Claims

1. A fusion protein comprising: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; (iii) a first uracil glycosylase inhibitor (UGI) domain; and (iv) a second uracil glycosylase inhibitor (UGI) domain.

2. The fusion protein of claim 1, wherein the fusion protein comprises the structure:

NH_2-[cytidine deaminase domain]-[napDNAbp]-[first UGI domain]-[second UGI domain]-COOH;

NH_2-[first UGI domain]-[second UGI domain]-[cytidine deaminase domain]-[napDNAbp]-COOH;

NH_2-[napDNAbp]-[cytidine deaminase domain]-[first UGI domain]-[second UGI domain]-COOH; or

NH_2-[first UGI domain]-[second UGI domain]-[napDNAbp]-[cytidine deaminase domain]-COOH;

wherein each instance of "]-[" comprises an optional linker.

3. The fusion protein of claim 1 or 2 further comprising (v) a nuclear localization sequence (NLS).

4. The fusion protein of claim 3, wherein the NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 614), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 615), or KRTADGSEFEPKKKRKV (SEQ ID NO: 740).

5. The fusion protein of any one of claims 1-4, wherein at least one of (1) the cytidine deaminase domain and the napDNAbp are linked via a linker, (2) the first UGI domain and the NLS are linked via a linker, (3) the second UGI domain and the NLS are linked via a linker, (4) the napDNAbp and the first UGI domain are linked via a linker, (5) the napDNAbp and the second UGI domain are linked via a linker, and (6) the first UGI domain and the second UGI domain are linked via a linker.

6. The fusion protein of claim 5, wherein any one of the linkers comprises the amino acid sequence: $(GGGS)_n$ (SEQ ID NO: 613), $(GGGGS)_n$ (SEQ ID NO: 607), $(G)_n$ (SEQ ID NO: 608), $(EAAAK)_n$ (SEQ ID NO: 609), $(GGS)_n$ (SEQ ID NO: 610), $(SGGS)_n$ (SEQ ID NO: 606), SGSETPGTSESATPES (SEQ ID NO: 604), $SGGS(GGS)_n$ (SEQ ID NO: 612), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605), $(XP)_n$ (SEQ ID NO: 611), or a combination thereof, wherein n is independently an integer between 1 and 30, inclusive, and wherein X is any amino acid.

7. The fusion protein of any one of claims 1-6, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is a CasX, a CasY, a Cpf1, a Cpf1 nickase, a dCpf1, a C2c1, a C2c2, a C2c3, a Cas9, a dCas9, a Cas9 nickase, or an Argonaute protein.

8. The fusion protein of any one of claims 3-7, wherein the fusion protein comprises the structure: NH_2-[cytidine deaminase domain]-[napDNAbp]-[first UGI domain]-[second UGI domain]-[NLS]-COOH, wherein each instance of "]-[" comprises an optional linker.

9. The fusion protein of any one of claims 3-7, wherein the second UGI domain and the NLS are linked via a linker comprising the amino acid sequence: SGGS (SEQ ID NO: 606).

10. The fusion protein of any one of claims 3-9, wherein the napDNAbp domain and the first UGI domain are linked via a linker comprising the amino acid sequence: SGGS (SEQ ID NO: 606).

11. The fusion protein of any one of claims 1-10, wherein the cytidine deaminase domain and the napDNAbp are linked via a linker comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 605).

12. The fusion protein of any one of claims 1-11, wherein the napDNAbp and the first UGI domain are linked via a linker comprising the amino acid sequence: $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2.

13. The fusion protein of any one of claims 1-12, wherein the first UGI domain and the second UGI domain are linked via a linker comprising the amino acid sequence: $SGGS(GGS)_n$ (SEQ ID NO: 612), wherein n is 2.

14. A fusion protein comprising: (i) a nucleic acid programmable DNA binding protein (napDNAbp); (ii) a cytidine deaminase domain; and (iii) a uracil glycosylase inhibitor (UGI) domain, wherein the napDNAbp is a CasX protein, a CasY protein, a Cpf1 protein, a C2c1 protein, a C2c2 protein, a C2c3 protein, or an Argonaute protein.

15. The fusion protein of claim 14, wherein the fusion protein comprises two, three, four, or five UGI domains.

16. The fusion protein of claim 14 or 15, wherein each of the UGI domains comprises an amino acid sequence that is 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 SEQ ID NO: 134.

17. The fusion protein of claim 14 or 15, wherein each of the UGI domains comprises an amino acid sequence as set forth in SEQ ID NO: 134.

18. The fusion protein of any one of claims 14-17, wherein the nucleic acid programmable DNA binding protein (napDNAbp) is:

(a) a CasY protein;

(b) a Cpf1 protein or a Cpf1 mutant protein;

(c) a C2c1 protein;

(d) a C2c2 protein;

(e) a C2c3 protein;

(f) an Argonaute protein; or

(g) a CasX protein.

19. The fusion protein of any one of claims 14-18, wherein the napDNAbp is:

(a) a CasY protein that comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 31;

(b) a Cpf1 protein or a Cpf1 mutant protein that comprises an amino acid sequence that is at least 90% identical to any one of SEQ ID NOs: 9-16;

(c) a C2c1 protein that comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 26;

(d) a C2c2 protein that comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 27;

(e) a C2c3 protein that comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 28;

(f) an Argonaute protein that comprises an amino acid sequence that is at least 90% identical to SEQ ID NO: 25; or

(g) a CasX protein that comprises (i) an amino acid sequence that is at least 90% identical to SEQ ID NO: 29 or 30; or (ii) the amino acid sequence of SEQ ID NO: 29 or 30.

20. The fusion protein of any one of claims 14-19, wherein the cytidine deaminase domain

(i) is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase;

(ii) comprises an amino acid sequence that is 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 an amino acid sequence of any one of SEQ ID NOs: 49-84;

(iii) comprises an amino acid sequence of any one of SEQ ID NOs: 49-84;

(iv) is a rat APOBEC1 (rAPOBEC1) deaminase comprising one or more mutations selected from the group consisting of W90Y, R126E, and R132E of SEQ ID NO: 76, or one or more corresponding mutations in another APOBEC deaminase;

(v) is a human APOBEC1 (hAPOBEC1) deaminase comprising one or more mutations selected from the group consisting of W90Y, Q126E, and R132E of SEQ ID NO: 74, or one or more corresponding mutations in another APOBEC deaminase;

(vi) is a human APOBEC3G (hAPOBEC3G) deaminase comprising one or more mutations selected from the group consisting of W285Y, R320E, and R326E of SEQ ID NO: 60, or one or more corresponding mutations in another APOBEC deaminase;

(vii) is an activation-induced deaminase (AID); or

(viii) is a cytidine deaminase 1 from *Petromyzon marinus* (pmCDA1).

21. The fusion protein of any one of claims 14-20, wherein the cytidine deaminase domain is an APOBEC family deaminase selected from the group consisting of APOBEC1 deaminase, APOBEC2 deaminase, APOBEC3A deaminase, APOBEC3B deaminase, APOBEC3C deaminase, APOBEC3D deaminase, APOBEC3F deaminase, APOBEC3G deaminase, and APOBEC3H deaminase.

22. The fusion protein of any one of claims 14-21 further comprising a nuclear localization sequence (NLS).

23. A complex comprising the fusion protein of any one of claims 1-22 and a guide RNA bound to the napDNAbp of the fusion protein.

24. The complex of claim 23, wherein the RNA is a guide RNA that is from 10-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

25. The complex of claim 24, wherein the target sequence is in the genome of a eukaryote, such as the genome of a mammal.

26. The complex of claim 25, wherein the mammal is a human.

27. An *in vitro* method comprising contacting a nucleic acid molecule with the complex of any one of claims 23-26, wherein the guide RNA comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the genome of an organism.

28. A polynucleotide encoding the fusion protein of any one of claims 1-22.

29. A vector comprising a polynucleotide of claim 28, wherein the vector comprises a heterologous promoter driving expression of the polynucleotide.

30. A cell comprising the fusion protein of any one of claims 1-22, the complex of any one of claims 23-26, the polynucleotide of claim 28, or the vector of claim 29.

31. The fusion protein of any one of claims 1-22, the complex of any one of claims 23-26, the polynucleotide of claim 28, the vector of claim 29, or the cell of claim 30 for use as a medicament.

32. Use of the fusion protein of any one of claims 1-22, the complex of any one of claims 23-26, the polynucleotide of claim 28, the vector of claim 29, or the cell of claim 30 for the manufacture of a medicament for the treatment of a disease, disorder, or condition, wherein the disease, disorder or condition is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, HIV, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), desmin-related myopathy (DRM), a neoplastic disease associated with a mutant PI3KCA protein, a mutant CTNNB1 protein, a mutant HRAS protein, or a mutant p53 protein.

33. A pharmaceutical composition comprising the fusion protein of any one of claims 1-22, the complex of any one of claims 23-26, the vector of claim 29, or the cell of claim 30.

34. A method for producing a ribonucleoprotein (RNP) complex, the method comprising:

(i) complexing the fusion protein of any one of claims 1-22 with an RNA in an aqueous solution, thereby forming a complex comprising the base editor and the RNA in the aqueous solution; and

(ii) contacting the complex of (i) with a cationic lipid.

35. The method of claim 34, wherein the fusion protein and the RNA of (i) are complexed at a molar ratio from 1:1 to 1:1.5.

36. The method of claim 34 or 35, wherein the complex in the aqueous solution of (i) is contacted with the cationic lipid of (ii) at a volumetric ratio that is from 1:2 to 2:1.

37. A method for purifying a fusion protein, the method comprising:

(i) expressing the fusion protein of any one of claims 1-22 in a cell, wherein the fusion protein further comprises an affinity tag;

(ii) lysing the cell of (i), thereby generating a lysate; and

(iii) subjecting the lysate to affinity chromatography, thereby producing an eluent comprising a purified fusion protein.

38. An *in vitro* method for editing a nucleobase pair of a double-stranded DNA sequence, the method comprising:

contacting a target region of the double-stranded DNA sequence with a complex comprising a nucleobase editor and a guide nucleic acid, wherein the target region comprises a target nucleobase pair;

whereby a first nucleobase of the target nucleobase pair is converted to a second nucleobase; and

whereby a third nucleobase complementary to the first nucleobase is replaced by a fourth nucleobase complementary to the second nucleobase; and

wherein the nucleobase editor comprises a UGI domain; and

wherein the nucleobase editor comprises a CasX protein, a CasY protein, a Cpf1 protein, a C2c1 protein, a C2c2 protein, a C2c3 protein, or an Argonaute protein.

39. The method of claim 38, wherein the nucleobase editor comprises the fusion protein of any one of claims 1-22.

40. The method of claim 38 or 39, wherein the nucleobase editor comprises two or more UGI domains.

41. The method of any one of claims 38-40, wherein the step of contacting causes less than 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, or less than 10% indel formation in the double-stranded DNA sequence.

42. The method of any one of claims 38-41, wherein the efficiency of generating the intended edited base pair is at least 5%, 10%, 20%, 30%, 40%, or 50%.

43. The method of any one of claims 38-42, wherein the target nucleobase pair comprises a T to 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 the disease or disorder.

44. The method of claim 43, wherein the disease or disorder is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, HIV, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), desmin-related myopathy (DRM), a neoplastic disease associated with a mutant PI3KCA protein, a mutant CTNNB1 protein, a mutant HRAS protein, or a mutant p53 protein.

45. A kit comprising a nucleic acid construct, comprising

(a) a nucleic acid sequence encoding the fusion protein of any one of claims 1-22; and

(b) a heterologous promoter that drives expression of the sequence of (a), and

(c) an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.