one of embodiments 129-132, wherein the SaCas9 domain comprises amino acid substitutions E782K, N968K, and R1015H, or corresponding amino acid substitutions thereof.

134. The base editor system of any one of embodiments 116-118, wherein the Cas9 domain comprises a St1Cas9 domain.

135. The base editor system of embodiment 134, wherein the St1Cas9 domain has specificity for a NNACCA PAM.

136. The base editor system of any one of embodiments 109-135, wherein the adenosine deaminase domain is a modified adenosine deaminase domain that does not occur in nature.

137. The base editor system of embodiment 136, wherein the adenosine deaminase domain comprises a TadA domain.

138. The base editor system of embodiment 137, wherein the TadA domain comprises the amino acid sequence of TadA 7.10..

139. The base editor system of any one of embodiments 109-138, wherein the base editor system further comprises a zinc finger domain.

140. The base editor system of embodiment 139, wherein the zinc finger domain comprises recognition helix sequences RNEHLEV, QSTTLKR, and RTEHLAR or recognition helix sequences RGEHLRQ, QSGTLKR, and RNDKLVP.

141. The base editor system of embodiment 139 or 140, wherein the zinc finger domain is zf1ra or zf1rb.

142. The base editor system of any one of embodiments 109-141, wherein the base editor system further comprises a nuclear localization signal (NLS).

143. The base editor system of any one embodiments 109-142, wherein the base editor system further comprises one or more linkers.

144. The base editor system of embodiment 143, wherein two or more of the polynucleotide programmable DNA binding domain, the adenosine deaminase domain, the zinc finger domain, and the NLS are connected via a linker.

145. The base editor system of embodiment 144, wherein the linker is a peptide linker, thereby forming a base editing fusion protein.

146. The base editor system of embodiment 145, wherein the peptide linker comprises an amino acid sequence selected from the group consisting of SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS, SGGSSGGSSGSETPGTSESATPES, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESS GGSSGGS, PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEG SAP GTSTEPSEGSAPGTSESATPESGPGSEPATS, (SGGS)n, (GGGS)n, (GGGGS)n, (G)n, (EAAAK)n, (GGS)n, SGSETPGTSESATPES, and (XP)n.

147. The base editor system of embodiment 145 or 146, wherein the base editing fusion protein comprises the amino acid sequence selected from the group consisting of MPKKKRKVSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEG WNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHS RIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFR MRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSH EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSS TDSGGSSGGSSGSETPGTSESATPESDLVLGLAIGIGSVGVGILNKVTGEIIHKNSR IFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPY QLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENS KQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQ TQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGI LIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVK NEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDI EQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSS IFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQK ANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTG KTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDS MDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTLYA

SRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALI
IAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTL
KSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYT
QDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINDKGKEVPCN
PFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVS
PWRADVYFNKTTGKYEILGLKYADLQFDKGTGTYKISQEKYNDIKKKEGVDSD
SEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGE
ALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFPKK
KRKVEGADKRTADGSEFESPKKKRKV,
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS

FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKV,
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG
ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVS
SGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRNEHLEVHTRT
HTGEKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRTEHLARHLK
THLRGSSAQ, or
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFG

ARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKA
QKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVL
HYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGS
SGSETPGTSESATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVR
LFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINP
YEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNS
KALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEIL
VNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSS
EDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIF
NRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDII
ELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFIN
RNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK
GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNN
LNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIAS
FYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGEGADKRTADGSEFESPKKKRKVS
SGNSNANSRGPSFSSGLVPLSLRGSHSRPGERPFQCRICMRNFSRGEHLRQHTRT
HTGEKPFQCRICMRNFSQSGTLKRHLRTHTGEKPFQCRICMRNFSRNDKLVPHL
KTHLRGSSAQ.

148. The base editor system of any one embodiments 109-147, wherein the guide polynucleotide comprises two individual polynucleotides, wherein the two individual polynucleotides are two DNAs, two RNAs or a DNA and an RNA.

149. The base editor system of any one of embodiments 109-148, wherein the guide polynucleotides comprise a crRNA and a tracrRNA, wherein the crRNA comprises a nucleic acid sequence complementary to a target sequence in the *MECP2* polynucleotide.

150. The base editor system of embodiment 149, wherein the target sequence comprises sequence selected from the group consisting of CCATGTCCAGCCTTCAGGCA, TCCATGTCCAGCCTTCAGGC, TTCCATGTCCAGCCTTCAGG, GCTTCCATGTCCAGCCTTCA, CTTCCATGTCCAGCCTTCAGG, AGCTTCCATGTCCAGCCTTCAGG, AGCTTCCATGTCCAGCCTTC, CTTAAGCTTCCATGTCCAGC, AGCAAAAGGCTTTTCCCTGG, GAGCAAAAGGCTTTTCCCTG, AGAGCAAAAGGCTTTTCCCT, TAGAGCAAAAGGCTTTTCCC, TAGAGCAAAAGGCTTTTCCCT, TTTAGAGCAAAAGGCTTTTCCCT, CCATGAAGTCAAAATCATTA, ACCATGAAGTCAAAATCATT, TACCATGAAGTCAAAATCAT, TTACCATGAAGTCAAAATCAT, AGTTACCATGAAGTCAAAATCAT, CTTTTCACTTCCTGCCGGGG, TCAGCCCCGTTTCTTGGGAA, GCTTTCAGCCCCGTTTCTTG, CGGCTTTCAGCCCCGTTTCTT, CCCGGCTTTCAGCCCCGTTTCTT, GCACTTCTTGATGGGGAGTA, TCTTGCACTTCTTGATGGGG, CTTGCACTTCTTGATGGGGAG, and GTCTTGCACTTCTTGATGGGGAG.

151. The base editor system of embodiment 149 or 150, wherein the base editor system comprises a single guide RNA (sgRNA).

152. The base editor system of embodiment 151, wherein the sgRNA comprises CUUUUCACUUCCUGCCGGGG.

EXAMPLES

[0449] These examples are provided for illustrative purposes only and not to limit the scope of the claims provided herein.

EXAMPLE 1. A•T to G•C DNA base editing for correction of *Mecp2* RTT-associated mutations in cells

[0450] Six of the eight most prevalent RTT-causing Mecp2 mutations were targeted for reversion to wild-type sequence using A•T to G•C DNA base editors (ABEs) that employ Cas9 moieties with validated protospacer adjacent motif (PAM) sequence preferences. To determine which guide RNA (gRNA) and ABE-Cas9 platform is able to most efficiently and precisely correct a targeted Mecp2 mutation, an Mecp2 allele bearing RTT targetable mutations (Table 66), including R255X was genomically integrated in HEK293T cells by lentivirus transduction. Editing efficiencies of gRNAs and ABE-Cas9 editors for a given mutation were measured. Five days after transfection of DNA encoding ABE-Cas9 editors and gRNAs, cells were lysed and analyzed for base editing at the desired site by miSeq analysis.

[0451] Table 6: 6 RTT mutations.

Causal mutation	Target sequence	PAM		Cas9 PAM-enabling mutations	editing position
c.316C>T	CCATGTCCAGCCTTCAGGCA	GGG	SpCas9		3
(2.77%)	TCCATGTCCAGCCTTCAGGC	AGG	SpCas9		4
R106W	TTCCATGTCCAGCCTTCAGG	CAG	SpCas9	xCas9	5
	GCTTCCATGTCCAGCCTTCA	GGC	SpCas9	MQKSER/ICKSER/VRKSER; Nureki	7
	CTTCCATGTCCAGCCTTCAGG	CAGGGT	SaCas9		6
	AGCTTCCATGTCCAGCCTTCAGG	CAGGGT	SaCas9		8
	AGCTTCCATGTCCAGCCTTC	CTTA	Cpf1		8
	CTTAAGCTTCCATGTCCAGC	TTTG	Cpf1		12
c.397C>T	AGCAAAAGGCTTTTCCCTGG	GGA	SpCas9	VRQR	4
(4.56%)	GAGCAAAAGGCTTTTCCCTG	GGG	SpCas9		5
R133C	AGAGCAAAAGGCTTTTCCCT	GGG	SpCas9		6
	TAGAGCAAAAGGCTTTTCCC	TGG	SpCas9		7
	TAGAGCAAAAGGCTTTTCCCT	GGGGAT	SaCas9		7
	TTTAGAGCAAAAGGCTTTTCCCT	GGGGAT	SaCas9		9
c.473C>T	CCATGAAGTCAAAATCATTA	GGG	SpCas9		3
(8.81%)	ACCATGAAGTCAAAATCATT	AGG	SpCas9		4
T158M	TACCATGAAGTCAAAATCAT	TAG	SpCas9	xCas9; ScCas9	5
	TTACCATGAAGTCAAAATCAT	TAGGGT	SaCas9		6
	AGTTACCATGAAGTCAAAATCAT	TAGGGT	SaCas9		8
c.763C>T (6.68%) R255X	CTTTTCACTTCCTGCCGGGG	CGT	SpCas9	LRSVQL/LRKIQK/LRSVQK; Nureki	7
c.808C>T	TCAGCCCCGTTTCTTGGGAA	TGG	SpCas9		3
(5.8%)	GCTTTCAGCCCCGTTTCTTG	GGA	SpCas9	VRQR	7
R270X	CGGCTTTCAGCCCCGTTTCTT	GGGAAT	SaCas9		9
	CCCGGCTTTCAGCCCCGTTTCTT	GGGAAT	SaCas9		11
c.916C>T (5.17%)	GCACTTCTTGATGGGGAGTA	CGG	SpCas9		3
R306C	TCTTGCACTTCTTGATGGGG	AGT	SpCas9	LRSVQL/LRKIQK/LRSVQK	7
	CTTGCACTTCTTGATGGGGAG	TACGGT	SaCas9	KKH	6
	GTCTTGCACTTCTTGATGGGGAG	TACGGT	SaCas9	KKH	8

[0452] In one set of experiments, R255X was targeted using a gRNA comprising the nucleic acid sequence CUUUUCACUUCCUGCCGGGG, which targets the base editor to the nucleic

acid sequence CTTTTCACTTCCTGCCGGGG comprising the RTT mutation 763C>T (R255X).

[0453] Adenosine deaminase base editors were created with specificity for NGT PAM having amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337I, T1337V, T1337F, and T1337M in SpCas9. The base editors showed activity as measured by precise correction of R255X.

[0454] Adenosine deaminase base editors were created with specificity for NGT PAM having amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M in SpCas9. The base editors showed activity as measured by precise correction of R255X (FIG. 1).

[0455] Adenosine deaminase base editors were created with specificity for NGT PAM having amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337 and, and one or more of L1111R, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M in SpCas9. The base editors showed activity as measured by precise correction of R255X (FIG. 2). pCAG promoter was used for the data on the left (light gray) and pCMV promoter was used for the data on the right (dark gray).

[0456] Adenosine deaminase base editors with specificity for NGT PAM were generated as shown in Table 7.

Table 7. NGT PAM variant

	NGTN variant	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
Variant 1	LRKIQK	L	R	K	I	-	Q	K
Variant 2	LRSVQK	L	R	S	V	-	Q	K
Variant 3	LRSVQL	L	R	S	V	-	Q	L
Variant 4	LRKIRQK	L	R	K	I	R	Q	K
Variant 5	LRSVRQK	L	R	S	V	R	Q	K
Variant 6	LRSVRQL	L	R	S	V	R	Q	L

[0457] The base editors showed activity as measured by precise correction of R255X (FIG. 3). In particular, variant 6, also termed LRSVRQL showed 4% base editing, the most of all the

variants. Variant 6 contains A1322R and T1337L. Without being bound by theory, it was proposed that these amino acid substitutions were important for NGT PAM specificity.

[0458] Additional mutations from other characterized PAM variants were introduced into PAM variant 6. Amino acid residue T1337 was identified as being important for editing (FIG. 4). PAM variants based on variant 6 and modified to have T1337Q and T1337 showed increased editing for correction of R255X, compared to T1337L. A PAM variant based on variant 6 and modified to have D1332A was also identified as having increased editing for correction of R255X compared to variant 6.

[0459] To characterize the residues at T1337 and D1332 further, different amino acids were tested at these two positions with regard to editing efficiency for correction of R255X. Compared to other substitutions at position 1337, T1337 and T1337Q were the most important residues for editing (FIG. 5).

[0460] To evaluate T1337 specifically, individual amino acids in the PAM variants were reverted to wild-type. Without being bound by theory, this analysis can identify which residues are important for activity on NGT PAM. In particular, E1219V and R1335Q were shown to be important for activity associated with T1337, as reversion back to wild-type at each of these residues abolished significantly reduced editing (FIG. 6).

[0461] Molecular modeling was also used to yield a library of pairwise mutations for improved NGT PAM recognition. Amino acids were shuffled to create 55 new variants of PAM variant 6. Groups of mutation are depicted in Tables 8 and 9 below:

Table 8: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

Table 9: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

[0001] Targeted mutations for the 55 new PAM variants in Tables 8 and 9 reinforced the importance of T1337 (FIGS. 7 and 8). Variant 5 in Table 8 displayed significantly enhanced editing (FIG. 7). In particular, the sequence of variant 5 (Table 8) includes a reversion to T1337.

[0002] Additional pairwise mutations were generated for improved NGT PAM recognition, as shown in Table 10.

Table 10: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

[0003] Additional pairwise mutations were generated for improved NGT PAM recognition, as shown in Table 10. The substitution G1218R was additive for editing with E1219F, R1335V, and T1337 or T1337Q (FIG. 9).

EXAMPLE 2. Materials and Methods

The results provided in the Examples described herein were obtained using the following materials and methods.

Cloning.

[0462] PCR was performed using VeraSeq ULtra DNA polymerase (Enzymatics), or Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs). Base Editor (BE) plasmids were constructed using USER cloning (New England Biolabs). Deaminase genes were synthesized as gBlocks Gene Fragments (Integrated DNA Technologies). Cas9 genes used are listed below. Cas9 genes were obtained from previously reported plasmids. Deaminase and fusion genes were cloned into pCMV (mammalian codon-optimized) or pET28b (*E. coli* codon-optimized) backbones. sgRNA expression plasmids were constructed using site-directed mutagenesis.

[0463] Briefly, the primers listed herein above were 5′ phosphorylated using T4 Polynucleotide Kinase (New England Biolabs) according to the manufacturer's instructions. Next, PCR was performed using Q5 Hot Start High- Fidelity Polymerase (New England Biolabs) with the phosphorylated primers and the plasmid encoding Mecp2 as a template according to the manufacturer's instructions. PCR products were incubated with DpnI (20 U, New England Biolabs) at 37 °C for 1 hour, purified on a QIAprep spin column (Qiagen), and ligated using QuickLigase (New England Biolabs) according to the manufacturer's instructions. DNA vector amplification was carried out using Mach1 competent cells (ThermoFisher Scientific).

[0464] For gRNAs, the following scaffold sequence is presented: GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC

UUGAAAAAGU GGCACCGAGU CGGUGCUUUU. This scaffold was used for the PAMs shown in the tables herein, e.g., NGG, NGA, NGC, NGT PAMs; the gRNA encompasses the scaffold sequence and the spacer sequence (target sequence) for disease-associated genes (e.g., Tables 3A, 3B and 4) as provided herein or as determined based on the knowledge of the skilled practitioner and as would be understood to the skilled practitioner in the art. (See, e.g., Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., et al., "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., et al., "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), and Rees, H.A., et al., "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1).

In vitro deaminase assay on ssDNA.

[0465] All Cy3-labelled substrates were obtained from Integrated DNA Technologies (IDT). Deaminases were expressed in vitro using the TNT T7 Quick Coupled Transcription/Translation Kit (Promega) according to the manufacturer's instructions using 1 μg of plasmid. Following protein expression, 5 μl of lysate was combined with 35 μl of ssDNA (1.8 μM) and USER enzyme (1 unit) in CutSmart buffer (New England Biolabs) (50 mM potassium acetate, 29 mM Tris-acetate, 10 mM magnesium acetate, 100 μg ml−1 BSA, pH 7.9) and incubated at 37 °C for 2 h. Cleaved U-containing substrates were resolved from full-length unmodified substrates on a 10% TBE-urea gel (Bio-Rad).

Expression and purification of His6–rAPOBEC1-linker–dCas9 fusions.

[0466] *E. coli* BL21 STAR (DE3)-competent cells (ThermoFisher Scientific) were transformed with plasmids encoding pET28b-His6-rAPOBEC1-linker-dCas9 linkers. The resulting expression strains were grown overnight in Luria-Bertani (LB) broth containing 100 μg/ml of kanamycin at 37 °C. The cells were diluted 1:100 into the same growth medium and grown at 37 °C to OD600 = ~0.6. The culture was cooled to 4 °C over a period of 2 h, and isopropyl- β-d-1-thiogalactopyranoside (IPTG) was added at 0.5 mM to induce protein expression. After ~16 h, the cells were collected by centrifugation at 4,000g and were resuspended in lysis buffer (50 mM tris(hydroxymethyl)-aminomethane (Tris)- HCl (pH 7.5), 1 M NaCl, 20% glycerol, 10 mM tris(2-carboxyethyl)phosphine (TCEP, Soltec Ventures)). The cells were lysed by sonication (20 s pulse-on, 20 s pulse-off for 8 min total at 6 W output) and the lysate supernatant was isolated following centrifugation at 25,000g for 15 minutes. The lysate was incubated with His-Pur

nickel-nitriloacetic acid (nickel-NTA) resin (ThermoFisher Scientific) at 4 °C for 1 hour to capture the His-tagged fusion protein. The resin was transferred to a column and washed with 40 ml of lysis buffer. The His-tagged fusion protein was eluted in lysis buffer supplemented with 285 mM imidazole, and concentrated by ultrafiltration (Amicon-Millipore, 100-kDa molecular weight cut-off) to 1 ml total volume. The protein was diluted to 20 ml in low-salt purification buffer containing 50 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl (pH 7.0), 0.1 M NaCl, 20% glycerol, 10 mM TCEP and loaded onto SP Sepharose Fast Flow resin (GE Life Sciences). The resin was washed with 40 ml of this low-salt buffer, and the protein eluted with 5 ml of activity buffer containing 50 mM tris(hydroxymethyl)- aminomethane (Tris)-HCl (pH 7.0), 0.5 M NaCl, 20% glycerol, 10 mM TCEP. The eluted proteins were quantified by SDS–PAGE.

In vitro transcription of sgRNAs.

[0467] Linear DNA fragments containing the T7 promoter followed by the 20-bp sgRNA target sequence were transcribed in vitro with the TranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific) according to the manufacturer's instructions. sgRNA products were purified using the MEGAclear Kit (ThermoFisher Scientific) according to the manufacturer's instructions and quantified by UV absorbance.

Preparation of Cy3-conjugated dsDNA substrates.

[0468] Sequences of 80-nt unlabeled strands were ordered as PAGE-purified oligonucleotides from IDT. The 25-nt Cy3-labelled primer listed in the Supplementary Information is complementary to the 3′ end of each 80-nt substrate. This primer was ordered as an HPLC-purified oligonucleotide from IDT. To generate the Cy3-labelled dsDNA substrates, the 80-nt strands (5 μl of a 100 μM solution) were combined with the Cy3-labelled primer (5 μl of a 100 μM solution) in NEBuffer 2 (38.25 μl of a 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9 solution, New England Biolabs) with dNTPs (0.75 μl of a 100 mM solution) and heated to 95°C for 5 min, followed by a gradual cooling to 45°C at a rate of 0.1 °C per s. After this annealing period, Klenow exo–(5 U, New England Biolabs) was added and the reaction was incubated at 37°C for 1 h. The solution was diluted with buffer PB (250 μl, Qiagen) and isopropanol (50 μl) and purified on a QIAprep spin column (Qiagen), eluting with 50 μl of Tris buffer. Deaminase assay on dsDNA. The purified fusion protein (20 μl of 1.9 μM in activity buffer) was combined with 1 equivalent of appropriate sgRNA and incubated at ambient temperature for 5 min. The Cy3-labelled dsDNA substrate was added to final concentration of 125 nM and the resulting solution was incubated at 37 °C for 2 h. The dsDNA was separated from the fusion by the addition of buffer PB (100 μl, Qiagen) and isopropanol (25 μl) and purified on a EconoSpin micro spin column (Epoch Life Science), eluting with 20 μl of

CutSmart buffer (New England Biolabs). USER enzyme (1 U, New England Biolabs) was added to the purified, edited dsDNA and incubated at 37 °C for 1 h. The Cy3-labeled strand was fully denatured from its complement by combining 5 μl of the reaction solution with 15 μl of a DMSO-based loading buffer (5 mM Tris, 0.5 mM EDTA, 12.5% glycerol, 0.02% bromophenol blue, 0.02% xylene cyan, 80% DMSO). The full-length C-containing substrate was separated from any cleaved, U-containing edited substrates on a 10% TBE-urea gel (Bio-Rad) and imaged on a GE Amersham Typhoon imager.

Preparation of *in vitro*-edited dsDNA for high-throughput sequencing.

[0469] The oligonucleotides listed below were obtained from IDT. Complementary sequences were combined (5 μl of a 100 μM solution) in Tris buffer and annealed by heating to 95 °C for 5 min, followed by a gradual cooling to 45 °C at a rate of 0.1°C per s to generate 60-bp dsDNA substrates. Purified fusion protein (20 μl of 1.9 μM in activity buffer) was combined with 1 equivalent of appropriate sgRNA and incubated at ambient temperature for 5 min. The 60-mer dsDNA substrate was added to final concentration of 125 nM, and the resulting solution was incubated at 37 °C for 2 h. The dsDNA was separated from the fusion by the addition of buffer PB (100 μl, Qiagen) and isopropanol (25 μl) and purified on a EconoSpin micro spin column (Epoch Life Science), eluting with 20 μl of Tris buffer. The resulting edited DNA (1 μl was used as a template) was amplified by PCR using the high-throughput sequencing primer pairs provided below: and VeraSeq Ultra (Enzymatics) according to the manufacturer's instructions with 13 cycles of amplification. PCR reaction products were purified using RapidTips (Diffinity Genomics), and the purified DNA was amplified by PCR with primers containing sequencing adapters, purified, and sequenced on a MiSeq high-throughput DNA sequencer (Illumina) as previously described.

Cell culture.

[0470] HEK293T (ATCC CRL-3216) and U2OS (ATCC HTB-96) were maintained in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10% (v/v) fetal bovine serum (FBS), at 37 °C with 5% CO2. HCC1954 cells (ATCC CRL-2338) were maintained in RPMI-1640 medium (ThermoFisher Scientific) supplemented as described above. Immortalized cells containing the Mecp2 gene (Taconic Biosciences) were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS) and 200 μg ml−1 Geneticin (ThermoFisher Scientific).

Transfections.

[0471] HEK293T cells were seeded on 48-well collagen-coated BioCoat plates (Corning) and transfected at approximately 85% confluency. Briefly, 750 ng of BE and 250 ng of sgRNA

expression plasmids were transfected using 1.5 μl of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol. HEK293T cells were transfected using appropriate Amaxa Nucleofector II programs according to manufacturer's instructions (V kits using program Q-001 for HEK293T cells).

High-throughput DNA sequencing of genomic DNA samples.

[0472] Transfected cells were harvested after 3 days and the genomic DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. On-target and off-target genomic regions of interest were amplified by PCR with flanking high-throughput sequencing primer pair. PCR amplification was carried out with Phusion high-fidelity DNA polymerase (ThermoFisher) according to the manufacturer's instructions using 5 ng of genomic DNA as a template. Cycle numbers were determined separately for each primer pair as to ensure the reaction was stopped in the linear range of amplification. PCR products were purified using RapidTips (Diffinity Genomics). Purified DNA was amplified by PCR with primers containing sequencing adaptors. The products were gel purified and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher) and KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described (Pattanayak, Nature Biotechnol. 31, 839–843 (2013)).

Data analysis.

[0473] Sequencing reads were automatically demultiplexed using MiSeq Reporter (Illumina), and individual FASTQ files were analyzed with a custom Matlab. Each read was pairwise aligned to the appropriate reference sequence using the Smith-Waterman algorithm. Base calls with a Q-score below 31 were replaced with Ns and were thus excluded in calculating nucleotide frequencies. This treatment yields an expected MiSeq base-calling error rate of approximately 1 in 1,000. Aligned sequences in which the read and reference sequence contained no gaps were stored in an alignment table from which base frequencies could be tabulated for each locus. Indel frequencies were quantified with a custom Matlab script using previously described criteria (Zuris, et al., *Nature Biotechnol.* 33, 73–80 (2015)). Sequencing reads were scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels might occur. If no exact matches were located, the read was excluded from analysis. If the length of this indel window exactly matched the reference sequence the read was classified as not containing an indel. If the indel window was two or more bases longer or shorter than the reference sequence, then the sequencing read was classified as an insertion or deletion, respectively.

CLAIMS

What is claimed is:

1. A method of editing an *MECP2* polynucleotide comprising a single nucleotide polymorphism (SNP) associated with Rett Syndrome (RTT), the method comprising contacting the *MECP2* polynucleotide with a base editor in complex with one or more guide polynucleotides, wherein the base editor comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain, and wherein one or more of said guide polynucleotides target said base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

2. The method of claim 1, wherein the contacting is in a cell, a eukaryotic cell, a mammalian cell, or human cell.

3. The method of claim 1 or 2, wherein the cell is in vivo.

4. The method of claim 1 or 2, wherein the cell is ex vivo.

5. The method of any one of claims 1-4, wherein the alteration is one or more of R106W, R168*, R133C, T158M, R255*, R270*, and R306C.

6. The method of any one of claims 1-5, wherein the A•T to G•C alteration at the SNP associated with RTT changes a cysteine to an arginine, methionine to a threonine, or stop codon to arginine in the methyl CpG binding protein 2 (Mecp2) polypeptide.

7. The method of any one of claims 1-6, wherein the SNP associated with RTT results in expression of an Mecp2 polypeptide comprising an arginine at amino acid position 168, 133, 255, 270, or 306; or a threonine at position 158.

8. The method of any one of claims 1-7, wherein the polynucleotide programmable DNA binding domain is a *Streptococcus pyogenes* Cas9 (SpCas9) or variant thereof.

9. The method of any one of claims 1-8, wherein the polynucleotide programmable DNA binding domain comprises a modified SpCas9 having an altered protospacer-adjacent motif (PAM) specificity.

10. The method of claim 9, wherein the modified SpCas9 has specificity for the nucleic acid sequence 5'-NGT-3'.

11. The method of claim 10, wherein the modified SpCas9 comprises the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, , T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof.

12. The method of claim 10, wherein the modified SpCas9 comprises the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L,, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof.

13. The method of any one of claims 1-12, wherein the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant.

14. The method of claim 13, wherein the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof.

15. The method of any one of claims 1-14, wherein the adenosine deaminase domain is capable of deaminating adenosine in deoxyribonucleic acid (DNA).

16. The method of claim 15, wherein the adenosine deaminase is a TadA deaminase.

17. The method of claim 15 or 16, wherein the TadA deaminase is TadA*7.10.

18. The method of any one of claims 1-17, wherein the one or more guide RNAs comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), wherein the crRNA comprises a nucleic acid sequence complementary to a *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

19. The method of any one of claims 1-18, wherein the base editor is in complex with a single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

20. A cell produced by introducing into the cell, or a progenitor thereof:

A base editor, or a polynucleotide encoding said base editor, to said cell, wherein said base editor comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain; and

one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

21. The cell of claim 20, wherein the cell is a neuron.

22. The cell of claim 20 or 21, wherein the neuron expresses an Mecp2 polypeptide.

23. The cell of any one of claims 20-22, wherein the cell is from a subject having RTT.

24. The cell of any one of claims 20-23, wherein the cell is a mammalian cell.

25. The cell of any one of claims 20-24, wherein the cell is a human cell.

26. The cell of any one of claims 20-25, wherein the alteration is one or more of R106W, R168*, R133C, T158M, R255*, R270*, and R306C.

27. The cell of any one of claims 20-26, wherein the A•T to G•C alteration at the SNP associated with RTT changes a cysteine to an arginine, methionine to a threonine, or stop codon to arginine in the methyl CpG binding protein 2 (Mecp2) polypeptide.

28. The cell of any one of claims 20-27, wherein the SNP associated with RTT results in expression of an Mecp2 polypeptide comprising an arginine at amino acid position 168, 133, 255, 270, or 306; or a threonine at position 158.

29. The cell of any one of claims 20-28, wherein the polynucleotide programmable DNA binding domain is a *Streptococcus pyogenes* Cas9 (SpCas9) or variant thereof.

30. The cell of any one of claims 20-29, wherein the polynucleotide programmable DNA binding domain comprises a modified SpCas9 having an altered protospacer-adjacent motif (PAM) specificity.

31. The cell of claim 30, wherein the modified SpCas9 has specificity for the nucleic acid sequence 5'-NGT-3'.

32. The cell of claim 31, wherein the modified SpCas9 comprises the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, , T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof.

33. The cell of claim 31, wherein the modified SpCas9 comprises the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof.

34. The cell of any one of claims 20-33, wherein the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant.

35. The cell of claim 34, wherein the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof.

36. The cell of any one of claims 20-35, wherein the adenosine deaminase domain is capable of deaminating adenosine in deoxyribonucleic acid (DNA).

37. The cell of claim 36, wherein the adenosine deaminase is a TadA deaminase.

38. The cell of claim 36 or 37, wherein the TadA deaminase is TadA*7.10.

39. The cell of any one of claims 20-38, wherein the one or more guide RNAs comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), wherein the crRNA comprises a nucleic acid sequence complementary to a *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

40. The cell of any one of claims 20-39, wherein the base editor is in complex with a single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

41. A method of treating RTT in a subject comprising administering to said subject a cell of any one of claims 1-40.

42. A method of treating RTT in a subject comprising administering to said subject:

a base editor, or a polynucleotide encoding said base editor, to said subject, wherein said base editor comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain; and

one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration of the SNP associated with RTT.

43. The method of claim 42, wherein the subject is a mammal or a human.

44. The method of claim 42 or 43, comprising delivering the base editor, or polynucleotide encoding said base editor, and said one or more guide polynucleotides to a cell of the subject.

45. The method of any one of claims 42-44, wherein the cell is a neuron.

46. The method of any one of claims 42-45, wherein the alteration is one or more of R106W, R168*, R133C, T158M, R255*, R270*, and R306C.

47. The method of any one of claims 42-46, wherein the A•T to G•C alteration at the SNP associated with RTT changes a cysteine to an arginine, methionine to a threonine, or stop codon to arginine in the methyl CpG binding protein 2 (Mecp2) polypeptide.

48. The method of any one of claims 42-47, wherein the SNP associated with RTT results in expression of an Mecp2 polypeptide comprising an arginine at amino acid position 168, 133, 255, 270, or 306; or a threonine at position 158.

49. The method of any one of claims 42-48, wherein the polynucleotide programmable DNA binding domain is a *Streptococcus pyogenes* Cas9 (SpCas9) or variant thereof.

50. The method of any one of claims 42-49, wherein the polynucleotide programmable DNA binding domain comprises a modified SpCas9 having an altered protospacer-adjacent motif (PAM) specificity.

51. The method of claim 50, wherein the modified SpCas9 has specificity for the nucleic acid sequence 5'-NGT-3'.

52. The method of claim 51, wherein the modified SpCas9 comprises the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, and T1337M, or corresponding amino acid substitutions thereof.

53. The method of claim 52, wherein the modified SpCas9 comprises the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof.

54. The method of any one of claims 42-53, wherein the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant.

55. The method of claim 54, wherein the nickase variant comprises an amino acid substitution D10A or a corresponding amino acid substitution thereof.

56. The method of any one of claims 42-55, wherein the adenosine deaminase domain is capable of deaminating adenosine in deoxyribonucleic acid (DNA).

57. The method of claim 55 or 56, wherein the adenosine deaminase is a TadA deaminase.

58. The method of any one of claims 55-57, wherein the TadA deaminase is TadA*7.10.

59. The method of any one of claims 55-58, wherein the one or more guide RNAs comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), wherein the crRNA comprises a nucleic acid sequence complementary to a *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

60. The method of any one of claims 55-59, wherein the base editor is in complex with a single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an *Mecp2* nucleic acid sequence comprising the SNP associated with RTT.

61. A base editor comprising:

(i) a modified SpCas9 comprising the amino acid substitutions L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R and one or more of L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, , T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof; and

(ii) an adenosine deaminase.

62. A base editor comprising:

(i) a modified SpCas9 comprising the amino acid substitutions D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337, and one or more of L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I,

T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereof; and

(ii) an adenosine deaminase.

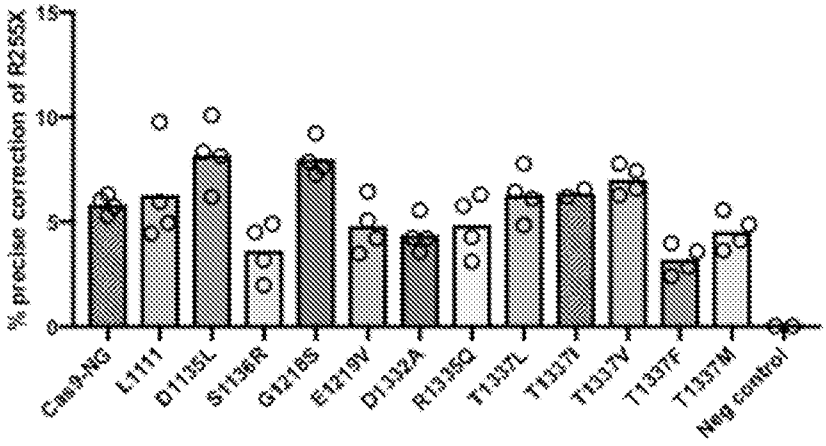
% precise correction of R255X
15
10
5
0
Cas9-NG
L1111
D1135L
S1136R
G1218S
E1219V
D1332A
R1335Q
T1337L
T1337I
T1337V
T1337F
T1337M
Neg control

FIG. 1

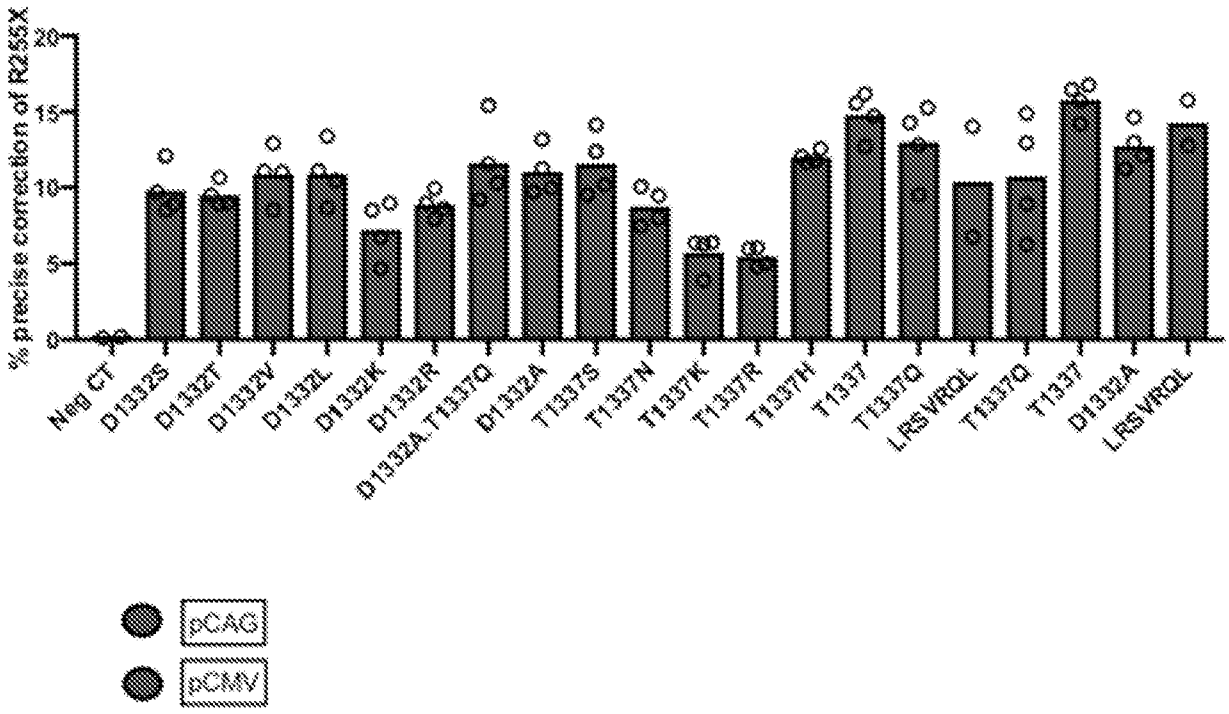
% precise correction of R255X
20
15
10
5
0
Neg CT
D1332S
D1332T
D1332V
D1332L
D1332K
D1332R
D1332A.T1337Q
D1332A
T1337S
T1337N
T1337K
T1337R
T1337H
T1337
T1337Q
LRSVRQL
T1337Q
T1337
D1332A
LRSVRQL
pCAG
pCMV

FIG. 2

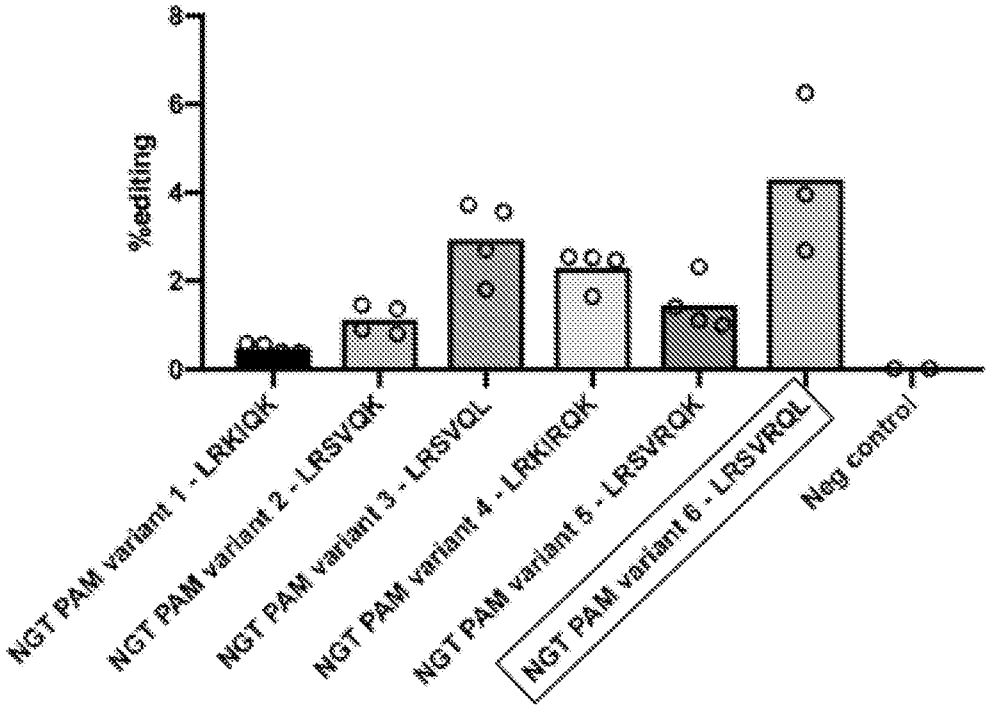
%editing
0
2
4
6
8
NGT PAM variant 1 - LRKIQK
NGT PAM variant 2 - LRSVQK
NGT PAM variant 3 - LRSVQL
NGT PAM variant 4 - LRKIRQK
NGT PAM variant 5 - LRSVRQK
NGT PAM variant 6 - LRSVRQL
Neg control

FIG. 3

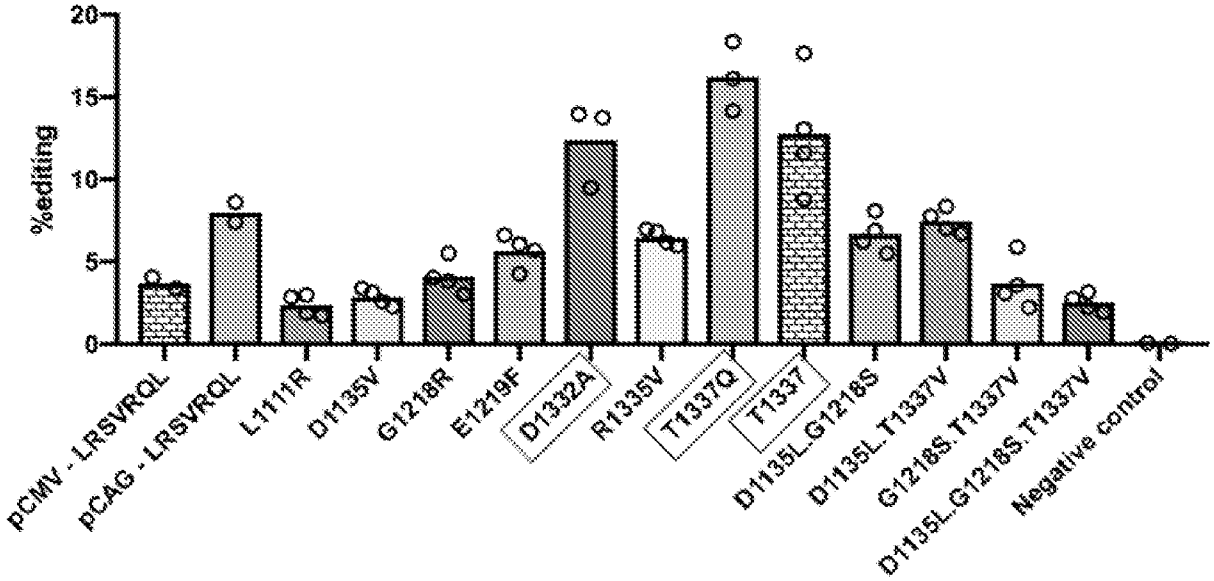
%editing
20
15
10
5
0
pCMV - LRSVRQL
pCAG - LRSVRQL
L1111R
D1135V
G1218R
E1219F
D1332A
R1335V
T1337Q
T1337
D1135L.G1218S
D1135L.T1337V
G1218S.T1337V
D1135L.G1218S.T1337V
Negative control

FIG. 4

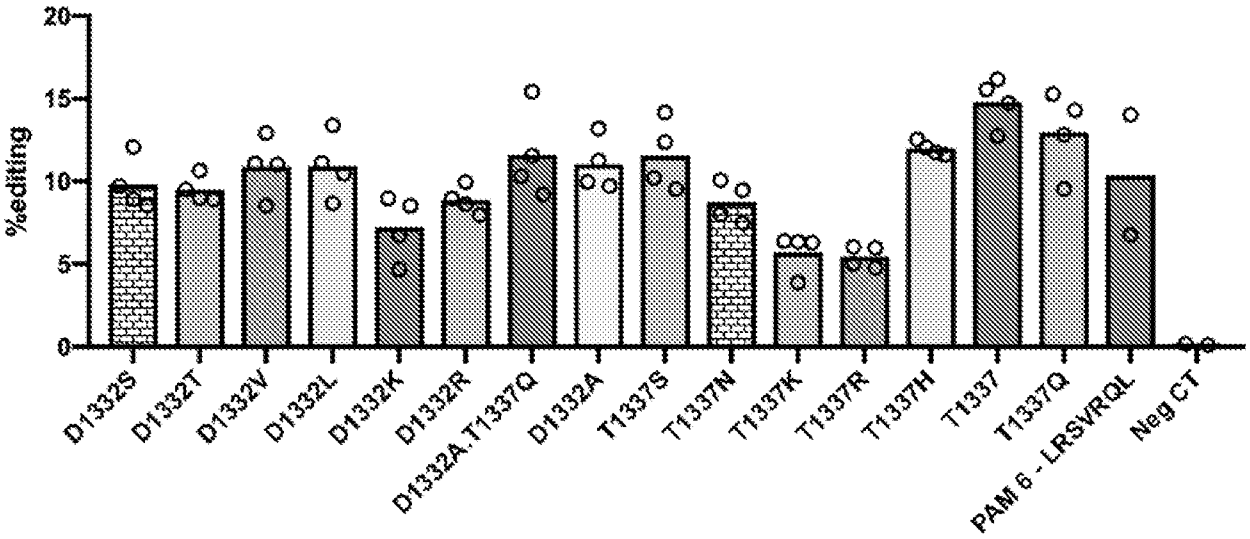
%editing
20
15
10
5
0
D1332S
D1332T
D1332V
D1332L
D1332K
D1332R
D1332A.T1337Q
D1332A
T1337S
T1337N
T1337K
T1337R
T1337H
T1337
T1337Q
PAM 6 - LRSVRQL
Neg CT

FIG. 5

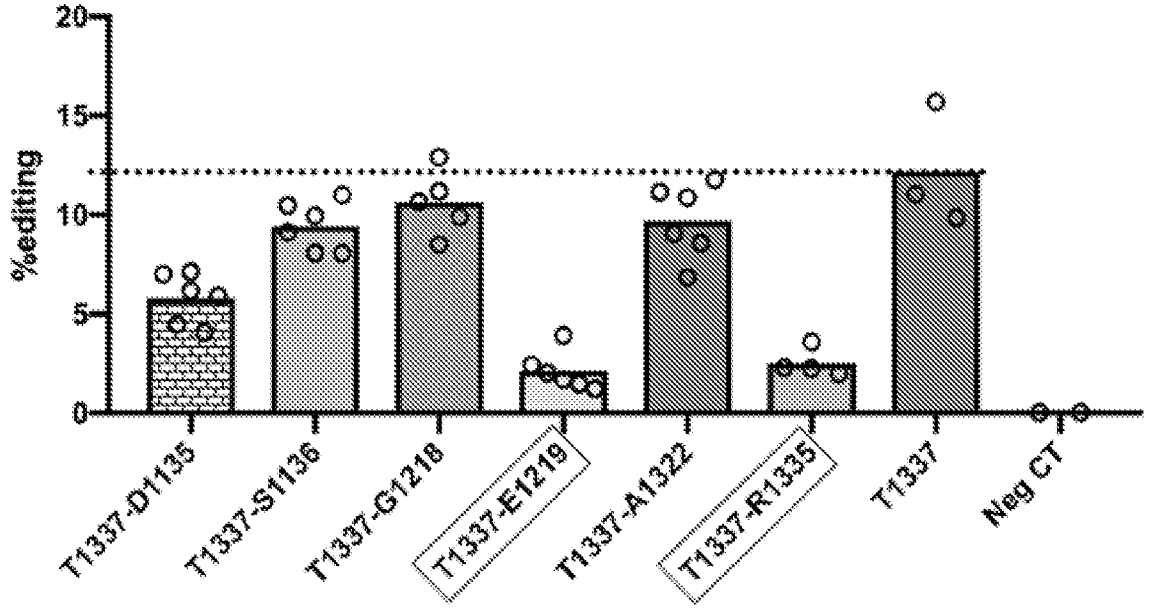
20
15
10
5
0
%editing
T1337-D1135
T1337-S1136
T1337-G1218
T1337-E1219
T1337-A1322
T1337-R1335
T1337
Neg CT

FIG. 6

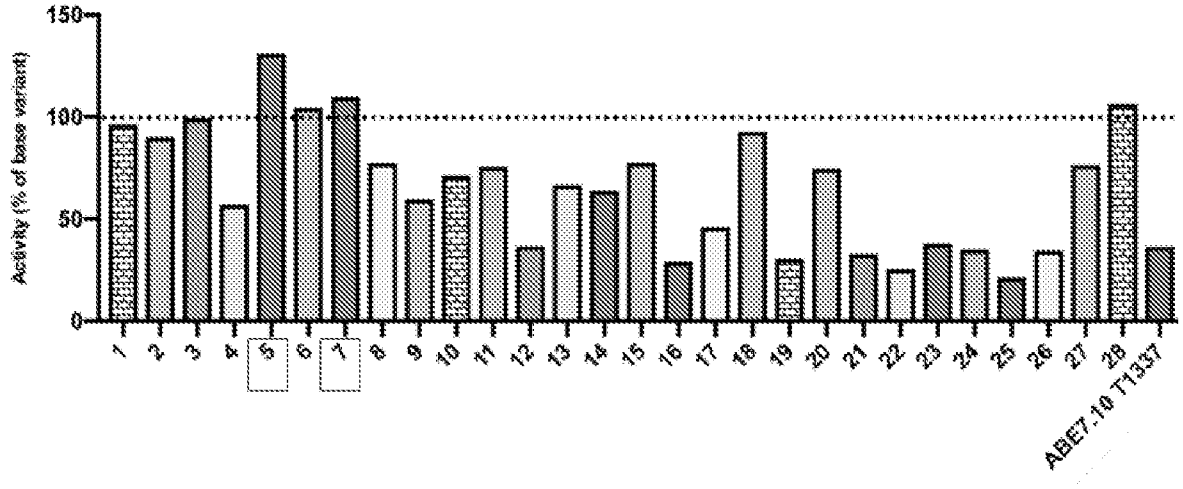
Activity (% of base variant)
150
100
50
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
ABE7.10 T1337

FIG. 7

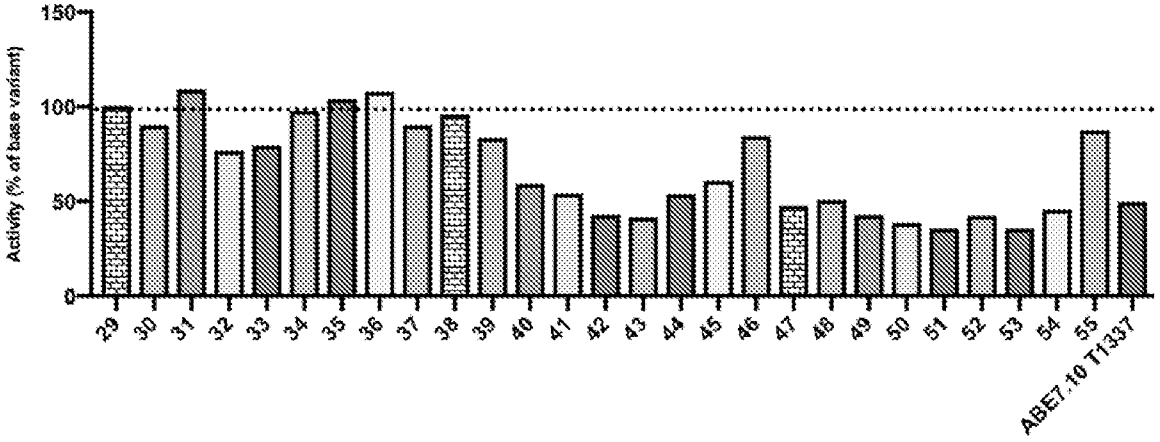
Activity (% of base variant)
150
100
50
0
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
ABE7.10 T1337

FIG. 8

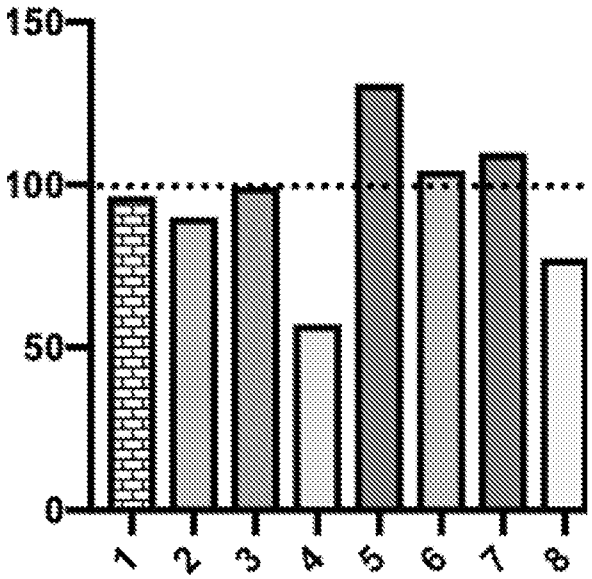

FIG. 9

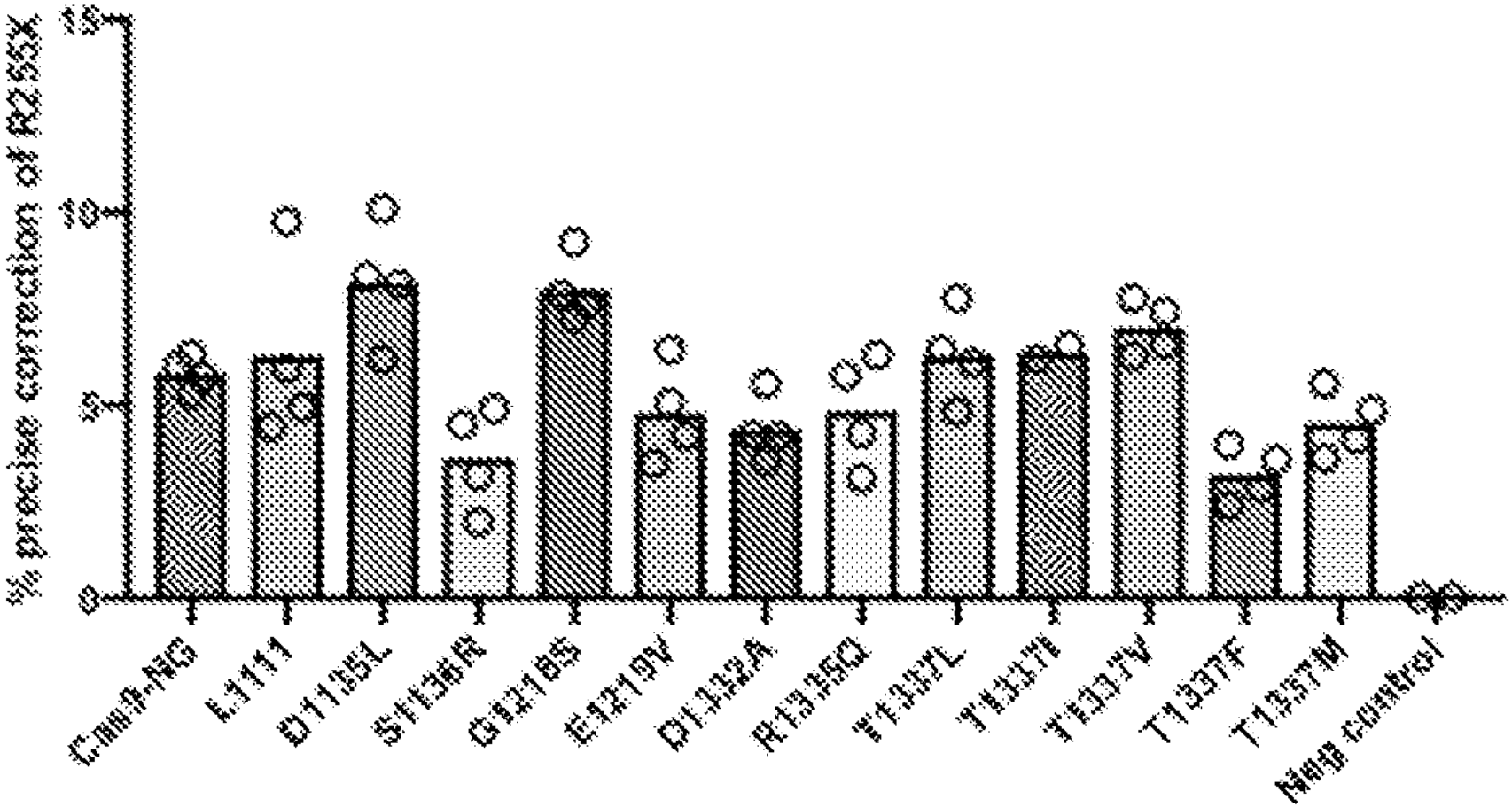
% precise correction of R255X
15
10
5
0
Cas9-NG
L1111
D1135L
S1136R
G1218S
E1219V
D1332A
R1335Q
T1337L
T1337I
T1337V
T1337F
T1337M
Neg control

FIG. 1