

COMPOSITIONS AND METHODS FOR NON-TOXIC CONDITIONING

CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims priority to U.S. App. No. 62/893,677 filed August 29,
5 2019, the entire contents of which are hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Gene therapy using autologous hematopoietic stem cell (HSCs) is desirable because it is not generally associated with the requirement to suppress the immune system and carries
10 reduced risk of host vs graft or graft vs host disease. However, ensuring engraftment of hematopoietic stem cell transplants in a host remains challenging. Conditioning (depletion of HSCs) prior to hematopoietic stem cell transplantation (HSCT) is used to promote engraftment. Indeed, conditioning efficacy is associated with improved engraftment. One disadvantage of myeloablative (toxic) conditioning is that it can lead to increasing levels of
15 chimerism.

Current methods for conditioning for autologous gene therapy for the treatment of, for example, β -thalassemia, sickle cell disease (SCD), or adenosine deaminase deficiency, typically involve the use of intravenous busulfan between 2-5 mg/kg for 2-4 days. Busulfan is a DNA alkylating drug originally designed to treat hematologic diseases, such as acute
20 myeloid leukemia (AML). However, busulfan carries the risk of significant side effects, including sterility, primary or secondary malignancy, and additional acute and chronic toxicities

Conditioning prior to HSCT is an unmet medical need. Thus, compositions and methods for conditioning to promote the engraftment of hematopoietic stem cell transplants
25 are urgently required.

SUMMARY OF THE INVENTION

As described below, the present invention features compositions and methods for non-toxic conditioning.

In one aspect, the invention provides methods of producing a hematopoietic stem cell or progenitor thereof for the treatment of a hemoglobinopathy, hematologic cancer, or myeloproliferative disease. In some embodiments, the method includes (a) expressing in the hematopoietic stem cell or progenitor thereof a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein

(napDNAbp) and a deaminase; and (b) contacting the hematopoietic stem cell or progenitor thereof with a guide RNA that targets a nucleic acid molecule encoding a cell surface protein selected from the group of CD117, CXCR4, CD135, CD90, CD45, and CD34, and introducing a mutation in the cell surface protein.

In some embodiments, the method includes (a) expressing in a hematopoietic stem cell or progenitor thereof comprising a CD117 protein a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; (b) contacting the hematopoietic stem cell or progenitor with a guide RNA capable of targeting a polynucleotide encoding the cell surface protein, thereby producing a hematopoietic stem cell or progenitor thereof for the treatment of a hemoglobinopathy, hematologic cancer, or myeloproliferative disease.

In some embodiments, the hemoglobinopathy is selected from the group of sickle cell anemia, thalassemia, Fanconi anemia, aplastic anemia, and Wiskott-Aldrich syndrome. In some embodiments, the hematologic cancer is selected from the group of acute myeloid leukemia, acute lymphoid leukemia, chronic myeloid leukemia, chronic lymphoid leukemia, multiple myeloma, diffuse large B-cell lymphoma, and non-Hodgkin's lymphoma. In some embodiments, the myeloproliferative disease is a myelodysplastic syndrome. In one aspect, the invention provides methods of producing a stem cell or progenitor thereof for the treatment of an immune deficiency. In some embodiments, the immune deficiency is severe combined immune deficiency (SCID).

In another aspect, the invention provides methods of identifying a mutation that alters antibody binding to a cell surface protein. In some embodiments, the method includes (a) expressing in a cell comprising a cell surface protein a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; (b) contacting the cell with a guide RNA capable of targeting a nucleic acid molecule encoding the cell surface protein and introducing a mutation in the cell surface protein; and (c) contacting the cell with an antibody that specifically binds a wild-type cell surface protein, but that exhibits reduced binding to the cell surface protein comprising the mutation, thereby identifying a mutation that alters antibody binding to the cell surface protein.

In some embodiments, the method further includes assaying a biological activity of the cell. In some embodiments, the cell surface protein is selected from the group of CD117, CXCR4, CD135, CD90, CD45, and CD34. In some embodiments, the cell surface protein is CD117. In some embodiments, the method further includes contacting the cell with one or

more additional guide RNAs that target a cell surface protein selected from the group of CXCR4, CD135, CD90, CD45, and CD34. In some embodiments, the cell is a hematopoietic stem cell or a progenitor thereof. In some embodiments, the mutation is a missense mutation. In some embodiments, the missense mutation fails to alter the biological activity of the cell surface protein. In some embodiments, one or more amino acid substitutions are introduced by the nucleobase editor polypeptide and the guide RNAs.

In some embodiments, the method includes (a) expressing in a hematopoietic stem cell or progenitor thereof a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; (b) contacting the cell with a guide RNA capable of targeting a nucleic acid molecule encoding a cell surface protein selected from the group of CD117, CXCR4, CD135, CD90, CD45, CD34, and introducing a mutation in the cell surface protein; and (c) contacting the cell with an antibody that specifically binds a wild-type cell surface protein, but that exhibits reduced binding to the cell surface protein comprising a mutation, thereby identifying a mutation that alters antibody binding to the cell surface protein.

In yet another aspect, the invention provides methods of base editing a gene encoding a cell surface protein expressed by a hematopoietic stem cell or a progenitor thereof. In some embodiments, the method includes (a) expressing in a hematopoietic stem cell or a progenitor thereof comprising a CD117 protein a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; and (b) contacting the cell with a guide RNA capable of targeting a nucleic acid molecule encoding the CD117 protein, thereby base editing the gene encoding the cell surface protein.

In one aspect, the invention provides methods of conditioning a subject concurrent with or subsequent to a hematopoietic stem cell transplant (HSCT). In some embodiments, the method includes (a) expressing in an isolated hematopoietic stem cell of the subject or of a donor a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; (b) contacting the hematopoietic stem cell with a guide RNA capable of targeting a nucleic acid molecule encoding a cell surface protein selected from the group of CD117, CXCR4, CD135, CD90, CD45, and CD34, thereby introducing a mutation in the cell surface protein and generating an edited hematopoietic stem cell; (c) administering the edited hematopoietic stem cell to the subject; and (d) administering to the subject an antibody, antibody drug conjugate, or chimeric antigen receptor expressing T cell (CAR-T) that selectively binds a wild-type version of the

cell surface protein, wherein the administering of step (d) is concurrent with or subsequent to step (c).

In some embodiments, the method includes (a) expressing in a hematopoietic stem cell of the subject a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; (b) contacting the hematopoietic stem cell with a guide RNA capable of targeting a nucleic acid molecule encoding a CD117 protein, thereby introducing a mutation in the CD117 protein and generating an edited hematopoietic stem cell; (c) administering the edited hematopoietic stem cell to the subject; and (d) administering to the subject an antibody, antibody drug conjugate, or chimeric antigen receptor expressing T cell (CAR-T) that selectively binds a wild-type version of CD117, wherein the administering of step (d) is concurrent with or subsequent to step (c).

In some embodiments, the deaminase domain is an adenosine deaminase or a cytidine deaminase. In some embodiments, the adenosine deaminase domain is a TadA deaminase domain. In some embodiments, the adenosine deaminase is a TadA*8 variant. In some embodiments, the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24. In some embodiments, the deaminase is a monomer or heterodimer. In some embodiments, the nucleobase editor polypeptide is an ABE8 base editor. In some embodiments, the ABE8 base editor is ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d, or ABE8.24-d. In some embodiments, the nucleobase editor polypeptide is an internal base editor (IBE) comprising the deaminase domain inserted at an internal location of the napDNAbp. In some embodiments, the nucleobase editor polypeptide further comprises one or more uracil glycosylase inhibitors (UGIs). In some embodiments, the nucleobase editor polypeptide further comprises one or more nuclear localization sequences (NLS).

In some embodiments, the mutation is a missense mutation. In some embodiments, the missense mutation fails to alter the biological activity of the cell surface protein. In some

embodiments, the CD117 cell surface protein comprising the mutation is capable of binding Stem Cell Factor (SCF). In some embodiments, the CD117 cell surface protein comprising the mutation is capable of SCF signaling. In some embodiments, the mutation is at least one amino acid substitution resulting from the modification of one or more single target nucleobases. In some embodiments, the single target nucleobase is a cytosine (C) and wherein the modification comprises conversion of the C to a thymine (T). In some embodiments, the single target nucleobase is an adenosine (A) and wherein the modification comprises conversion of the A to a guanine (G). In some embodiments, the at least one amino acid substitution is a naturally occurring mutation. In some embodiments, the at least one amino acid substitution is in domain 1, 2, 3, 4, or 5 of CD117. In some embodiments, the at least one amino acid substitution in CD117 is selected from the group of: T13A, S35P, I39V, H40R, K43G, K43R, S44P, D45G, I47V, D52G, E53G, I54V, R55G, L56P, L57P, T59A, F63P, V64A, K65E, K65R, W66R, T67A, D72G, E73G, T74A, N75D, N75G, E76G, N77G, N77S, N77Y, K78E, K78R, Q79R, N80G, N80S, E81G, E81D, W82R, I83T, I83V, T84A, E85G, K86E, E88G, T90A, N99G, K100G, H101R, K116R, V120A, S123P, L124P, Y125H, K127G, K127R, E128G, D129G, D129E, N130G, N130D, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, K149E, K149G, Q152R, K154E, K154G, R161G, F162P, F162L, I163T, I163V, D165G, M171T, I172T, I172V, K173G, S174G, K176G, Q190R, E191G, K193G, V195A, L196P, S197P, E198G, K199G, F200P, I201T, I201V, L202P, V213A, V214A, S215P, V216A, K218G, K218R, S220G, Y221C, E225G, E227G, E228G, T230A, S240G, Y243C, K247G, R248G, Q256R, E257G, E257D, K258E, K258G, Y259C, Y259H, N260G, D266G, N268D, N268G, Y269C, T276A, I277T, I277V, S279P, R281G, V282A, S285P, N293G, N294G, T295A, F296P, S298P, N300G, N300S, T302A, T303A, T304A, M318V, T322A, V323A, F324L, N326G, D327G, D332G, I334V, K342E, K342G, K342R, Q347R, Y350H, M351T, R353G, T354A, T354I, K358G, E360G, D361G, K364G, E366G, N367G, H378R, T380A, R381G, K383G, T385A, T389A, D398G, V399A, N400G, V407A, Y408H, E414G, I415V, T417A, Y418C, Y418H, D419G, R420G, L421P, V422A, N423S, M425V, E435G, I438M, D439G, V454A, L455P, V457A, V459A, Q460R, T461A, N463G, S464P, S465P, F469P, K471E, K471G, L472P, V473A, Q475R, S476G, I478M, I478V, D479G, S481G, F483P, K484G, N486G, N486S, T488A, Y494C, N495D, D496G, K499E, Y503C, Y503H, F504P, and N505G. In some embodiments, the at least one amino acid substitution in CD117 is selected from the group of: T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G,

N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, and T488A. In some embodiments, the at least one amino acid substitution in CD117 is a naturally occurring mutation selected from the group of: T13A, N77S, D129E, N130D, D131N, T144A, Y221C, E257D, T322A, T354I, D419G.

In some embodiments, the is a hematopoietic stem cell, a common myeloid progenitor, proerythroblast, or erythroblast. In some embodiments, the cell is a CD34⁺ cell. In some embodiments, the cell is from a subject having hemoglobinopathy, hematologic cancer, or a myeloproliferative disease. In some embodiments, the cell is from a subject having sickle cell disease (SCD). In some embodiments, the cell is from a subject having hereditary persistence of fetal hemoglobin (HPFH). In some embodiments, the cell is from a subject having severe combined immune deficiency (SCID). In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell.

In some embodiments, the deaminase domain is fused to the napDNAbp. In some embodiments, the deaminase domain is inserted at an internal location of the napDNAbp. In some embodiments, the napDNAbp is a nuclease inactive or nickase variant. In some embodiments, the napDNAbp comprises a Cas9, Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polynucleotide or a portion thereof. In some embodiments, the napDNAbp comprises a Cas9 polynucleotide or a portion thereof. In some embodiments, the napDNAbp comprises a dead Cas9 (dCas9) or a Cas9 nickase (nCas9). In some embodiments, the napDNAbp is a modified *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a modified *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof. In some embodiments, the napDNAbp comprises a variant of SpCas9 having an altered protospacer-adjacent motif (PAM) specificity. In some embodiments, the altered PAM has specificity for the nucleic acid sequence 5'-NGC-3'.

In some embodiments, the deaminase domain is capable of deaminating cytidine or adenine in DNA. In some embodiments, the deaminase domain is a cytidine deaminase domain. In some embodiments, the cytidine deaminase is an APOBEC deaminase domain or a derivative thereof. In some embodiments, the deaminase domain is an adenosine deaminase domain. In some embodiments, the adenosine deaminase domain is TadA

deaminase domain. In some embodiments, the adenosine deaminase is a TadA*8 variant. In some embodiments, the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24. In some embodiments, the deaminase is a monomer or a heterodimer. In some embodiments, the guide polynucleotide comprises a nucleic acid sequence comprising at least 10 contiguous nucleotides that are complementary to a CD117 nucleic acid sequence. In some embodiments, the one or more guide polynucleotides comprises a nucleic acid sequence selected from Table 23. In some embodiments, the one or more guide polynucleotides comprise a nucleic acid sequence that hybridizes to the complement of a CD117 target sequence selected from the group of:

CAAGCTATCTTCTTAGGGA; GCAAGCTATCTTCTTAGGGAA;

ACTTACGACAGGCTCGTGAA; TGACCAATTATTCCTCAAG;

AGTGACCAATTATTCCTCA; ACTACAGTATTTGTAAACGA;

AAGACAACGACACGCTGGTC; and GGCTGTTATGCACTGATCCG. In some

embodiments, the guide RNAs target the following scaffold nucleic acid sequence:

GTTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
CACCGAGTCGGTGC.

In some embodiments, the antibody is an anti-CD117 antibody. In some embodiments, the antibody binds domain 1, 2, 3, 4, or 5 of CD117. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the hemoglobinopathy is sickle cell disease (SCD). In some embodiments, the hemoglobinopathy is hereditary persistence of fetal hemoglobin (HPFH). In some embodiments, the sickle cell disease is associated with a mutation in a beta globin (HBB) polynucleotide. In some embodiments, the HPFH is associated with a mutation in a Hemoglobin Subunit Gamma 1 (HBG1) and/or Hemoglobin Subunit Gamma 2 (HBG2) polynucleotide.

In another aspect, the invention provides a base editor system including a fusion protein or a polynucleotide encoding the fusion protein, wherein the fusion protein comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase domain, and a guide polynucleotide comprising a nucleic acid sequence selected from Table 23.

In yet another aspect, the invention provides a base editor system including a fusion protein or a polynucleotide encoding the fusion protein, wherein the fusion protein comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase domain, and a guide polynucleotide comprising a nucleic acid sequence that hybridizes to the

complement of a target nucleic acid sequence selected from the group of:

CAAGCTATCTTCTTAGGGA; GCAAGCTATCTTCTTAGGGAA;
 ACTTACGACAGGCTCGTGAA; TGACCAATTATTCCCTCAAG;
 AGTGACCAATTATTCCCTCA; ACTACAGTATTTGTAAACGA;
 AAGACAACGACACGCTGGTC; and GGCTGTTATGCACTGATCCG. In some
 embodiments, the fusion protein further comprises one or more uracil glycosylase inhibitors
 (UGIs). In some embodiments, the fusion protein further comprises one or more nuclear
 localization sequences (NLS).

In some embodiments, the base editor system is capable of modifying one or more
 single target nucleobases to effect at least one amino acid substitution in a CD117
 polypeptide. In some embodiments, the single target nucleobase is a cytosine (C) and
 wherein the modification comprises conversion of the C to a thymine (T). In some
 embodiments, the single target nucleobase is an adenosine (A) and wherein the modification
 comprises conversion of the A to a guanine (G). In some embodiments, the at least one
 amino acid substitution is a naturally occurring mutation. In some embodiments, the at least
 one amino acid substitution is in domain 1, 2, 3, 4, or 5 of CD117. In some embodiments, the
 at least one amino acid substitution is selected from the group consisting of: T13A, S35P,
 I39V, H40R, K43G, K43R, S44P, D45G, I47V, D52G, E53G, I54V, R55G, L56P, L57P,
 T59A, F63P, V64A, K65E, K65R, W66R, T67A, D72G, E73G, T74A, N75D, N75G, E76G,
 N77G, N77S, N77Y, K78E, K78R, Q79R, N80G, N80S, E81G, E81D, W82R, I83T, I83V,
 T84A, E85G, K86E, E88G, T90A, N99G, K100G, H101R, K116R, V120A, S123P, L124P,
 Y125H, K127G, K127R, E128G, D129G, D129E, N130G, N130D, D131G, D131N, T132A,
 T144A, N145D, N145G, N145Y, N145S, Y146C, K149E, K149G, Q152R, K154E, K154G,
 R161G, F162P, F162L, I163T, I163V, D165G, M171T, I172T, I172V, K173G, S174G,
 K176G, Q190R, E191G, K193G, V195A, L196P, S197P, E198G, K199G, F200P, I201T,
 I201V, L202P, V213A, V214A, S215P, V216A, K218G, K218R, S220G, Y221C, E225G,
 E227G, E228G, T230A, S240G, Y243C, K247G, R248G, Q256R, E257G, E257D, K258E,
 K258G, Y259C, Y259H, N260G, D266G, N268D, N268G, Y269C, T276A, I277T, I277V,
 S279P, R281G, V282A, S285P, N293G, N294G, T295A, F296P, S298P, N300G, N300S,
 T302A, T303A, T304A, M318V, T322A, V323A, F324L, N326G, D327G, D332G, I334V,
 K342E, K342G, K342R, Q347R, Y350H, M351T, R353G, T354A, T354I, K358G, E360G,
 D361G, K364G, E366G, N367G, H378R, T380A, R381G, K383G, T385A, T389A, D398G,
 V399A, N400G, V407A, Y408H, E414G, I415V, T417A, Y418C, Y418H, D419G, R420G,
 L421P, V422A, N423S, M425V, E435G, I438M, D439G, V454A, L455P, V457A, V459A,

Q460R, T461A, N463G, S464P, S465P, F469P, K471E, K471G, L472P, V473A, Q475R, S476G, I478M, I478V, D479G, S481G, F483P, K484G, N486G, N486S, T488A, Y494C, N495D, D496G, K499E, Y503C, Y503H, F504P, and N505G. In some embodiments, the at least one amino acid substitution in CD117 is selected from the group of: T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, and T488A. In some embodiments, the at least one amino acid substitution is a naturally occurring mutation selected from the group consisting of: T13A, N77S, D129E, N130D, D131N, T144A, Y221C, E257D, T322A, T354I, D419G

In some embodiments, the napDNAbp is a nuclease inactive or nickase variant. In some embodiments, the napDNAbp comprises a Cas9, Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polynucleotide or a portion thereof. In some embodiments, the napDNAbp comprises a Cas9 polynucleotide or a portion thereof. In some embodiments, the napDNAbp comprises a dead Cas9 (dCas9) or a Cas9 nickase (nCas9). In some embodiments, the napDNAbp is a modified *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a modified *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof. In some embodiments, the napDNAbp comprises a variant of SpCas9 having an altered protospacer-adjacent motif (PAM) specificity. In some embodiments, the altered PAM has specificity for the nucleic acid sequence 5'-NGC-3'. In some embodiments, the deaminase domain is capable of deaminating cytidine or adenine in DNA. In some embodiments, the deaminase domain is a cytidine deaminase domain. In some embodiments, the cytidine deaminase is an APOBEC deaminase or a derivative thereof. In some embodiments, the deaminase domain is an adenosine deaminase domain. In some embodiments, the adenosine deaminase domain is a TadA deaminase domain. In some embodiments, the adenosine deaminase is a TadA*8 variant. In some embodiments, the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24. In some

embodiments, the deaminase is a monomer or heterodimer.

In one aspect, the invention provides a polynucleotide encoding any of the base editor systems as provided herein. In another aspect, the invention provides a cell produced by any of the methods as provided herein. In yet another aspect, the invention provides a cell produced by introducing into a cell, or a progenitor thereof, any of the base editor systems or any of the polynucleotides as provided herein. In some embodiments, the cell is produced *ex vivo* or *in vitro*. In some embodiments, the cell is a hematopoietic stem cell, a common myeloid progenitor, proerythroblast, or erythroblast. In some embodiments, the cell is a CD34⁺ cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell.

In one aspect, the invention provides an isolated cell or population of cells propagated or expanded from any of the cells as provided herein.

In yet another aspect, the invention provides pharmaceutical compositions including an effective amount of any of the cells as provided herein.

In another aspect, the invention provides methods of treating a subject with a hemoglobinopathy, hematologic cancer, or myeloproliferative disease. In some embodiments, the method includes administering to the subject any of the pharmaceutical compositions as provided herein. In some embodiments, the method includes administering to the subject a conditioning regimen comprising an antibody, antibody drug conjugate or chimeric antigen receptor expressing T-cell that selectively binds a cell surface protein and any of the cells as provided herein, thereby treating the hemoglobinopathy, hematologic cancer, or myeloproliferative disease. In some embodiments, the cell surface protein is CD117 antibody. In some embodiments, the antibody, antibody drug conjugate or chimeric antigen receptor expressing T-cell selectively binds domain 1, 2, 3, 4, or 5 of CD117. In some embodiments, the antibody is a monoclonal antibody. In some embodiments, the antibody is administered sequentially or concurrently. In some embodiments, the hemoglobinopathy is sickle cell disease or hereditary persistence of fetal hemoglobin (HPFH). In another aspect, the invention provides methods of treating a subject with an immune deficiency. In some embodiments, the immune deficiency is severe combined immune deficiency (SCID).

In some embodiments, the cell is autologous to the subject. In some embodiments, the cell is allogenic to the subject. In some embodiments, the subject is a mammal. In some embodiments, the mammal is a human.

In one aspect, the invention provides a kit comprising any of the cells as provided

herein, any of the base editor systems as provided herein, any of the polynucleotides as provided herein, or any of the pharmaceutical compositions as provided herein. In some embodiments, the kit includes written instructions for the use of the kit in the treatment of a hemoglobinopathy, hematologic cancer, myeloproliferative disease, or immune deficiency.

The description and examples herein illustrate embodiments of the present disclosure in detail. Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein. It is to be understood that this disclosure is not limited to the particular embodiments described herein and as such can vary.

5 Those of skill in the art will recognize that there are numerous variations and modifications of this disclosure, which are encompassed within its scope.

The practice of some 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
10 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*
15 *Manual of Basic Technique and Specialized Applications*, 6th Edition (R.I. Freshney, ed. (2010)).

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

Although various features of the present disclosure can be described in the context of
20 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. The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

25 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 in view of the accompanying drawings as described hereinbelow.

30

Definitions

The following definitions supplement those in the art and are directed to the current application and are not to be imputed to any related or unrelated case, *e.g.*, to any commonly owned patent or application. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present disclosure, the preferred materials and methods are described herein. Accordingly, the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

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

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.

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.

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, *e.g.*, within 5-fold, within 2-fold of a value.

- 5 Where particular values are described in the application and claims, unless otherwise stated, the term “about” means within an acceptable error range for the particular value should be assumed.

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

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
 15 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.

20 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%,
 25 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 naturally-occurring deaminase. 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 TadA deaminase is an *E. coli* TadA (ecTadA)
 30 deaminase or a fragment thereof.

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:eaa04774 (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.

A wild type TadA(wt) adenosine deaminase has the following sequence (also termed TadA reference sequence):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRQEIKAKKQKQSSTD.

In some embodiments, the adenosine deaminase comprises an alteration in the following sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

(also termed TadA*7.10).

In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 166. In particular embodiments, a variant of the above-referenced sequence comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. The alteration Y123H refers to the alteration H123Y in TadA*7.10 reverted back to Y123H TadA(wt). In other embodiments, a variant of the TadA*7.10 sequence comprises a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R.

In other embodiments, the invention provides adenosine deaminase variants that include deletions, *e.g.*, TadA*8, comprising a deletion of the C-terminus beginning at residue 149, 150, 151, 152, 153, 154, 155, 156, or 157, relative to TadA*7.10, the TadA reference

sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a TadA (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a TadA (*e.g.*, TadA*8) monomer comprising the following alterations: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In still other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains each having one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer comprising a wild-type TadA adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer comprising a wild-type TadA adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.* TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R +

Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

5 In other embodiments, the adenosine deaminase variant is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a
10 heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.* TadA*8) comprising a combination of the following alterations: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R +
15 T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; or I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In one embodiment, the adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

20 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 N-
25 terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In particular embodiments, an adenosine deaminase heterodimer comprises an
30 TadA*8 domain and an adenosine deaminase domain selected from one of the following:
Staphylococcus aureus (*S. aureus*) TadA:
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLLQOPTAH
AEHIAIERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVGADDPKGGCSGS
LMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFKNLRANKKSTN

Bacillus subtilis (*B. subtilis*) TadA:

MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML
 VIDEACKALGTWRLEGATLYVTLEPCPMCAGAVLSRVEKVVFAGFDPKGGCSGTLMN
 LLQEERFNHQAEVVSQVLEEECGGMLSAFFRELRRKKKAARKNLSE

5 *Salmonella typhimurium* (*S. typhimurium*) TadA:

MPPAFITGVTSLSDELVDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEG
 WNRPIGRHDPTAHAEIMALRQGGGLVLQNYRLDITLYVTLEPCVMCAGAMVHSRIG
 RVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIIEGVLRDECATLLSDFFRMRRQEIK
 ALKKADRAEGAGPAV

10 *Shewanella putrefaciens* (*S. putrefaciens*) TadA:

MDEYWMQVAMQMAEKAEAAAGEVPVGAVLVKDGQIATGYNLSISQHDPTAHAEI
 LCLRSAGKKLENYRLDITLYVTLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGT
 VVNLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

15

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAH
 AEIIALRNGAKNIQNYRLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYK
 TGAIGSRFHFFDDYKMNHTLEITSGVLAEESQKLSTFFQKRREEKKIEKALLKSLSDK

Caulobacter crescentus (*C. crescentus*) TadA:

20

MRTDESEDQDHRMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAH
 DPTAHAEIAAMRAAAKLGNYRLTDLTLVVTLTPCAMCAGAI SHARIGRVVFGADD
 PKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:

25

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGNLREGSN
 DPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAILARLERVVFGCYDP
 KGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFRDLRRRKKAKATPALF
 IDERKVPPEP

TadA*7.10

30

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD.

By “Adenosine Deaminase Base Editor 8 (ABE8) polypeptide” or “ABE8” is meant a base editor as defined herein comprising an adenosine deaminase variant comprising an alteration at amino acid position 82 and/or 166 of the following reference sequence: MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA

LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY P
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

In some embodiments, ABE8 comprises further alterations, as described herein, relative to the reference sequence.

5 By “Adenosine Deaminase Base Editor 8 (ABE8) polynucleotide” is meant a polynucleotide encoding an ABE8.

“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, 10 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 an oral route.

By “agent” is meant any small molecule chemical compound, antibody, nucleic acid 15 molecule, or polypeptide, or fragments thereof.

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, a 25% change, a 40% change, and a 50% or greater change in expression levels.

20 By “ameliorate” is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of a disease.

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 25 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.

As used herein, the term “antibody” refers to an immunoglobulin molecule that 30 specifically binds to, or is immunologically reactive with, a particular antigen, and includes polyclonal, monoclonal, genetically engineered, and otherwise modified forms of antibodies, including but not limited to chimeric antibodies, humanized antibodies, heteroconjugate antibodies (*e.g.*, bi- tri- and quad-specific antibodies, diabodies, triabodies, and tetrabodies), and antigen binding fragments of antibodies, including, for example, Fab', F(ab')₂, Fab, Fv,

rlgG, and scFv fragments. Unless otherwise indicated, the term "monoclonal antibody" (mAb) is meant to include both intact molecules, as well as antibody fragments (including, for example, Fab and F(ab')₂ fragments) that are capable of specifically binding to a target protein. As used herein, the Fab and F(ab')₂ fragments refer to antibody fragments that lack the Fc fragment of an intact antibody. Examples of these antibody fragments are described herein.

By "base editor (BE)," or "nucleobase editor polypeptide (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 one or more domains having base editing activity. In another embodiment, the protein domains having base editing activity are 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 domains having base editing activity are capable of deaminating a base within a nucleic acid molecule. In some embodiments, the base editor is capable of deaminating one or more bases 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 capable of deaminating a cytosine (C) and 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, the base editor is an adenosine base editor (ABE) and a cytidine base editor (CBE). In some embodiments, the base editor is a nuclease-inactive Cas9 (dCas9) fused to an adenosine deaminase. In some embodiments, the Cas9 is a circular permutant Cas9 (*e.g.*, spCas9 or saCas9). Circular permutant Cas9s are known in the art and described, for example, in Oakes *et al.*, Cell 176, 254–267, 2019. In some embodiments, the base editor is fused to an inhibitor of base excision repair, for example, a UGI domain, or a dISN domain. In some embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase and an inhibitor of base excision repair, such as a UGI or dISN domain. In other embodiments the base editor is an abasic base editor.

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 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), 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:eaa04774 (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.

In some embodiments, base editors are generated (*e.g.*, ABE8) by cloning an adenosine deaminase variant (*e.g.*, TadA*8) into a scaffold that includes a circular permutant Cas9 (*e.g.*, spCAS9) and a bipartite nuclear localization sequence. Circular permutant Cas9s are known in the art and described, for example, in Oakes *et al.*, *Cell* 176, 254–267, 2019. Exemplary circular permutant sequences are set forth below, in which the bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

CP5 (with MSP “NGC=Pam Variant with mutations Regular Cas9 likes NGG” PID=Protein Interacting Domain and “D10A” nickase):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKDWDPKKYGGFMQPTVAYSVLVVAKEK
GKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKR
M
LASAKFLQKGNEALPSKYVNFLYLASHYEKLKGSPEQNEQQLFVEQHKHYLDEIIIEQISE
FSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR

STKEVL DATLIHQSI TGLYETRIDL SQLGGDGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAI
 GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
 RRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTI
 YHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLE
 5 ENPINASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
 EDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
 IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
 DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILT
 FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
 10 LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGT YHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREM
 IEERLKYAHLEFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHK
 PENIVIAMARENQT TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQ
 15 NGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
 KYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPK
 LESEFVYGDYKVYDVRKMI AKSEQEGADKRTADGSEFESPKKKRKV*

In some embodiments, the ABE8 is selected from a base editor from **Table 10, 11 or**
 20 **13 *infra***. In some embodiments, ABE8 contains an adenosine deaminase variant evolved
 from TadA. In some embodiments, the adenosine deaminase variant of ABE8 is a TadA*8
 variant as described in **Table 8, 10, 11, or 13 *infra***. In some embodiments, the adenosine
 deaminase variant is the TadA*7.10 variant (*e.g.*, TadA*8) comprising one or more of an
 alteration selected from the group consisting of Y147T, Y147R, Q154S, Y123H, V82S,
 25 T166R, and/or Q154R. In various embodiments, ABE8 comprises TadA*7.10 variant (*e.g.*
 TadA*8) with a combination of alterations selected from the group of Y147T + Q154R;
 Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S
 + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S +
 Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R +
 30 T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y +
 V82S + Y123H + Y147R + Q154R.

In some embodiments ABE8 is a monomeric construct. In some embodiments, ABE8
 is a heterodimeric construct. In some embodiments the ABE8 base editor comprises the
 sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

By way of example, the adenine base editor ABE to be used in the base editing
5 compositions, systems and methods described herein has the 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
10 encompassed.

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTACAT
GACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTG
ACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCC
15 ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGT
CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACA
GCCGACGGAAGCGAGTTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGT
ATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGT
GCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCA
20 CACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCC
TGTATGTGCACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGT
GTTTCGGAGACGCGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATG
AACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTCTTTTA
GAATGCGGAGACAGGAGATCAAGGCCCAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGG
25 AGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCC
GGAGGATCCTCTGAGGTGGAGTTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGG
CACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTG
GAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTG
GTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTTCGAGCCTTGCGTGATGTGCGCCG
30 GCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTGTTGGCGTGAGGAACGCAAAAACCGGCGCCGAGG
CTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCA
GATGAATGTGCCGCCCTGCTGTGCTATTTCTTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGG
CCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGA
GAGCGCAACACCTGAAAGCAGCGGGGCGAGCAGCGGGGGTCAAGACAAGAAGTACAGCATCGGCCTGGCC
35 ATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGG
TGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAGCGGCGA
AACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGT
CCTTCCTGGTGGAGAGGATAAGAAGCAGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGC

CTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGAC
 CTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGCCACTTCCTGATCGAGGGCGACC
 TGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTGA
 GAAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
 5 CGGCTGGAATCTGATCGCCAGCTGCCGGCGAGAAGAAGAATGGCCTGTTTCGAAACCTGATTGCCC
 TGAGCCTGGGCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAG
 CAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGACCTGTTT
 CTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCA
 AGGCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGC
 10 TCTCGTGGGCGAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
 GGCTACATTGACGGCGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAGATGG
 ACGGCACCGAGGAACCTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAA
 CGGCAGCATCCCCACCATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGCAGGAAGATTTTTTAC
 CCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCGCGATCCCCTACTACGTGGGCC
 15 CTCTGGCCAGGGGAAACAGCAGATTGCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAA
 CTTTCGAGGAAGTGGTGGACAAGGGCGCTTCGCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
 AACCTGCCCCACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGC
 TGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGC
 CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
 20 AAAATCGAGTGCTTCGACTCCGTGGAATCTCCGGCGTGAAGATCGGTTCAACGCCTCCCTGGGCACAT
 ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA
 AGATATCGTGCTGACCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
 CACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGAGATACACCGGCTGGGGCAGGCTGAGCC
 GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
 25 CTTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA
 GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTA
 AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
 GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
 ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
 30 CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
 ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
 TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGCAAGAGCGACAACGTGCCCTCCGAAG
 AGGTGCTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTT
 CGACAATCTGACCAAGGCCGAGAGAGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
 35 CTGGTGGAAACCCGGCAGATCACAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
 ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCG
 GAAGGATTTCCAGTTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCTTACCTGAAC
 GCGCTCGTGGGAACCGCCCTGATCAAAAAGTACCTAAGCTGGAAGCGAGTTCTGTACGGCGACTACA
 AGGTGTACGAGCTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
 40 CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
 CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGGCACCGTGC

GGAAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAA
 AGAGTCTATCCTGCCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
 TACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGT
 CCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA
 5 TCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGA AAAAGGACCTGATCATCAAGCTGCCTAAG
 TACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAA
 ACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGG
 CTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATC
 GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAAGTGTGTCCGCCT
 10 ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAA
 TCTGGGAGCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCCGGAAGAGGTACACCAGCACAAA
 GAGGTGCTGGACGCCACCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
 AGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
 GAAAGTCTAACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTT
 15 CTAGTTGCCAGCCATCTGTGTGTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCAC
 TGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATGTCTGAGTAGGTGTCTATTCTATTCTGGGGGT
 GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
 CTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTA
 ATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGA
 20 AGCATAAAGTGTAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
 CCGCTTTCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
 TTTGCGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTCGGTGCGGCGA
 GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
 TGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
 25 CCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
 AGATAACAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCTGCCGCTTACCGGAT
 ACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
 GGTGTAGTCTGTTGCTCCAGCTGGGCTGTGTGCACGAACCCCCGTTAGCCCGACCGCTGCGCCTTA
 TCCGGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
 30 ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
 CACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
 TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCA
 GAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAATC
 ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
 35 AGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
 CACCTATCTCAGCGATCTGTCTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
 GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
 GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
 CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTTTGGCGAACGT
 40 TGTGTCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTCAGCTCCGGTTCC
 CAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGA

TCGTTGTCAGAAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
 TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
 ATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA
 AAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
 5 TTCGATGTAACCACTCGTGCACCAACTGATCTTCAGCATCTTTACTTTACCAGCGTTTCTGGGTGA
 GCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
 TCTTCCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
 TATTTAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTGACGGA
 TCGGGAGATCGATCTCCCGATCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
 10 GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
 AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTCGCTGCTTCGCGAT
 GTACGGGGCCAGATATACCGGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCAT
 TAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCC
 CAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCAT
 15 TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

By way of example, a 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
 20 sequences having at least 95% or greater identity to the BE4 nucleic acid sequence are also
 encompassed.

1 ATATGCCAAG TACGCCCCCT ATTGACGTCA ATGACGGTAA ATGGCCCGCC TGGCATTATG
 61 CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA CATCTACGTA TTAGTCATCG
 121 CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG GCGTGGATAG CGGTTTGACT
 25 181 CACGGGGATT TCCAAGTCTC CACCCCATTG ACGTCAATGG GAGTTTGTTC TGGCACCAAA
 241 ATCAACGGGA CTTTCCAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCAA ATGGGCGGTA
 301 GCGGTGTACG GTGGGAGGTC TATATAAGCA GAGCTGGTTT AGTGAACCGT CAGATCCGCT
 361 AGAGATCCGC GGCCGCTAAT ACGACTCACT ATAGGGAGAG CCGCCACCAT GAGCTCAGAG
 421 ACTGGCCCAG TGGCTGTGGA CCCACATTG AGACGGCGGA TCGAGCCCCA TGAGTTTGAG
 30 481 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
 35 781 CCCCAGATC GACAAGGCCT GCGGGATTG 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
 40 1081 GCCACCGGGT TGAAATCTGG TGGTCTTCTT GGTGGTCTA GCGGCAGCGA GACTCCCGGG

1141 ACCTCAGAGT CCGCCACACC CGAAAGTTCT GGTGGTTCTT CTGGTGGTTC TGATAAAAAG
 1201 TATTCTATTG GTTTAGCCAT CGGCACTAAT TCCGTTGGAT GGGCTGTCAT AACCGATGAA
 1261 TACAAAGTAC CTTCAAAGAA ATTTAAGGTG TTGGGGAACA CAGACCGTCA TTCGATTAAA
 1321 AAGAACTCTTA TCGGTGCCCT CCTATTGCAT AGTGGCGAAA CGGCAGAGGC GACTCGCCTG
 5 1381 AAACGAACCG CTCGGAGAAG GTATACACGT CGCAAGAACC GAATATGTTA CTTACAAGAA
 1441 ATTTTTAGCA ATGAGATGGC CAAAGTTGAC GATTCTTTCT TTCACCGTTT GGAAGAGTCC
 1501 TTCCTGTGCG AAGAGGACAA GAAACATGAA CGGCACCCCA TCTTTGGAAA CATAGTAGAT
 1561 GAGGTGGCAT ATCATGAAAA GTACCCAACG ATTTATCACC TCAGAAAAAA GCTAGTTGAC
 1621 TCAACTGATA AAGCGGACCT GAGGTTAATC TACTTGGCTC TTGCCCATAT GATAAAGTTC
 10 1681 CGTGGGCACT TTCTCATTGA GGGTGATCTA AATCCGGACA ACTCGGATGT CGACAAACTG
 1741 TTCATCCAGT TAGTACAAAC CTATAATCAG TTGTTTGAAG AGAACCTTAT AAATGCAAGT
 1801 GCGGTGGATG CGAAGGCTAT TCTTAGCGCC CGCCTCTCTA AATCCCGACG GCTAGAAAAC
 1861 CTGATCGCAC AATTACCCGG AGAGAAGAAA AATGGGTTGT TCGGTAACCT TATAGCGCTC
 1921 TCACTAGGCC TGACACCAAA TTTTAAGTCG AACTTCGACT TAGCTGAAGA TGCCAAATTG
 15 1981 CAGCTTAGTA AGGACACGTA CGATGACGAT CTCGACAATC TACTGGCACA AATTGGAGAT
 2041 CAGTATGCGG ACTTATTTTT 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
 20 2281 GCGGAGCGA GTCAAGAGGA ATTCTACAAG TTTATCAAAC CCATATTAGA GAAGATGGAT
 2341 GGGACGGAAG AGTTGCTTGT AAAACTCAAT CGCGAAGATC TACTGCGAAA GCAGCGGACT
 2401 TTCGACAACG GTAGCATTCC ACATCAAATC CACTTAGGCG AATTGCATGC TATACTTAGA
 2461 AGGCAGGAGG ATTTTATATCC GTTCCTCAAA GACAATCGTG AAAAGATTGA GAAAATCCTA
 2521 ACCTTTCGCA TACCTTACTA TGTGGGACCC CTGGCCCCGAG GGAACCTCTG GTTCGCATGG
 25 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
 30 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 CCTGTTTCGAC
 3121 GATAAGGTTA TGAACAGTT AAAGAGGCGT CGCTATACGG GCTGGGGACG ATTGTGCGCG
 35 3181 AAACCTATCA ACGGATAAG 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
 40 3481 CGCGAAAATC AAACGACTCA GAAGGGGCAA AAAACAGTC 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
 5 3841 CAGCTCCTAA ATGCGAAACT GATAACGCAA AGAAAGTTCG ATAACCTAAC TAAAGCTGAG
 3901 AGGGGTGGCT TGTCTGAACT TGACAAGGCC GGATTTATTA AACGTCAGCT CGTGGAAACC
 3961 CGCCAAATCA CAAAGCATGT TGCACAGATA CTAGATTCCC GAATGAATAC GAAATACGAC
 4021 GAGAACGATA AGCTGATTCT GGAAGTCAAA GTAATCACTT TAAAGTCAAA ATTGGTGTCTG
 4081 GACTTCAGAA AGGATTTTCA ATTCTATAAA GTTAGGGAGA TAAATAACTA CCACCATGCG
 10 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
 15 4441 AAAGTTTGT CCATGCCCCA AGTCAACATA GTAAAGAAAA CTGAGGTGCA GACCGGAGGG
 4501 TTTTCAAAGG AATCGATTCT TCCAAAAAGG AATAGTGATA AGCTCATCGC TCGTAAAAAG
 4561 GACTGGGACC CGAAAAAGTA CGGTGGCTTC GATAGCCCTA CAGTTGCCTA TTCTGTCTTA
 4621 GTAGTGGCAA AAGTTGAGAA GGGAAAATCC AAGAACTGA AGTCAGTCAA AGAATTATTG
 4681 GGGATAACGA TTATGGAGCG CTCGTCTTTT GAAAAGAACC CCATCGACTT CCTTGAGGCG
 20 4741 AAAGTTTACA 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 GCAAATTTTC GAATTCAGTA AGAGAGTCAT CCTAGCTGAT
 25 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 TTTGTCTACAG
 5281 CTTGGGGGTG ACTCTGGTGG TTCTGGAGGA TCTGGTGGTT CTACTAATCT GTCAGATATT
 30 5341 ATTGAAAAGG AGACCGGTAA GCAACTGGTT ATCCAGGAAT CCATCCTCAT GCTCCCAGAG
 5401 GAGGTGGAAG AAGTCATTGG GAACAAGCCG GAAAGCGATA TACTCGTGCA CACCGCCTAC
 5461 GACGAGAGCA CCGACGAGAA TGTCATGCTT CTGACTAGCG ACGCCCTGA ATACAAGCCT
 5521 TGGGCTCTGG TCATACAGGA TAGCAACGGT GAGAACAAGA TTAAGATGCT CTCTGGTGGT
 5581 TCTGGAGGAT CTGGTGGTTC TACTAATCTG TCAGATATTA TTGAAAAGGA GACCGGTAAG
 35 5641 CAACTGGTTA TCCAGGAATC CATCCTCATG CTCCCAGAGG AGGTGGAAGA AGTCATTGGG
 5701 AACAAGCCGG AAAGCGATAT ACTCGTGCAC ACCGCCTACG ACGAGAGCAC CGACGAGAAT
 5761 GTCATGCTTC TGAAGTACGA CGCCCCTGAA TACAAGCCTT GGGCTCTGGT CATAAGGAT
 5821 AGCAACGGTG AGAACAAGAT TAAGATGCTC TCTGGTGGTT CTCCAAGAA GAAGAGGAAA
 5881 GTCTAACCGG TCATCATCAC CATCACCATT GAGTTTAAAC CCGCTGATCA GCCTCGACTG
 40 5941 TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT GCCCCTCCCC CGTGCCTTCC TTGACCCTGG
 6001 AAGGTGCCAC TCCCCTGTC CTTTCCTAAT AAAATGAGGA AATTGCATCG CATTGTCTGA

6061 GTAGGTGTCA TTCTATTCTG GGGGGTGGGG TGGGGCAGGA CAGCAAGGGG GAGGATTGGG
 6121 AAGACAATAG CAGGCATGCT GGGGATGCGG TGGGCTCTAT GGCTTCTGAG GCGGAAAGAA
 6181 CCAGCTGGGG CTCGATACCG TCGACCTCTA GCTAGAGCTT GGCCTAATCA TGGTCATAGC
 6241 TGTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
 5 6301 TAAAGTGTA AGCCTAGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT
 6361 CACTGCCCCG TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC
 6421 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCTCGCTC ACTGACTCGC
 6481 TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT
 6541 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG
 10 6601 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG
 6661 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT
 6721 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCGACC CTGCCGCTTA
 6781 CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTCTCAT AGCTCACGCT
 6841 GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC
 15 6901 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG 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
 20 7201 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGG TCTGACGCTC
 7261 AGTGAACGA AAATCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA
 7321 CCTAGATCCT TTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA
 7381 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT
 7441 TTCGTTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT
 25 7501 TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTACCG GCTCCAGATT
 7561 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT
 7621 CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA
 7681 ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG
 7741 GTATGGCTTC ATTCAGCTCC GGTTCCTAAC GATCAAGGCG AGTTACATGA TCCCCATGT
 30 7801 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAAGT AAGTTGGCCG
 7861 CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG
 7921 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC
 7981 GCGGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGGCCA CATAGCAGAA
 8041 CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTGCGGGCG AAAACTCTCA AGGATCTTAC
 35 8101 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT
 8161 TTACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAGG
 8221 GAATAAGGGC GACACGAAA TGTTGAATAC TCATACTCTT CCTTTTCAA TATTATTGAA
 8281 GCATTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
 8341 AACAAATAGG GGTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC GACGGATCGG
 40 8401 GAGATCGATC TCCCGATCCC CTAGGGTCGA CTCTCAGTAC AATCTGCTCT GATGCCGCAT
 8461 AGTTAAGCCA GTATCTGCTC CCTGCTTGTG TGTTGGAGGT CGCTGAGTAG TGCGCGAGCA

8521 AAATTTAAGC TACAACAAGG CAAGGCTTGA CCGACAATTG CATGAAGAAT CTGCTTAGGG
 8581 TTAGGCGTTT TCGCTGCTT CGCGATGTAC GGGCCAGATA TACGCGTTGA CATTGATTAT
 8641 TGACTIONTAA TTAATAGTAA TCAATTACGG GGTCAATTAGT TCATAGCCCA TATATGGAGT
 8701 TCCGCGTTAC ATAACCTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC
 5 8761 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
 8821 GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATC

In some embodiments, the cytidine base editor is BE4 having a nucleic acid sequence selected from one of the following:

Original BE4 nucleic acid sequence:

10 ATGagctcagagactggcccagtggtgtggacccacattgagacggcggatcgagccccatgagtt
 tgaggtattcttcgatccgagagagctccgcaaggagacctgctgctttacgaaattaattgggggg
 gccggcactccatttggcgacatacatcacagaactaacaagcacgtcgaagtcaacttcacgag
 aagttcacgacagaaagatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggag
 ccgcaatgtagtagggccatcactgaattcctgtcaaggtatccccacgtcactctgtttatttaca
 15 tcgcaaggctgtaccaccacgctgacccccgcaatcgacaaggcctgcgggatttgatctcttcaggt
 gtgactatccaaattatgactgagcaggagtcaggatactgctggagaaactttgtgaattatagccc
 gagtaatgaagccactggcctaggtatccccatctgtgggtacgactgtacgttcttgaaactgtact
 gcatcatactgggcctgcctccttgtctcaacattctgagaaggaagcagccacagctgacattcttt
 accatcgctcttcagttctgtcattaccagcgactgccccacacattctctgggccaccgggttgaa
 20 atctggtggttcttctggtggttctagcggcagcgagactcccggaacctcagagtcgccacacccg
 aaagtctggtggttcttctggtggttctgataaaaagtattctattggtttagccatcggcactaat
 tccgttggtatgggctgtcataaccgatgaatacaaaagtaccttcaaagaaatttaagggtgttggggaa
 cacagaccgtcattcgattaaaaagaatcttatcggtgcctcctattcgatagtggcgaaacggcag
 aggcgactcgcctgaaacgaaccgctcggagaaggtatacacgtcgcaagaaccgaatatgttactta
 25 caagaaatttttagcaatgagatggccaaagttgacgattctttctttcaccgtttggaagagtcctt
 ccttgtcgaagaggacaagaaacatgaacggcaccccatctttggaacatagtagatgaggtggcat
 atcatgaaaagtacccaacgatttatcacctcagaaaaaagctagttgactcaactgataaagcggac
 ctgaggttaattctacttggctcttgcccatatgataaagttccgtgggcactttctcattgaggggtga
 tctaaatccggacaactcggatgtcgacaaactgttcatccagttagtagtaaaacctataatcagttgt
 30 ttgaagagaacctataaatgcaagtggcgtggatgcgaaggctattcttagcgcccgctctctaaa
 tcccgacggctagaaaacctgatcgcaaatcccgagagagaagaaaaatgggttggttcggtaacct
 tatagcgctctcactaggcctgacaccaaattttaagtcgaacttcgacttagctgaagatgccaaat
 tgcagcttagtaaggacacgtacgatgacgatctcgacaatctactggcacaaattggagatcagtat
 gcggacttatttttggctgcaaaaaaccttagcgatgcaatcctcctatctgacatactgagagttaa
 35 tactgagattaccaaggcgccgttatccgcttcaatgatcaaaaggtagcatgaacatcccaagact
 tgacacttctcaaggccctagtcgctcagcaactgcctgagaaatataaggaaatattctttgatcag
 tcgaaaaacgggtacgcaggttatattgacggcggagcgagtcgaagggaattctacaagtttatcaa

accatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcggaagatctactgc
gaaagcagcggacttttcgacaacggttagcattccacatcaaaccacttaggcgaattgcatgctata
cttagaaggcaggaggtttttatccgttcctcaaagacaatcgtgaaaagattgagaaaatcctaac
ctttcgcatacctactatgtgggacccctggcccagggaactctcggttcgcatggatgacaagaa
5 agtccgaagaaacgattactccatggaatgttgaggaagttgtcgataaagggtgcgtcagctcaatcg
ttcatcgagaggatgaccaactttgacaagaatttaccgaacgaaaaagtattgcctaagcacagttt
actttacgagtattttcacagtgtacaatgaactcacgaaagttaagtatgtcactgagggcatgcgta
aaccgccttttctaagcggagaacagaagaagcaatagtagatctgttattcaagaccaaccgcaa
gtgacagttaagcaattgaaagaggactactttaagaaaattgaatgcttcgattctgtcgagatctc
10 cggggtagaagatcgatttaatgcgtcacttggtacgtatcatgacctcctaaagataattaaagata
aggacttcttgataacgaagagaatgaagatatcttagaagatatagtggtgactcttaccctcttt
gaagatcgggaaatgattgaggaaagactaaaaacatacgtcacctgttcgacgataagggttatgaa
acagttaaagaggcgtcgctatacgggctggggacgattgtcgcgaaacttatcaacgggataagag
acaagcaaagtggtaaaactattctcgatttttctaagagcgacggcttcgccaataggaactttatg
15 cagctgatccatgatgactctttaaccttcaaagaggatatacaaaaggcacaggtttccggacaagg
ggactcattgcacgaacatattgcaatcttgctgggttcgccagccatcaaaaaggcactactccaga
cagtcaaagtagtgatgagctagttaaggatcatgggacgtcacaaaccggaaaacattgtaatcgag
atggcacgcgaaaatcaaacgactcagaaggggcaaaaaacagtcgagagcggatgaagagaataga
agagggtatttaaagaactgggcagccagatcttaaaggagcatcctgtggaaaataccaattgcaga
20 acgagaaactttacctctattacctacaaaatggaaggacatgtatgttgatcaggaactggacata
aaccgtttatctgattacgacgtcgatcacattgtacccaatcctttttgaaggacgattcaatcga
caataaagtgttacacgctcggataagaaccgagggaaaagtgacaatgttccaagcaggaagtgcg
taaagaaaatgaagaactattggcggcagctcctaaatgcgaaactgataacgcaaagaaagttcgat
aacttaactaaagctgagaggggtggcttgctgaacttgacaaggccgatttattaaacgtcagct
25 cgtggaaaccgcgcaaatcacaaagcatgttgacagatactagattcccgaatgaatacgaatacgcg
acgagaacgataagctgattcgggaagtcaaagtaatcactttaaaagtcaaaattggtgtcggaacttc
agaaaggattttcaattctataaagttagggagataaataactaccaccatgcgcacgacgcttatct
taatgccgtcgtagggaccgcactattaagaaatacccgaagctagaaaagtgagtttgtgtatggtg
attacaaagtttatgacgtccgtaagatgatcgcgaaaagcgaacaggagataggcaaggctacagcc
30 aaatacttcttttattctaacattatgaatttctttaagacggaaatcactctggcaaaccgagagat
acgcaaaccgacctttaattgaaaccaatggggagacaggtgaaatcgtatgggataagggccgggact
tcgcgacggtgagaaaagttttgtccatgccccagtcaacatagtaaagaaaactgaggtgcagacc
ggagggttttcaaaggaatcgattcttccaaaaaggaatagtataagctcatcgctcgtaaaaagga
ctgggacccgaaaaagtacggtggcttcgatagccctacagttgcctattctgtcctagtagtgga
35 aagttgagaagggaaaaatccaagaaactgaagtcagtcgaagaattattggggataacgattatggag
cgctcgtcttttgaagaaaccccatcgacttcttgaggcgaaaggttacaaggaagtaaaaaagga
tctcataattaaactaccaaagtatagtctgtttgagttagaaaatggccgaaaacggatgttggtta

gcgccggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaatttcctgtattta
 cggtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaactttttgttgagca
 gcacaaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcaccttagctg
 atgccaatctggacaaagtattaagcgcatacaacaagcacagggataaaccatacgtgagcaggcg
 5 gaaaatattatccatttgtttactcttaccacctcggcgctccagccgattcaagtattttgacac
 aacgatagatcgcacaaacgatacacttctaccaaggaggtgctagacgcgacactgattcaccaatcca
 tcacgggattatatgaaactcggatagatttgtcacagcttgggggtgactctggtggttctggagga
 tctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggttatccagga
 atccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcg
 10 tgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgccccctgaatac
 aagccttgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttc
 tggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggtta
 tccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgat
 atactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccc
 15 tgaatacaagccttgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctg
 gtggttctAAAAAGGACGGCGGACGGATCAGAGTTCGAGAGTCCGAAAAAAAAACGAAAGGTGCAAtaa

BE4 Codon Optimization 1 nucleic acid sequence:

ATGTCATCCGAAACCGGGCCAGTGGCCGTAGACCCAACACTCAGGAGGCGGATAGAACCCCATGAGTT
 TGAAGTGTCTTCGACCCAGAGAGCTGCGCAAAGAGACTTGCCTCCTGTATGAAATAAATTGGGGGG
 20 GTCGCCATTCAATTTGGAGGCACACTAGCCAGAATACTAACAAACACGTGGAGGTAAATTTTATCGAG
 AAGTTTACCACCGAAAGATACTTTTGCCCCAATACACGGTGTTCATTACCTGGTTTCTGTCATGGAG
 TCCATGTGGAGAATGTAGTAGAGCGATAACTGAGTTCCTGTCTCGATATCCTCACGTCACGTTGTTTA
 TATACATCGCTCGGCTTTATCACCATGCGGACCCGCGGAACAGGCAAGGTCTTCGGGACCTCATATCC
 TCTGGGGTGACCATCCAGATAATGACGGAGCAAGAGAGCGGATACTGCTGGCGAAAACCTTTGTAACTA
 25 CAGCCCAAGCAATGAGGCACACTGGCCTAGATATCCGCATCTCTGGGTTCGACTGTATGTCCTTGAAC
 TGTACTGCATAATTCTGGGACTTCCGCCATGCTTGAACATTCTGCGGCGGAAACAACCACAGCTGACC
 TTTTTCACGATTGCTCTCCAAAGTTGTCACTACCAGCGATTGCCACCCACATCTTGTGGGCTACTGG
 ACTCAAGTCTGGAGGAAGTTCAGGCGGAAGCAGCGGTCTGAAACGCCCGGAACCTCAGAGAGCGCAA
 CGCCCGAAAGCTCTGGAGGGTCAAGTGGTGGTAGTGATAAGAAATACTCCATCGGCCTCGCCATCGGT
 30 ACGAATTCTGTCGGTTGGGCCGTTATCACCGATGAGTACAAGGTCCCTTCTAAGAAATTCAAGGTTTT
 GGGCAATACAGACCGCCATTCTATAAAAAAAAAACCTGATCGGCGCCCTTTTGTGTTGACAGTGGTGAGA
 CTGCTGAAGCGACTCGCCTGAAGCGAACTGCCAGGAGGCGGTATACGAGGCGAAAAACGAATTTGT
 TACCTCCAGGAGATTTTCTCAAATGAAATGGCCAAGGTAGATGATAGTTTTTTTACCGCTTGGAAGA
 AAGTTTTCTCGTTGAGGAGGACAAAAAGCACGAGAGGCACCCAATCTTTGGCAACATAGTCGATGAGG
 35 TCGCATACCATGAGAAATATCCTACGATCTATCATCTCCGCAAGAAGCTGGTCGATAGCACGGATAAA
 GCTGACCTCCGGCTGATCTACCTTGCTCTTGCTCACATGATTAAATTCAGGGGCCATTTTCTGATAGA
 AGGAGACCTCAATCCCGACAATTCTGATGTGACAAACTGTTTATTTCAGCTCGTTCAGACCTATAATC

AACTCTTTGAGGAGAACCCCATCAATGCTTCAGGGGTGGACGCAAAGGCCATTTTGTCCGCGCGCTTG
 AGTAAATCACGACGCCTCGAGAATTTGATAGCTCAACTGCCGGGTGAGAAGAAAAACGGGTGTTTTGG
 GAATCTCATAGCGTTGAGTTTGGGACTTACGCCAAACTTTAAGTCTAACTTTGATTTGGCCGAAGATG
 CCAAATTGCAGCTGTCCAAAGATACCTATGATGACGACTTGGATAACCTTCTTGCGCAGATTGGTGAC
 5 CAATACGCGGATCTGTTTCTTGCCGCAAAAAATCTGTCCGACGCCATACTCTTGTCGATATACTGCG
 CGTCAATACTGAGATAACTAAGGCTCCCTCAGCGCGTCCATGATTAAGATACGATGAGCACCACC
 AAGATCTCACTCTGTTGAAAGCCCTGGTTCGCCAGCAGCTTCCAGAGAAGTATAAGGAGATATTTTTTC
 GACCAATCTAAAAACGGCTATGCGGGTTACATTGACGGTGGCGCCTCTCAAGAAGAATTCTACAAGTT
 TATAAAGCCGATACTTGAGAAAATGGACGGTACAGAGGAATTGTTGGTTAAGCTCAATCGCGAGGACT
 10 TGTTGAGAAAAGCAGCGCACATTTGACAATGGTAGTATTCCACACCAGATTCATCTGGGCGAGTTGCAT
 GCCATTCTTAGAAGACAAGAAGATTTTTATCCGTTTCTGAAAGATAACAGAGAAAAGATTGAAAAGAT
 ACTTACCTTTTCGCATACCGTATTATGTAGGTCCCTGGCTAGAGGGAACAGTCGCTTCGCTTGGATGA
 CTCGAAAATCAGAAGAAACAATAACCCCTGGAATTTTGAAGAAGTGGTAGATAAAGGTGCGAGTGCC
 CAATCTTTTATTGAGCGGATGACAAATTTTGACAAGAATCTGCCTAACGAAAAGGTGCTTCCCAAGCA
 15 TTCCCTTTTGTATGAATACTTTACAGTATATAATGAACTGACTAAAGTGAAGTACGTTACCGAGGGGA
 TGCAGAAAGCCAGCTTTTCTCAGTGGCGAGCAGAAAAAGCAATAGTTGACCTGCTGTTCAAGACGAAT
 AGGAAGGTTACCGTCAAACAGCTCAAAGAAGATTACTTTAAAAAGATCGAATGTTTTGATTGAGTTGA
 GATAAGCGGAGTAGAGGATAGATTTAACGCAAGTCTTGGAACCTTATCATGACCTTTTGAAGATCATCA
 AGGATAAAGATTTTTTGGACAACGAGGAGAATGAAGATATCCTGGAAGATATAGTACTTACCTTGACG
 20 CTTTTTGAAGATCGAGAGATGATCGAGGAGCGACTTAAGACGTACGCACATCTCTTTGACGATAAGGT
 TATGAAACAATTGAAACGCCGGCGGTATACTGGCTGGGGCAGGCTTTCTCGAAAGCTGATTAATGGTA
 TCCGCGATAAGCAGTCTGGAAAGACAATCCTTGACTTTCTGAAAAGTGATGGATTTGCAAATAGAAAC
 TTTATGCAGCTTATACATGATGACTCTTTGACGTTCAAGGAAGACATCCAGAAGGCACAGGTATCCGG
 CCAAGGGGATAGCCTCCATGAACACATAGCCAACCTGGCCGGCTCACCAGCTATTAAGGGAATAT
 25 TGCAAACCGTTAAGGTTGTTGACGAACTCGTTAAGGTTATGGGCCGACACAAACCAGAGAATATCGTG
 ATTGAGATGGCTAGGGAGAATCAGACCACTCAAAAAGGTCAGAAAAATTTCTCGCGAAAGGATGAAGCG
 AATTGAAGAGGGAATCAAAGAAGCTTGGCTCTCAAATTTTGAAGAGCACCCGGTAGAAAACACTCAGC
 TGCAGAATGAAAAGCTGTATCTGTATTATCTGCAGAATGGTCGAGATATGTACGTTGATCAGGAGCTG
 GATATCAATAGGCTCAGTGACTACGATGTGACCACATCGTTTCTCAATCTTTCCTGAAAGATGACTC
 30 TATCGACAACAAAGTGTTGACGCGATCAGATAAGAACCGGGGAAAATCCGACAATGTACCCTCAGAAG
 AAGTTGTCAAGAAGATGAAAAACTATTGGAGACAATTGCTGAACGCCAAGCTCATAACACAACGCAAG
 TTCGATAACTTGACGAAAGCCGAAAGAGGTGGGTGTGTCAGAATTGGACAAAGCTGGCTTTATTAAGCG
 CCAATTGGTGGAGACCCGGCAGATTACGAAACACGTAGCACAAATTTTGGATTACGAATGAATACCA
 AATACGACGAAAACGACAAATTGATACGCGAGGTGAAAGTGATTACGCTTAAGAGTAAGTTGGTTTTCC
 35 GATTTGAGGAAGGATTTTTCAGTTTTACAAAGTAAGAGAAATAAACAACTACCACCACGCCCATGATGC
 TTACCTCAACGCGGTAGTTGGCACAGCTCTTATCAAAAAATATCCAAAGCTGGAAAGCGAGTTTCGTTT
 ACGGTGACTATAAAGTATACGACGTTTCGGAAGATGATAGCCAAATCAGAGCAGGAAATTGGGAAGGCA

ACCGCAAAATACTTCTTCTATTCAAACATCATGAACTTCTTTAAGACGGAGATTACGCTCGCGAACGG
 CGAAATACGCAAGAGGCCCTCATAGAGACTAACGGCGAAACCGGGGAGATCGTATGGGACAAAGGAC
 GGGACTTTGCGACCGTTAGAAAAGTACTTTCAATGCCACAAGTGAATATTGTTAAAAAGACAGAAGTA
 CAAACAGGGGGGTTCAGTAAGGAATCCATTTTGCCCAAGCGGAACAGTGATAAATTGATAGCAAGGAA
 5 AAAAGATTGGGACCCTAAGAAGTACGGTGGTTTCGACTCTCCTACCGTTGCATATTCAGTCCTTG TAG
 TTGCGAAAGTGAAAAAGGGGAAAAGTAAGAAGCTTAAGAGTGTTAAAGAGCTTCTGGGCATAACCATA
 ATGGAACGGTCTAGCTTCGAGAAAAATCCAATTGACTTTCTCGAGGCTAAAGGTTACAAGGAGGTAAA
 AAAGGACCTGATAATTAACTCCCAAAGTACAGTCTCTTCGAGTTGGAGAATGGGAGGAAGAGAATGT
 TGGCATCTGCAGGGGAGCTCCAAAAGGGGAACGAGCTGGCTCTGCCTTCAAAATACGTGAACTTTCTG
 10 TACCTGGCCAGCCACTACGAGAACTCAAGGGTTCTCCTGAGGATAACGAGCAGAAACAGCTGTTTGT
 AGAGCAGCACAAGCATTACCTGGACGAGATAATTGAGCAAATTAGTGAGTTCTCAAAAAGAGTAATCC
 TTGCAGACGCGAATCTGGATAAAGTTCTTTCCGCCTATAATAAGCACCGGGACAAGCCTATACGAGAA
 CAAGCCGAGAACATCATTACCTCTTTACCCTTACTAATCTGGGCGCGCCGGCCCTTCAAATACTT
 CGACACCACGATAGACAGGAAAAGGTATACGAGTACCAAAGAAGTACTTGACGCCACTCTCATCCACC
 15 AGTCTATAACAGGGTTGTACGAAACGAGGATAGATTTGTCCCAGCTCGGCGGCGACTCAGGAGGGTCA
 GCGGCTCCGGTGGATCAACGAATCTTTCCGACATAATCGAGAAAGAAACCGGCAAACAGTTGGTGAT
 CCAAGAATCAATCCTGATGCTGCCTGAAGAAGTAGAAGAGGTGATTGGCAACAAACCTGAGTCTGACA
 TTCTTGCCACACCGGTATGACGAGAGCACGGACGAGAACGTTATGCTTCTCACTAGCGACGCCCT
 GAGTATAAACCATGGGCGCTGGTCATCCAAGATTCCAATGGGGAAAACAAGATTAAGATGCTTAGTGG
 20 TGGGTCTGGAGGGAGCGGTGGGTCCACGAACCTCAGCGACATTATTGAAAAAGAGACTGGTAAACAAC
 TTGTAATACAAGAGTCTATTCTGATGTTGCCTGAAGAGGTGGAGGAGGTGATTGGGAACAAACCGGAG
 TCTGATATACTTGTTTCATACCGCCTATGACGAATCTACTGATGAGAATGTGATGCTTTTaaACGTGAGA
 CGCTCCCGAGTACAAACCTGGGCTCTGGTGATTACAGGACAGCAATGGTGAGAAATAAGATTAAAATGT
 TGAGTGGGGGCTCAAAGCGCACGGCTGACGGTAGCGAATTTGAGAGCCCCAAAAAAAACGAAAGGTC
 25 GAAtaa

BE4 Codon Optimization 2 nucleic acid sequence:

ATGAGCAGCGAGACAGGCCCTGTGGCTGTGGATCCTACACTGCGGAGAAGAATCGAGCCCCACGAGTT
 CGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCG
 GCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACCTTTATCGAG
 30 AAGTTTACGACCGAGCGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTC
 CCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCA
 TCTATATCGCCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGACTGCGCGACCTGATCAGC
 AGCGGAGTGACCATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACCTTCGTGAACTA
 CAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAAC
 35 TGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACC
 TTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGG
 ACTTAAGAGCGGAGGATCTAGCGGCGGCTCTAGCGGATCTGAGACACCTGGCACAAGCGAGTCTGCCA

CACCTGAGAGTAGCGGCGGATCTTCTGGCGGCTCCGACAAGAAGTACTCTATCGGACTGGCCATCGGC
ACCAACTCTGTTGGATGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCT
GGGCAACACCGACCGGCACAGCATCAAGAAGAATCTGATCGGCGCCCTGCTGTTGACTCTGGCGAAA
CAGCCGAAGCCACCAGACTGAAGAGAACC GCCAGGCGGAGATACACCCGGCGGAAGAACCGGATCTGC
5 TACCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGA
GTCCTTCTGTTGGAAGAGGACAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGATGAGG
TGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAG
GCCGACCTGAGACTGATCTACCTGGCTCTGGCCACATGATCAAGTTCCGGGGCCACTTTCTGATCGA
GGCGATCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACC
10 AGCTGTTTCGAGGAAAAACCCCATCAACGCCTCTGGCGTGGACGCCAAGGCTATCCTGTCTGCCAGACTG
AGCAAGAGCAGAAGGCTGGAACCTGATCGCCAGCTGCCTGGCGAGAAGAAGAATGGCCTGTTTCG
CAACCTGATTGCCCTGAGCCTGGGACTGACCCCTAACTTCAAGAGCAACTTCGACCTGGCCGAGGATG
CCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAATCTGCTGGCCAGATCGGCGAT
CAGTACGCCGACTTGTCTGCGCCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGATATCCTGAG
15 AGTGAACACCGAGATCACAAAGGCCCTCTGAGCGCTCTATGATCAAGAGATACGACGAGCACCACC
AGGATCTGACCCTGCTGAAGGCCCTCGTTAGACAGCAGCTGCCAGAGAAGTACAAAGAGATTTTCTTC
GATCAGTCCAAGAACGGCTACGCCGGCTACATTGATGGCGGAGCCAGCCAAGAGGAATTCTACAAGTT
CATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAAGTGTGGTCAAGCTGAACAGAGAGGACC
TGCTGCGGAAGCAGCGGACCTTCGACAATGGCTCTATCCCTCACCAGATCCACCTGGGAGAGCTGCAC
20 GCCATTCTGCGGAGACAAGAGGACTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGAT
CCTGACCTTCAGGATCCCTACTACGTGGGACCACTGGCCAGAGGCAATAGCAGATTTCGCTGGATGA
CCAGAAAAGAGCGAGGAAACCATCACACCCTGGAAGTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCT
CAGTCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCTAACGAGAAGGTGCTGCCAAGCA
CTCCCTGCTGTATGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAATACGTGACCGAGGGAA
25 TGAGAAAAGCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCCATTGTGGATCTGCTGTTCAAGACCAAC
CGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGTTCGACAGCGTGGA
AATCAGCGGCGTGGAAGATCGGTTCAATGCCAGCCTGGGCACATACCACGACCTGCTGAAAATTATCA
AGGACAAGGACTTCCTGGACAACGAAGAGAACGAGGACATTCTCGAGGACATCGTGCTGACCCTGACA
CTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACATACGCCACCTGTTTCGACGACAAAGT
30 GATGAAGCAACTGAAGCGGAGGCGGTACACAGGCTGGGGCAGACTGTCTCGGAAGCTGATCAACGGCA
TCCGGGATAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAAAC
TTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCAGGTGTCCGG
CCAAGGCGATTCTCTGCACGAGCACATTGCCAACCTGGCCGGATCTCCCGCCATTAAGAAGGGCATCC
TGCAGACAGTGAAGGTGGTGGACGAGCTTGTGAAAGTGATGGGCAGACACAAGCCCCGAGAACATCGTG
35 ATCGAAATGGCCAGAGAGAACCAGACCACAGAAAGGGCCAGAAGAACAGCCGCGAGAGAATGAAGCG
GATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGC
TGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGACGGGATATGTACGTGGACCAAGAGCTG

GACATCAACCGGCTGAGCGACTACGATGTGGACCATATCGTGCCCCAGAGCTTTCTGAAGGACGACTC
 CATCGATAACAAGGTCTTGACCAGAAGCGACAAGAACCGGGCAAGAGCGATAACGTGCCCTCCGAAG
 AGGTGGTCAAGAAGATGAAGAATACTGGCGACAGCTGCTGAACGCCAAGCTGATTACCCAGCGGAAG
 TTCGATAACCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTTGATAAGGCCGGCTTCATTAAGCG
 5 GCAGCTGGTGGAAACCCGGCAGATCACCAACACGTGGCACAGATTCTGGACTCCCGGATGAACACTA
 AGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTCATCACCTGAAGTCTAAGCTGGTGTCC
 GATTTCCGGAAGGATTTCCAGTTCTACAAAGTGCGGGAAATCAACAATAACCATCACGCCCACGACGC
 CTACCTGAATGCCGTTGTTGGAACAGCCCTGATCAAGAAGTATCCCAAGCTGGAAAGCGAGTTCTGTGT
 ACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAACAAGAGATCGGCAAGGCT
 10 ACCGCCAAGTACTTTTTCTACAGCAACATCATGAACTTTTTCAAGACAGAGATCACCTTGGCCAACGG
 CGAGATCCGGAAGAACCCCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCA
 GAGATTTTGCACAGTGCAGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAGAAAACCGAGGTG
 CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCTAAGCGGAACAGCGATAAGCTGATCGCCAGAAA
 GAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGATAGCCCTACCGTGGCCTATTCTGTGTCTGGTGG
 15 TGGCCAAAGTGGAAGGGCAAGTCCAAAAGCTCAAGAGCGTGAAAGAGCTGCTGGGGATCACCATC
 ATGGAAGAAGCAGCTTTGAGAAGAACCCGATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTCAA
 GAAGGACCTCATCATCAAGCTCCCCAAGTACAGCCTGTTTCGAGCTGGAAAATGGCCGAAGCGGATGC
 TGGCCTCAGCAGGCGAACTGCAGAAAGGCAATGAAGTGGCCCTGCCTAGCAAATACGTCAACTTCCTG
 TACCTGGCCAGCCACTATGAGAAGCTGAAGGGCAGCCCCGAGGACAATGAGCAAAAGCAGCTGTTTGT
 20 GGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCC
 TGGCCGACGCTAACCTGGATAAGGTGCTGTCTGCCTATAACAAGCACCGGGACAAGCCTATCAGAGAG
 CAGGCCGAGAAATATCATCCACCTGTTTACCCTGACCAACCTGGGAGCCCCCTGCCGCTTCAAGTACTT
 CGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACACTGATCCACC
 AGTCTATCACCGGCTGTACGAAACCCGGATCGACCTGTCTCAGCTCGGCGGCGATTCTGGTGGTTCT
 25 GGCGGAAGTGGCGGATCCACCAATCTGAGCGACATCATCGAAAAAGAGACAGGCAAGCAGCTCGTGAT
 CCAAGAAATCCATCCTGATGCTGCCTGAAGAGGTTGAGGAAGTGATCGGCAACAAGCCTGAGTCCGACA
 TCCTGGTGACACCGCTACGATGAGAGCACCGATGAGAACGTCATGCTGCTGACAAGCGACGCCCT
 GAGTACAAGCCTTGGGCTCTCGTGATTGAGGACAGCAATGGGGAGAACAAGATCAAGATGCTGAGCGG
 AGGTAGCGGAGGAGTGGCGGAAGCACAAACCTGTCTGATATCATTGAAAAAGAAACCGGGAAGCAAC
 30 TGGTCATTCAAGAGTCCATTCTCATGCTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAACCCGAG
 AGCGATATTCTGGTCCACACAGCCTATGACGAGTCTACAGACGAAAACGTGATGCTCCTGACCTCTGA
 CGCTCCCGAGTATAAGCCCTGGGCACTTGTTATCCAGGACTCTAACGGGGAAAAACAAAATCAAAATGT
 TGTCGGCGGCAGCAAGCGGACAGCCGATGGATCTGAGTTTCGAGAGCCCCAAGAAGAAACGGAAGGTg
 GAGtaa

- 35 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. In another embodiment, the base editing activity is cytidine deaminase activity, *e.g.*, converting target C•G to T•A and adenosine or adenine deaminase activity, *e.g.*, converting A•T to G•C.

5 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, a deaminase domain (*e.g.*, cytidine deaminase or adenosine deaminase) for deaminating nucleobases in the target nucleotide sequence; and (2) one or more guide polynucleotides (*e.g.*, guide RNA) in conjunction with
10 the polynucleotide programmable nucleotide binding domain. In various embodiments, the base editor (BE) system comprises a nucleobase editor domain selected from an adenosine deaminase or a cytidine deaminase, and a domain having nucleic acid sequence specific binding activity. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable DNA binding domain and a deaminase
15 domain for deaminating one or more nucleobases in a target nucleotide sequence; and (2) one or more guide RNAs 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
20 adenosine base editor (ABE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE) or a cytidine base editor (CBE).

 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
25 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease. An exemplary Cas9, is *Streptococcus pyogenes* Cas9 (spCas9), the amino acid sequence of which is provided below:
MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAET
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVD
30 EVAYHEKYPTIYHLRKKLADSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK

DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
 LTTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 5 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGHSLSHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIDKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 10 GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSD
 KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSSFENPI
 DFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIR
 15 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQL
 GGD (single underline: HNH domain; double underline: RuvC domain)

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
 20 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.
 25 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 –NH₂ can be maintained.

30 The term “coding sequence” or “protein coding sequence” as used interchangeably
 herein refers to a segment of a polynucleotide that codes for a protein. Coding sequences can
 also be referred to as open reading frames. The region or sequence is bounded nearer the 5’
 end by a start codon and nearer the 3’ end with a stop codon. Stop codons useful with the
 base editors described herein include the following:

Glutamine **CAG → TAG** Stop codon

CAA → TAA

Arginine **CGA → TGA**

Tryptophan **TGG → TGA**

5 **TGG → TAG**

TGG → TAA

As used herein, the terms "condition" and "conditioning" refer to processes by which a patient is prepared for receipt of a transplant containing hematopoietic stem cells. Such procedures promote the engraftment of a hematopoietic stem cell transplant (for instance, as
 10 inferred from a sustained increase in the quantity of viable hematopoietic stem cells within a blood sample isolated from a patient following a conditioning procedure and subsequent hematopoietic stem cell transplantation). According to the methods described herein, a patient may be conditioned for hematopoietic stem cell transplant therapy by administration to the patient of an antibody or antigen-binding fragment thereof capable of binding an
 15 antigen expressed by hematopoietic stem cells, such as CD117, CXCR4, CD135, CD90, CD45, and/or CD34. Such antibodies are expected to act via complement-mediated cytotoxicity and antibody-dependent cell-mediated cytotoxicity. As described herein, the transplanted cells have been edited so that the antibody no longer binds the antigen (e.g., CD117, CXCR4, CD135, CD90, CD45, and/or CD34). Administration of an antibody,
 20 antigen-binding fragment thereof, drug-antibody conjugate, or chimeric antigen receptor expressing T-cell (CAR-T) capable of binding one or more antigens (e.g., CD117, CXCR4, CD135, CD90, CD45, CD34) to a patient in need of hematopoietic stem cell transplant therapy can promote the engraftment of a hematopoietic stem cell graft, for example, by selectively depleting endogenous hematopoietic stem cells, thereby creating a vacancy filled
 25 by an exogenous hematopoietic stem cell transplant.

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*
 30 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.

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 deaminase (*e.g.*, engineered adenosine deaminase, evolved adenosine deaminase) 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.

“Detect” refers to identifying the presence, absence or amount of the analyte to be detected. In one embodiment, a sequence alteration in a polynucleotide or polypeptide is detected. In another embodiment, the presence of indels is detected.

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 enzyme linked immunosorbent assay (ELISA)), biotin, digoxigenin, or haptens.

By “disease” is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. Exemplary diseases include diseases amenable to treatment with hematopoietic stem cell transplantation, such as β -thalassemia, sickle cell disease (SCD), or adenosine deaminase deficiency.

5 By “effective amount” is meant the amount of an agent or active compound, *e.g.*, a base editor as described herein, that is required to ameliorate the symptoms of a disease relative to an untreated patient or an individual without disease, *i.e.*, a healthy individual, or is the amount of the agent or active compound sufficient to elicit a desired biological response. The effective amount of active compound(s) used to practice the present invention for
10 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 in a cell (*e.g.*,
15 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. Such therapeutic effect need not be sufficient to alter a pathogenic gene in all cells of a subject, tissue or organ, but only to alter the pathogenic gene 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
20 one or more symptoms of a disease.

In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*, of a nucleobase editor comprising a nCas9 domain and a deaminase domain (*e.g.*, adenosine deaminase, cytidine deaminase) refers to the amount that is sufficient to induce editing of a target site specifically bound and edited by the nucleobase editors described herein. As will
25 be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, may vary depending on various factors as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and/or on the agent being used.

In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*,
30 of a fusion protein comprising a nCas9 domain and a deaminase domain may refer to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the fusion protein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, a nuclease, a hybrid protein, a protein

dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and/or on the agent being used.

- 5 By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, at least about 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.
- 10 By "guide RNA" or "gRNA" is meant a polynucleotide which is 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 (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
- 15 RNAs (sgRNAs), although "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
- 20 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
- 25 US20160208288, entitled "Switchable Cas9 Nucleases and Uses Thereof," and US 9,737,604, 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." An extended gRNA will 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
- 30 sequence that complements a target site, which mediates binding of the nuclease/RNA complex to the target site, providing the sequence specificity of the nuclease:RNA complex.

As used herein, the term "hematopoietic stem cells" ("HSCs") refers to immature blood cells having the capacity to self-renew and to differentiate into mature blood cells containing diverse lineages including but not limited to granulocytes (*e.g.*, promyelocytes,

neutrophils, eosinophils, basophils), erythrocytes (*e.g.*, reticulocytes, erythrocytes), thrombocytes (*e.g.*, megakaryoblasts, platelet producing megakaryocytes, platelets), monocytes (*e.g.*, monocytes, macrophages), dendritic cells, microglia, osteoclasts, and lymphocytes (*e.g.*, NK cells, B-cells and T-cells). Such cells may include CD34⁺ cells.

5 CD34⁺ cells are immature cells that express the CD34 cell surface marker. In humans, CD34⁺ cells are believed to include a subpopulation of cells with the stem cell properties defined above, whereas in mice, HSCs are CD34⁻. In addition, HSCs also refer to long term repopulating HSCs (LT-HSC) and short term repopulating HSCs (ST-HSC). LT-HSCs and ST-HSCs are differentiated, based on functional potential and on cell surface marker

10 expression. For example, human HSCs are CD34⁺, CD38⁻, CD45RA⁻, CD90⁺, CD49F⁺, and lin⁻(negative for mature lineage markers including CD2, CD3, CD4, CD7, CD8, CD10, CD11B, CD19, CD20, CD56, CD235A). In mice, bone marrow LT-HSCs are CD34⁻, SCA-1⁺, C-kit⁺, CD135⁻, Slamf1/CD150⁺, CD48⁻, and lin⁻ (negative for mature lineage markers including Ter119, CD11b, Gr1, CD3, CD4, CD8, B220, IL7ra), whereas ST-HSCs are

15 CD34⁺, SCA-1⁺, C-kit⁺, CD135⁻, Slamf1/CD150⁺, and lin⁻(negative for mature lineage markers including Ter119, CD11b, Gr1, CD3, CD4, CD8, B220, IL7ra). In addition, ST-HSCs are less quiescent and more proliferative than LT-HSCs under homeostatic conditions. However, LT-HSC have greater self renewal potential (*i.e.*, they survive throughout adulthood, and can be serially transplanted through successive recipients), whereas ST-HSCs

20 have limited self renewal (*i.e.*, they survive for only a limited period of time, and do not possess serial transplantation potential). Any of these HSCs can be used in the methods described herein. ST-HSCs are particularly useful because they are highly proliferative and thus, can more quickly give rise to differentiated progeny.

As used herein, the term "hematopoietic stem cell functional potential" refers to the

25 functional properties of hematopoietic stem cells which include 1) multi-potency (which refers to the ability to differentiate into multiple different blood lineages including, but not limited to, granulocytes (*e.g.*, promyelocytes, neutrophils, eosinophils, basophils), erythrocytes (*e.g.*, reticulocytes, erythrocytes), thrombocytes (*e.g.*, megakaryoblasts, platelet producing megakaryocytes, platelets), monocytes (*e.g.*, monocytes, macrophages), dendritic

30 cells, microglia, osteoclasts, and lymphocytes (*e.g.*, NK cells, B-cells and T-cells), 2) self-renewal (which refers to the ability of hematopoietic stem cells to give rise to daughter cells that have equivalent potential as the mother cell, and further that this ability can repeatedly occur throughout the lifetime of an individual without exhaustion), and 3) the ability of hematopoietic stem cells or progeny thereof to be reintroduced into a transplant recipient

whereupon they home to the hematopoietic stem cell niche and re-establish productive and sustained hematopoiesis."

By "heterodimer" is meant a fusion protein comprising two domains, such as a wild type TadA domain and a variant of TadA domain (*e.g.*, TadA*8) or two variant TadA domains (*e.g.*, TadA*7.10 and TadA*8 or two TadA*8 domains).

"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.

By "increases" is meant a positive alteration of at least 10%, 25%, 50%, 75%, or 100%.

The terms "inhibitor of base repair", "base repair inhibitor", "IBR" or their grammatical equivalents refer 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, hOGG1, hNEIL1, T7 EndoI, T4PDG, UDG, hSMUG1, and hAAG. In some embodiments, the base repair inhibitor is an inhibitor of Endo V or 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 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.

5 An "intein" is a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing." In some embodiments, an intein of a precursor protein (an intein
10 containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein (*e.g.*, split intein-N and split intein-C). For example, in cyanobacteria, DnaE, the catalytic subunit α of DNA polymerase III, is encoded by two separate genes, *dnaE-n* and *dnaE-c*. The intein encoded by the *dnaE-n* gene may be herein referred as "intein-N." The intein encoded by the *dnaE-c* gene may be herein referred
15 as "intein-C."

Other intein systems may also be used. For example, a synthetic intein based on the *dnaE* intein, the Cfa-N (*e.g.*, split intein-N) and Cfa-C (*e.g.*, split intein-C) intein pair, has been described (*e.g.*, in Stevens *et al.*, J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5, incorporated herein by reference). Non-limiting examples of intein pairs that may be used in
20 accordance with the present disclosure include: Cfa DnaE intein, Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (*e.g.*, as described in U.S. Patent No. 8,394,604, incorporated herein by reference). Exemplary nucleotide and amino acid sequences of inteins are provided.

DnaE Intein-N DNA:

25 TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGAT
TGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTGCGATAACAATGGTAACATTTATACTC
AGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCTGAATACTGTCTGGAGGAT
GGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCC
TATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

DnaE Intein-N Protein:

30 CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR
GEQEVFEYCLEDGSLIRATKDHKFMFTVDGQMLPIDEIFERELDLMRVDNL PN

DnaE Intein-C DNA:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGA
TATTGGAGTCGAAAGAGATCACAACCTTGCTCTGAAGAACGGATTCATAG CTTCTAAT

Intein-C: MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

5 **Cfa-N DNA:**

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGAT
TGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACAC
AGCCCATTTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGAT
GGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCC
10 AATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

Cfa-N Protein:

CLSYDTEILTVEYGFPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED
GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

Cfa-C DNA:

15 ATGAAGAGGACTGCCGATGGATCAGAGTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGAT
AATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACA
ACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

Cfa-C Protein:

MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLKNGLVASN

20 Intein-N and intein-C may be fused to the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split Cas9, i.e., to form a structure of N--[N-terminal portion of the split Cas9]-[intein-N]-C. In
25 some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, i.e., to form a structure of N-[intein-C]--[C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (*e.g.*, split Cas9) is known in the art, *e.g.*, as described in Shah *et al.*, Chem Sci. 2014; 5(1):446-461, incorporated herein by reference. Methods for designing and using
30 inteins are known in the art and described, for example by WO2014004336, WO2017132580, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

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.

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.

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.

By “CD117 (C-kit; SCFR) polypeptide” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_000213 that binds an anti-CD117 antibody. In some embodiments, an CD117 polypeptide or fragment thereof has SCF signaling activity. An exemplary CD117 polypeptide sequence follows:

5 >NP_000213.1 mast/stem cell growth factor receptor Kit isoform 1 precursor [Homo sapiens]
MRGARGAWDFLCVLLLLLRVQTGSSQPSVSPGEPSPPSIHPGKSDLIVRVGDEIRLLCTDPGFVKWTFEIL
LDETENENKQNEWITEKAEATNTGKYTCTNKHGLSNSIYVFVRDPAKLFLVDRSLYKEDNDTLVRCPLTD
PEVTNYSLKGCQKPLPKDLRFIPDPKAGIMIKSVKRAYHRLCLHCSVDQEGKSVLSEKFILKVRPAFKA
10 VPVVSVKASYLLREGEFTVTCTIKDVSSSVYSTWKRENSQTKLQEKYNSWHHGFNYERQATLTISSA
RVNDSGVFMCYANNTFGSANVTTLLEVVDKGFNIFPMINTTVFVNDGENVDLIVEYEAFFPKPEHQQWIY
MNRTFTDKWEDYPKSENESENIRYVSELHLTRLKGTEGGTYTFLVNSDVNAAIAFNIVYVNTKPEILTYDR
LVNGMLQCVAAGFPEPTIDWYFCPGTEQRCSASVLPVDVQTLNSSGPPFGKLVVQSSIDSSAFKHNGTVE
CKAYNDVGKTSAYFNFAFKGNKEQIHPHTLFTPLLI GFVIVAGMMCIIVMILTYKYLQKPMYEVQWKVV
15 EEINGNNYVYIDPTQLPYDHKWEFPRNRLSFGKTLGAGAFGKVVEATAYGLIKSDAAMTVAVKMLKPSAH
LTEREALMSELKVL SYLGNHMNIVNLLGACTIGGPTLVITEYCCYGDLLNFLRRKRDSFICSKQEDHAEA
ALYKNLLHSKESSCSDSTNEYMDMKPGVS YVVP TKADKRRSVRIGSYIERDVT PAIMEDDELALDLEDLL
SFSYQVAKGMAFLASKNCIHRDLAARNILLTHGRITKICDFGLARDIKNDSNYVVKGNARLPVKWMAPE
IFNCVYTFESDVWSYGI FLWELFSLGSSPYPGMPVDSKFYKMIKEGFRMLSPEHAPAEMYDIMKTCWDAD
20 PLKRPTFKQIVQLIEKQISESTNHIYSNLANCSPNRQKPVVDH SVRINSVGSTASSQPLL VHDDV

By “CD117 polynucleotide” is meant a nucleic acid molecule that encodes a CD117 polypeptide. An exemplary CD117 polynucleotide sequence follows:

>NM_000222.2 Homo sapiens KIT proto-oncogene, receptor tyrosine kinase (KIT), transcript variant 1, mRNA
25 TCTGGGGGCTCGGCTTTGCCGCGCTCGCTGCACTTGGGCGAGAGCTGGAACGTGGACCAGAGCTCGGATC
CCATCGCAGCTACCGCGATGAGAGGCGCTCGCGGCGCCTGGGATTTTCTCTGCGTTCTGCTCCTACTGCT
TCGCGTCCAGACAGGCTCTTCTCAACCATCTGTGAGTCCAGGGAACCGTCTCCACCATCCATCCATCCA
GGAAAATCAGACTTAATAGTCCGCGTGGGCGACGAGATTAGGCTGTTATGCACTGATCCGGGCTTTGTCA
AATGGACTTTTGAGATCCTGGATGAAACGAATGAGAATAAGCAGAATGAATGGATCACGGAAAAGGCAGA
30 AGCCACCAACACCGGCAAATACACGTGCACCAACAAACACGGCTTAAGCAATCCATTTATGTGTTTGT
AGAGATCCTGCCAAGCTTTTCCTTGTGACCGCTCCTGTATGGGAAAGAAGACAACGACACGCTGGTCC
GCTGTCTCTCACAGACCCAGAAGTGACCAATTATTCCTCAAGGGGTGCCAGGGGAAGCCTCTCCCAA
GGACTTGAGGTTTATTCTGACCCCAAGCGGGCATCATGATCAAAAGTGTAACGCGCCTACCATCGG
CTCTGTCTGCATTGTTCTGTGGACCAGGAGGGCAAGTCAGTGCTGTGCGAAAAATTATCCTGAAAGTGA
35 GGCCAGCCTTCAAAGCTGTGCTGTGTGTCTGTGTCCAAAGCAAGCTATCTTCTTAGGGAAGGGGAAGA
ATTCACAGTGACGTGCACAATAAAGATGTGTCTAGTTCTGTGTACTCAACGTGGAAAAGAGAAAACAGT
CAGACTAACTACAGGAGAAATATAATAGCTGGCATCACGGTGACTTCAATTATGAACGTCAGGCAACGT
TGACTATCAGTTAGCGAGAGTTAATGATTCTGGAGTGTTTCATGTGTATGCCAATAATACTTTTGATC
AGCAAATGTCACAACAACCTTGAAGTAGTAGATAAAGATTATTAATATCTTCCCATGATAAACACT

ACAGTATTTGTAAACGATGGAGAAAATGTAGATTTGATTGTTGAATATGAAGCATTCCCCAACCTGAAC
ACCAGCAGTGGATCTATATGAACAGAACCTTCACTGATAAATGGGAAGATTATCCCAAGTCTGAGAATGA
AAGTAATATCAGATACGTAAGTGAACCTCATCTAACGAGATTAAAAGGCACCGAAGGAGGCACTTACACA
TTCCTAGTGTCCAATTCTGACGTCAATGCTGCCATAGCATTTAATGTTTATGTGAATACAAAACCGA
5 TCCTGACTTACGACAGGCTCGTGAATGGCATGCTCCAATGTGTGGCAGCAGGATTCCCAGAGCCCA
AGATTGGTATTTTTGTCCAGGAAGTGAAGCAGAGATGCTCTGCTTCTGTACTGCCAGTGGATGTGCAGACA
CTAAACTCATCTGGGCCACCCTTTGGAAAGCTAGTGGTTCAGAGTCTATAGATTCTAGTGCATTCAAGC
ACAATGGCAGGTTGAATGTAAGGCTTACAACGATGTGGGCAAGACTTCTGCCTATTTTAACCTTGCATT
TAAAGGTAACAACAAAGAGCAAATCCATCCCCACACCCTGTTCACTCCTTTGCTGATTGGTTTCGTAATC
10 GTAGCTGGCATGATGTGCATTATTGTGATGATTCTGACCTACAAATATTTACAGAAACCCATGTATGAAG
TACAGTGGAAAGTTGTTGAGGAGATAAATGGAAACAATTATGTTTACATAGACCCAAACAACCTTCCTTA
TGATCACAAATGGGAGTTTCCAGAAACAGGCTGAGTTTGGGAAAACCTGGGTGCTGGAGCTTTCGGG
AAGGTTGTTGAGGCAACTGCTTATGGCTTAATTAAGTCAGATGCGGCCATGACTGTCGCTGTAAAGATGC
TCAAGCCGAGTGCCCATTTGACAGAACGGGAAGCCCTCATGTCTGAACTCAAAGTCTGAGTTACCTTGG
15 TAATCACATGAATATTGTGAATCTACTTGGAGCCTGCACCATTGGAGGGCCACCCCTGGTCAATTACAGAA
TATTGTTGCTATGGTGATCTTTGAATTTTTGAGAAGAAAACGTGATTCATTTATTTGTTCAAAGCAGG
AAGATCATGCAGAAGCTGCACTTTATAAGAATCTTCTGCATTCAAAGGAGTCTTCTGTCAGCGATAGTAC
TAATGAGTACATGGACATGAAACCTGGAGTTTCTTATGTTGTCCCAACCAAGGCCGACAAAAGGAGATCT
GTGAGAATAGGCTCATACATAGAAAGAGATGTGACTCCCGCCATCATGGAGGATGACGAGTTGGCCCTAG
20 ACTTAGAAGACTTGCTGAGCTTTTCTTACCAGGTGGCAAAGGGCATGGCTTTCCTCGCCTCCAAGAATTG
TATTCACAGAGACTTGGCAGCCAGAAATATCCTCCTTACTCATGGTCGGATCACAAAGATTTGTGATTTT
GGTCTAGCCAGAGACATCAAGAATGATTCTAATTATGTGGTTAAAGGAAACGCTCGACTACCTGTGAAGT
GGATGGCACCTGAAAGCATTTTCAACTGTGTATACACGTTTGAAAGTGACGTCTGGTCTATGGGATTTT
TCTTTGGGAGCTGTTCTCTTTAGGAAGCAGCCCTATCCTGGAATGCCGGTCGATTCTAAGTTCTACAAG
25 ATGATCAAGGAAGGCTTCCGGATGCTCAGCCCTGAACACGCACCTGCTGAAATGTATGACATAATGAAGA
CTTGCTGGGATGCAGATCCCCTAAAAGACCAACATTCAAGCAAATTTGTTCAAGCTAATTGAGAAGCAGAT
TTCAGAGAGCACCAATCATATTTACTCCAACCTTAGCAAACCTGCAGCCCCAACCGACAGAAGCCCGTGGTA
GACCATTCTGTGCGGATCAATTCTGTGCGCAGCACCGCTTCTCCTCCCAGCCTCTGCTTGTGCACGACG
ATGTCTGAGCAGAATCAGTGTGTTGGGTACCCCTCCAGGAATGATCTCTTCTTTTGGCTTCCATGATGGT
30 TATTTTCTTTCTTTCAACTTGCATCCAACCTCCAGGATAGTGGGCACCCCACTGCAATCCTGTCTTTCTG
AGCACACTTTAGTGGCCGATGATTTTTTGTATCAGCCACCATCCTATTGCAAAGGTTCCAACGTATATA
TTCCCCAATAGCAACGTAGCTTCTACCATGAACAGAAAACATTCTGATTTGGAAAAAGAGAGGGGAGGTATG
GACTGGGGGCCAGAGTCTTTCCAAGGCTTCTCCAATTCTGCCCAAAAATATGGTTGATAGTTTACCTGA
ATAAATGGTAGTAATCACAGTTGGCCTTCAGAACCATCCATAGTAGTATGATGATACAAGATTAGAAGCT
35 GAAAACCTAAGTCTTTTATGTGGAACAGAACATCATTAGAACAAAGGACAGAGTATGAACACCTGGGC
TTAAGAAATCTAGTATTTTCTGCTGGGAATGAGACATAGGCCATGAAAAAATGATCCCCAAGTGTGAAC
AAAAGATGCTCTTCTGTGGACCACTGCATGAGCTTTTATACTACCGACCTGGTTTTAAATAGAGTTTGC
TATTAGAGCATTGAATTGGAGAGAAGGCCTCCCTAGCCAGCACTTGTATATACGCATCTATAAATTGTCC
GTGTTTATACATTTGAGGGGAAAACACCATAAGGTTTCTGTTTCTGTATACAACCTGGCATTATGTCCAC
40 TGTGTATAGAAGTAGATTAAAGAGCCATATAAGTTTGAAGGAAACAGTTAATACCATTTTTTAAGGAAACA
ATATAACCACAAAGCACAGTTTGAACAAAATCTCCTCTTTTAGCTGATGAACTTATTCTGTAGATTCTGT

GGAACAAGCCTATCAGCTTCAGAATGGCATTGTACTCAATGGATTTGATGCTGTTTGACAAAGTTACTGA
 TTCCTGTCATGGCTCCACAGGAGTGGGAAAACACTGCCATCTTAGTTTGGATTCTTATGTAGCAGGAAA
 TAAAGTATAGGTTTAGCCTCCTTCGCAGGCATGTCCTGGACACCGGGCCAGTATCTATATATGTGTATGT
 ACGTTTGTATGTGTGTAGACAAATATTTGGAGGGGTATTTTGGCCCTGAGTCCAAGAGGGTCCTTTAGTA
 5 CCTGAAAAGTAACTTGGCTTTCATTATAGTACTGCTCTTGTTCCTTTTACATAGCTGTCTAGAGTAGC
 TTACCAGAAGCTTCCATAGTGGTGCAGAGGAAGTGAAGGCATCAGTCCCTATGTATTTGCAGTTCACCT
 GCACTTAAGGCACTCTGTTATTTAGACTCATCTTACTGTACCTGTTCCCTTAGACCTTCCATAATGCTACT
 GTCTCACTGAAACATTTAAATTTTACCCTTTAGACTGTAGCCTGGATATTATTCTGTAGTTTACCTCTT
 TAAAAACAAAACAAAACAAAACAAAACACTCCCTTCCTCACTGCCCAATATAAAAGGCAAATGTGTACA
 10 TGGCAGAGTTTGTGTGTTGCTTGAAAGATTGAGGTATGTTGCCTTTATGGTTTCCCCCTTCTACATTTT
 TTAGACTACATTTAGAGAACTGTGGCCGTTATCTGGAAGTAACCATTTGCACTGGAGTTCTATGCTCTCG
 CACCTTTCCAAAGTTAACAGATTTTGGGGTTGTGTTGTACCCAAGAGATTGTTGTTGCCATACTTTGT
 CTGAAAAATTCTTTGTGTTTCTATTGACTTCAATGATAGTAAGAAAAGTGGTTGTTAGTTATAGATGTC
 TAGGTACTTCAGGGGCACTTCATTGAGAGTTTTGTCTTGATATTCTTGAAAGTTTATATTTTATAATT
 15 TTTTCTTACATCAGATGTTTCTTTGCAGTGGCTTAATGTTTGAAATTATTTTGTGGCTTTTTTTGTAAAT
 ATTGAAATGTAGCAATAATGTCTTTTGAATATTCCCAAGCCCATGAGTCCTTGAAATATTTTATATA
 TACAGTAACCTTATGTGTAAATACATAAGCGGCGTAAGTTTAAAGGATGTTGGTGTCCACGTGTTTTAT
 TCCTGTATGTTGTCCAATTGTTGACAGTTCTGAAGAATTCTAATAAAATGTACATATATAAATCAAAAAA
 AAAAAAAAAA

20 By “C-X-C chemokine receptor type 4 (CXCR4) polypeptide” is meant a polypeptide
 or fragment thereof having at least about 95% amino acid sequence identity to an amino acid
 sequence provided at GenBank Accession NP_001008540 that binds an anti- CXCR4
 antibody. An exemplary CXCR4 polypeptide sequence follows:

>NP_001008540.1 C-X-C chemokine receptor type 4 isoform a [Homo sapiens]
 25 MSIPLPQLLIYTSNYTEEMSGSDYDSMKPECFREENANFNKIFLPTIYSIIFLTGIVGNGLVILVMGYQ
 KKLRSMTDKYRLHLSVADLLEFVITLPFWAVDAVANWYFGNFLCKAVHVIYTVNLYSSVLILAFISLDRL
 AIVHATNSQRPRKLLAEKVYVGVWIPALLLTIPDFIFANVSEADDRYICDRFYPNDLWVVVFQFQHIMV
 GLILPGVILSCYCIISKLHSHKQKRAKLTIVLILAFFACWLPYYIGISIDSFILLEIKQGEF
 ENTVHKWISITEALAFFHCCLNPILYAFLGAKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTES
 30 SSFHSS

By “CXCR4 polynucleotide” is meant a nucleic acid molecule that encodes a CXCR4
 polypeptide. An exemplary CXCR4 polynucleotide sequence follows:

>NM_003467.2 Homo sapiens C-X-C motif chemokine receptor 4 (CXCR4), transcript variant 2,
 mRNA
 35 AACTTCAGTTTGTGGCTGCGGCAGCAGGTAGCAAAGTGACGCCGAGGGCCTGAGTGCTCCAGTAGCCAC
 CGCATCTGGAGAACCAGCGGTTACCATGGAGGGGATCAGTATATACACTTCAGATAACTACACCGAGGAA
 ATGGGCTCAGGGGACTATGACTCCATGAAGGAACCTGTTTCCGTGAAGAAAATGCTAATTTCAATAAAA
 TCTTCCTGCCACCATCTACTCCATCATCTTCTTAAGTGGCATTGTGGCAATGGATTGGTCATCCTGGT
 CATGGGTTACCAGAAGAACTGAGAAGCATGACGGACAAGTACAGGCTGCACCTGTCAAGTGGCCGACCTC
 40 CTCTTTGTATCATCAGCTTCCCTTCTGGGAGTTGATGCCGTGGCAACTGGTACTTTGGGAACCTCCTAT

- GCAAGGCAGTCCATGTCATCTACACAGTCAACCTCTACAGCAGTGTCCATCCTGGCCTTCATCAGTCT
 GGACCGCTACCTGGCCATCGTCCACGCCACCAACAGTCAGAGGCCAAGGAAGCTGTTGGCTGAAAAGGTG
 GTCTATGTTGGCGTCTGGATCCCTGCCCTCCTGCTGACTATTCCCGACTTCATCTTTGCCAACGTCAGTG
 AGGCAGATGACAGATATATCTGTGACCGCTTCTACCCCAATGACTTGTGGGTGGTTGTGTTCCAGTTTCA
 5 GCACATCATGGTTGGCCTTATCCTGCCTGGTATTGTCATCCTGTCCTGCTATTGCATTATCATCTCCAAG
 CTGTCACACTCCAAGGGCCACCAGAAGCGCAAGGCCCTCAAGACCACAGTCATCCTCATCCTGGCCTTCT
 TCGCCTGTTGGCTGCCTTACTACATTGGGATCAGCATCGACTCCTTCATCCTCCTGGAAATCATCAAGCA
 AGGGTGTGAGTTTGAGAACACTGTGCACAAGTGGATTTCATCACCAGAGGCCCTAGCTTTCTTCCACTGT
 TGTCTGAACCCCATCCTCTATGCTTTCCTTGGAGCCAAATTTAAAACCTCTGCCAGCACGCACTCACCT
 10 CTGTGAGCAGAGGGTCCAGCCTCAAGATCCTCTCCAAAGGAAAGCGAGGTGGACATTCATCTGTTTCCAC
 TGAGTCTGAGTCTTCAAGTTTTCACTCCAGCTAACACAGATGTAAAAGACTTTTTTTTATACGATAAATA
 ACTTTTTTTTTAAGTTACACATTTTTTTCAGATATAAAAGACTGACCAATATTGTACAGTTTTTATTGCTTGT
 TGGATTTTTGTCTTGTGTTTCTTTAGTTTTTGTGAAGTTAATTGACTTATTTATATAAATTTTTTTTGT
 TTCATATTGATGTGTGTCTAGGCAGGACCTGTGGCCAAGTTCCTAGTTGCTGTATGTCTCGTGGTAGGAC
 15 TGTAGAAAAGGGAAGTGAACATTCAGAGCGTGTAGTGAATCACGTAAAGCTAGAAATGATCCCCAGCTG
 TTTATGCATAGATAATCTCTCATTCCCGTGAACGTTTTTCTGTTCTTAAGACGTGATTTTGTCTGTAG
 AAGATGGCACTTATAACCAAAGCCCAAAGTGGTATAGAAATGCTGGTTTTTTCAGTTTTTCAGGAGTGGGT
 GATTTTCAGCACCTACAGTGTACAGTCTTGTATTAAGTTGTTAATAAAAGTACATGTAAACTTAAAAAA
 AAAAAAAAAA
- 20 By “CD135 polypeptide” is meant a polypeptide or fragment thereof having at least
 about 95% amino acid sequence identity to an amino acid sequence provided at GenBank
 Accession No. NP_004110 that binds an anti-CD135 antibody. An exemplary CD135
 polypeptide sequence follows:
- >NP_004110.2 receptor-type tyrosine-protein kinase FLT3 precursor [Homo sapiens]
- 25 MPALARDGGQLPLLVVFSAMIFGTITNQDLPVIKCVLINHKNNDSVKGSSSYPMVSESPEDLGCALRPQ
 SSGTVYEEAAVEVDVSASITLQVLVDAPGNISCLWVFKHSSLNCQPHFDLQNRGVVSMVILKMTETQAGE
 YLLFIQSEATNYTILFTVSIRNTLLYLRRPYFRKMENQDALVCISESVPEPIVEWVLCDSQGESCKEES
 PAVVKKEEKVLHELFGTDIRCCARNELGRECTRLFTIDLNQTPQTTLPLQLFLKVGEPWIRCKAVHVNHG
 FGLTWELNKALEEGNYFEMSTYSTNRMTIRILFAFVSSVARNDTGYTCSSSKHPSQSALVTIVEKGF
 30 NATNSSEDYEIDQYEEFCFSVRFKAYPQIRCTWTFSTRKSPCEQKGLDNGYSISKFCNHHKQPEYIFHA
 ENDDAQFTKMTLNIIRKPVLAEASASQASCFSDGYPLPSWTWKKCDKSPNCTEEITEGVNRRKANRK
 VFGQWVSSSTLNMSEAIGFLVKCCAYNSLGTSCETILLNSPGPFPIQDNISFYATIGVCLLFIVVLTLL
 LICHKYYKKQFRYESQLQMVQVTGSSDNEYFYVDFREYEDLKWEFPRENLEFGKVLGSGAFGKVMNATAY
 GISKTGVSIQVAVKMLKEKADSSEREALMSELKMMTQLGSHENIVNLLGACTLSGPIYLI FEYCCYGDLL
 35 NYLRSKREKFHRTWTEIFKEHNFSFYPTFQSHPNSSMPGSRVQIHPDSDQISGLHGNSFHSEDEIEYEN
 QKRLEEEEDLNVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNVLVTHGKVVKICDFGLARDIMSDS
 NYVVRGNARLPVKWMAPESLFEGIYTIKSDVWSYGILLWEIFSLGVNPYPGIPVDANFYKLIQNGFKMDQ
 PFYATEEIIYIMQSCWAFDSRKRPSFPNLTSFLGCQLADAEAMYQNVDRVSECPHTYQNRPPFSREMD
 LGLLSPQAQVEDS

By “CD135 polynucleotide” is meant a nucleic acid molecule that encodes a CD135 polypeptide. An exemplary CD135 polynucleotide sequence follows:

>NM_004119.2 Homo sapiens fms related tyrosine kinase 3 (FLT3), transcript variant 1, mRNA

ACCTGCAGCGCGAGGCGCGCCGCTCCAGGCGGCATCGCAGGGCTGGGCGGCGCGGCCTGGGGACCCCCGG

5 GCTCCGGAGGCCATGCCGCGCTTGGCGCGCGACGGCGGCCAGCTGCCGCTGCTCGTTGTTTTTCTGCAA

TGATATTTGGGACTATTACAAATCAAGATCTGCCTGTGATCAAGTGTGTTTTAATCAATCATAAGAACAA

TGATTCATCAGTGGGGAAGTCATCATCATATCCCATGGTATCAGAATCCCCGGAAGACCTCGGGTGTGCG

TTGAGACCCAGAGCTCAGGGACAGTGTACGAAGTGGCGCTGTGGAAGTGGATGTATCTGCTTCCATCA

CACTGCAAGTGTGGTTCGACGCCCCAGGGAACATTTCTGTCTCTGGGTCTTTAAGCACAGCTCCCTGAA

10 TTGCCAGCCACATTTTGATTTACAAACAGAGGAGTTGTTCCATGGTCATTTTGAAAATGACAGAAACC

CAAGCTGGAGAATACCTACTTTTTATTTCAGAGTGAAGCTACCAATTACACAATATTGTTTACAGTGAGTA

TAAGAAATACCTGCTTTACACATTAAGAAGACCTTACTTTAGAAAAATGGAAAACAGGACGCCCTGGT

CTGCATATCTGAGAGCGTTCAGAGCCGATCGTGGAATGGGTGCTTTGCGATTACAGGGGGAAAGCTGT

AAAGAAGAAAGTCCAGCTGTTGTTAAAAAGGAGGAAAAAGTGTTCATGAATTATTTGGGACGGACATAA

15 GGTGCTGTGCCAGAAATGAACTGGGCAGGGAATGCACCAGGCTGTTTACAATAGATCTAAATCAAACCTCC

TCAGACCACATTGCCACAATTATTTCTTAAAGTAGGGGAACCTTATGGATAAGGTGCAAAGCTGTTTCAT

GTGAACCATGGATTTCGGGCTCACCTGGGAATTAGAAAACAAAGCACTCGAGGAGGGCAACTACTTTGAGA

TGAGTACCTATTCAACAAACAGAACTATGATACGGATTCTGTTTGCTTTGTATCATCAGTGGCAAGAAA

CGACACCGGATACTACACTTGTTCCTCTTCAAAGCATCCCAGTCAATCAGCTTTGGTTACCATCGTAGAA

20 AAGGGATTTATAAATGCTACCAATTCAGTGAAGATTATGAAATTGACCAATATGAAGAGTTTTGTTTTT

CTGTCAGGTTTAAAGCCTACCCACAAATCAGATGTACGTGGACCTTCTCTCGAAAATCATTTCTTGTGA

GCAAAAGGGTCTTGATAACGGATACAGCATATCCAAGTTTTGCAATCATAAGCACCAGCCAGGAGAATAT

ATATTCCATGCAGAAAATGATGATGCCAATTTACCAAAATGTTTACGCTGAATATAAGAAGGAAACCTC

AAGTGCTCGCAGAAGCATCGGCAAGTCAGGCGTCTGTTCTCGGATGGATACCCATTACCATCTTGGAC

25 CTGGAAGAAGTGTTTCAGACAAGTCTCCCACTGCACAGAAGAGATCACAGAAGGAGTCTGGAATAGAAAG

GCTAACAGAAAAGTGTGTTGGACAGTGGGTGTCGAGCAGTACTCTAAACATGAGTGAAGCCATAAAAGGGT

TCCTGGTCAAGTGTGTGCATACAATCCCTTGGCACATCTTGTGAGACGATCCTTTTAACTCTCCAGG

CCCCTTCCCTTTCATCCAAGACAACATCTCATTCTATGCAACAATTGGTGTGTTGTCTCCTCTTCATTGTC

GTTTTAACCTGCTAATTTGTCACAAGTACAAAAGCAATTTAGGTATGAAAGCCAGCTACAGATGGTAC

30 AGGTGACCGGCTCCTCAGATAATGAGTACTTCTACGTTGATTTTCAGAGAATATGAATATGATCTCAAATG

GGAGTTTCCAAGAGAAAATTTAGAGTTTGGGAAGGTACTAGGATCAGGTGCTTTTGGAAAAGTATGAAC

GCAACAGCTTATGGAATTAGCAAAACAGGAGTCTCAATCCAGGTTGCCGTCAAATGCTGAAAGAAAAAG

CAGACAGCTCTGAAAGAGAGGCACCTCATGTGCAAGTCAAGATGATGACCCAGCTGGGAAGCCACGAGAA

TATTGTGAACCTGCTGGGGGCGTGACACTGTGAGGACCAATTTACTTGATTTTGAATACTGTTGCTAT

35 GGTGATCTTCTCAACTATCTAAGAAGTAAAAGAGAAAAATTTACAGGACTTGGACAGAGATTTTCAAGG

AACACAATTTTCAGTTTTTACCCACTTTCCAATCACATCCAAATTCAGCATGCCTGGTTCAAGAGAAGT

TCAGATACACCCGGACTCGGATCAAATCTCAGGGCTTCATGGGAATTCATTTCACTCTGAAGATGAAATT

GAATATGAAAACCAAAAAGGCTGGAAGAAGAGGAGGACTTGAATGTGCTTACATTTGAAGATCTTCTTT

GCTTTGCATATCAAGTTGCCAAAGGAATGGAATTTCTGGAATTTAAGTCGTGTGTTACAGAGACCTGGC

40 CGCCAGGAACGTGCTTGTCAACCCACGGGAAAGTGGTGAAGATATGTGACTTTGGATTGGCTCGAGATATC

ATGAGTGATTCCAACATATGTTGTCAGGGGCAATGCCCGTCTGCCTGTAAATGGATGGCCCCGAAAGCC
 TGTTTGAAGGCATCTACACCATTAAGAGTGATGTCTGGTCATATGGAATATTACTGTGGGAAATCTTCTC
 ACTTGGTGTGAATCCTTACCCTGGCATTCCGGTTGATGCTAACTTCTACAACTGATTCAAAATGGATTT
 AAAATGGATCAGCCATTTTATGCTACAGAAGAAATATACATTATAATGCAATCCTGCTGGGCTTTTGACT
 5 CAAGGAAACGGCCATCCTTCCCTAATTTGACTTCGTTTTTAGGATGTCAGCTGGCAGATGCAGAAGAAGC
 GATGTATCAGAATGTGGATGGCCGTGTTTCGGAATGTCCTCACACCTACCAAAACAGGCGACCTTTCAGC
 AGAGAGATGGATTTGGGGTACTCTCTCCGCAGGCTCAGGTGGAAGATTCGTAGAGGAACAATTTAGTTT
 TAAGGACTTCATCCCTCCACCTATCCCTAACAGGCTGTAGATTACCAAAACAAGATTAATTTTCATCACTA
 AAAGAAAATCTATTATCAACTGCTGCTTCACCAGACTTTTCTCTAGAAGCTGTCTGCGTTTACTCTTGTT
 10 TTCAAAGGGACTTTTGTAAATCAAATCATCCTGTCACAAGGCAGGAGGAGCTGATAATGAACTTTATTG
 GAGCATTGATCTGCATCCAAGGCCTTCTCAGGCTGGCTTGAGTGAATTGTGTACCTGAAGTACAGTATAT
 TCTTGTAATACATAAAACAAAAGCATTTTGCTAAGGAGAAGCTAATATGATTTTTTAAGTCTATGTTTT
 AAAATAATATGTAAATTTTTCAGCTATTTAGTGATATATTTTATGGGTGGGAATAAAATTTCTACTACAG
 AATTGCCCATTTATTGAATTATTTACATGGTATAATTAGGGCAAGTCTTAAGTGGAGTTACGAAACCCCT
 15 GAAATTGTGCACCCATAGCCACCTACACATTCCCTCCAGAGCACGTGTGCTTTTACCCCAAGATACAAGG
 AATGTGTAGGCAGCTATGGTTGTCACAGCCTAAGATTCTGCAACAACAGGGGTGTATTGGGGGAAGTT
 TATAATGAATAGGTGTTCTACCATAAAGAGTAATACATCACCTAGACACTTTGGCGGCCTTCCAGACTC
 AGGGCCAGTCAGAAGTAACATGGAGGATTAGTATTTTCAATAAAGTTACTCTTGTCACCAAAAAA

By “CD90 polypeptide” is meant a polypeptide or fragment thereof having at least
 20 about 95% amino acid sequence identity to an amino acid sequence provided at GenBank
 Accession No. NP_001298089 that binds an anti-CD90 antibody. An exemplary CD90
 polypeptide sequence follows:

>NP_001298089.1 thy-1 membrane glycoprotein isoform 1 preproprotein [Homo sapiens]
 MNLAISIALLLTVLQVSRGQKVTSLTACLVDQSLRLDCRHENTSSSPIQYEFSLTRETKKHVLFGTVGV
 25 EHTYRSRTNFTSKYNMKVLYLSAFTSKDEGTYTCALHHS GHSPPISSQNVTVLRDKLVKCEGISLLAQNT
 SWLLLLLLLSLLQATDFMSL

By “CD90 polynucleotide” is meant a nucleic acid molecule that encodes a CD90
 polypeptide. An exemplary CD90 polynucleotide sequence follows:

>NM_006288.5 Homo sapiens Thy-1 cell surface antigen (THY1), transcript variant 1, mRNA
 30 AGCAACCGGAGGCGGCGCGCTCTGGAGGAGGCTGCAGCAGCGGAAGACCCAGTCCAGATCCAGGACT
 GAGATCCCAGAACCATGAACCTGGCCATCAGCATCGCTCTCCTGCTAACAGTCTTGCAGGTCTCCCGAGG
 GCAGAAGGTGACCAGCCTAACGGCCTGCCTAGTGGACCAGAGCCTTCGTCTGGACTGCCGCCATGAGAAT
 ACCAGCAGTTCACCCATCCAGTACGAGTTCAGCCTGACCCGTGAGACAAAGAAGCACGTGCTCTTTGGCA
 CTGTGGGGGTGCCTGAGCACACATACCGCTCCCGAACCAACTTCACCAGCAAATACAACATGAAGGTCCT
 35 CTACTTATCCGCCTTCACTAGCAAGGACGAGGGCACCTACACGTGTGCACTCCACCACTCTGGCCATTCC
 CCACCCATCTCCTCCAGAACGTACAGTGCTCAGAGACAACTGGTCAAGTGTGAGGGCATCAGCCTGC
 TGGCTCAGAACCTCGTGGCTGCTGCTGCTCCTGCTCTCCCTCTCCCTCCTCCAGGCCACGGATTTCAT
 GTCCCTGTGACTGGTGGGGCCCATGGAGGAGACAGGAAGCCTCAAGTTCAGTGCAGAGATCCTACTTCT
 CTGAGTCAGCTGACCCCTCCCCCAATCCCTCAAACCTTGAGGAGAAGTGGGGACCCACCCCTCATCA
 40 GGAGTTCAGTGCTGCATGCGATTATCTACCCACGTCCACGCGCCACCTCACCTCTCCGCACACCTCT

GGCTGTCTTTTGTACTTTTGTTCAGAGCTGCTTCTGTCTGGTTTATTTAGGTTTATCCTTCCTTTT
 CTTTGAGAGTTCGTGAAGAGGGAAGCCAGGATTGGGGACCTGATGGAGAGTGAGAGCATGTGAGGGGTAG
 TGGGATGGTGGGTACCAGCCACTGGAGGGGTATCCTTGCCCATCGGGACCAGAAACCTGGGAGAGACT
 TGGATGAGGAGTGGTTGGGCTGTGCTGGGCCTAGCACGGACATGGTCTGTCTGACAGCACTCCTCGGC
 5 AGGCATGGCTGGTGCCTGAAGACCCAGATGTGAGGGCACCACCAAGAATTTGTGGCCTACCTTGTGAGG
 GAGAGAACTGAGCATCTCCAGCATTCTCAGCCACAACCAAAAAAATAAAAAGGGCAGCCCTCCTTACC
 ACTGTGGAAGTCCCTCAGAGGCCTTGGGGCATGACCCAGTGAAGATGCAGGTTTGACCAGGAAAGCAGCG
 CTAGTGGAGGGTTGGAGAAGGAGGTAAAGGATGAGGGTTCATCATCCCTCCCTGCCTAAGGAAGCTAAAA
 GCATGGCCCTGCTGCCCCCTCCCTGCCTCCACCCACAGTGGAGAGGGCTACAAAGGAGGACAAGACCCTCT
 10 CAGGCTGTCCCAAGCTCCCAAGAGCTTCCAGAGCTCTGACCCACAGCCTCCAAGTCAGGTGGGGTGGAGT
 CCCAGAGCTGCACAGGGTTTGGCCCAAGTTTCTAAGGAGGCACCTTCTCCCTCGCCCATCAGTGCCAG
 CCCCTGCTGGCTGGTGCCTGAGCCCCCTCAGACAGCCCCCTGCCCCGAGGCCTGCCTTCTCAGGGACTTC
 TCGGGGGCTGAGGCAAGCCATGGAGTGAGACCCAGGAGCCGGACACTTCTCAGGAAATGGCTTTTCCCA
 ACCCCCAGCCCCACCCGGTGGTTCTTCTGTTCTGTGACTGTGTATAGTGCCACCACAGCTTATGGCAT
 15 CTCATTGAGGACAAAGAAAAGTGCACAATAAAACCAAGCCTCTGGAATCTGTCTCGTGTCCACCTGGCC
 TTCGCTCCTCCAGCAGTGCCTGCCTGCCCCGCTTCGCTGGGGTCTCCACGGGTGAGGCTGGGGAACGCC
 ACCTCTTCTCTTCCCTGACTTCTCCCCAACCACTTAGTAGCAACGCTACCCAGGGGCTAATGACTGCA
 CACTGGGCTTCTTTTTCAGAAATGACCCTAACGAGACACATTTGCCCAAATAAACGAACATCCCATGTCTGC
 TGACTCACCTGGCTGGAACAACATGCTTACTGCCAACATGTGGGCCGAACCACATGGCCCTGGCTCTGGA
 20 ATGCACAAGTGGCTTTGCGTGAATCTGCGCTAAGCTATGCAGTCTGCTTTTCTTCTCAGCTCTGGTAGT
 TCTTCAGAAATGTACCCTCCAGGCACATCCACTATTGCGAGGGTGAGCACGAAGGGTGGGAGATGCCCAT
 GTCCTCAAGGCATCACTTCTAAATCCAAAAGCATCGGCAGGAGAAAGGACTGGGGACAAATACTGTCCC
 TTCGGGAGTAGGGAGGGAACACTGAGGCCCATCCCTGGCTCCTTCCCTAAAAGTAGAGTAAATGGAAGC
 GAGCATCCTGGGATTGGGGCAAGAGGGGGACCGAGGGTAGCTGTGGGTTCCACTGCTGTGAGAGTCA
 25 GAGAGGCAGCCCCAAGCCAGCCTCCCTGCTTTGCCAGGAATTTGGGGGAGGAAGGTGACAGCTGCCCAG
 AGGCTGACTCATCTGATATTTAGCACTGGGTAGGATGATTGTTTCTGAGCATTTTCTTAAAGGCCTCAG
 ATCTAAATTATGCCACCGGCTCCCACTCTTGTACTCTCCGTCAACTTCTCTGCCTTGCCTTCCACCCCT
 GTAGTTACCATAACAGAGGAGGAGGAGCTGTCTTGTCCCAGGTTGGGAGGCTGACAACCCCTTAGCAA
 GATGCTGCCAGCCCAGAGCTCTCCAAGGGAGGAACACCCCTGAGACTCAGGCCCTCTCCTTCAGCCCT
 30 GCTTGGGCTGCAAGCGCCGTGCCAAGGAAAGGCATCTTGGTGAGAAGAGCTGCTGTGGGGGAAGGGAGAT
 CAAATGCCAGAGAAATGTGGGGTGGCCACCCCTCAGGATAGTAAAAGAGTATGGAGGTATTTCTGGAAGG
 AAATGAGCGGCACTGTGTGAAGCCTCGCACCTGTGTGACACTTCTATGGGGTCTTGTGCACACTCTAGT
 ACTATGTCCCTGAAGAGTTTAGCAGCCACACTCTTAGAAGGGTGCTGGGAGATGGTGTGCCCTCTGCAG
 CCATGTTTAGGGGAGCGGAACCTGAGGCCACAGTGGGTGAGATTAGCTCAAGAAGCCACAGAGGCCACC
 35 AGAGGGCCACGGACTTCGGAAGGAGAAGAGAAGAAGAGGCATCAGGCCTCACAAACGCAAACCTACCCA
 GAGATGGGCACAGTGGCTCATGCCTGTAATCCACCACTTTGGGAAGAGGCGGATCGCTTGAGGTCAGGA
 GTTCGAGACTAGCCTCGAAACCCCTATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCCTGCGC
 CTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGCGGAGGTTGCAGTG
 AGCCAAGATTGCACCACTGCACTCCAGCCTGGGAGATAGAGTGAGACTCCATCTCAAAAAAATAAAAAAT
 40 AAAATAAACCTACCCGGAATGACCATGTGAGGACTGGGAGCCCGCAGACTTTCAGCCACAGGCCGCGAC
 AGCCGTGGGTCCCTCCCTGGTCAAGTCAGCAGGCCTTGTGGAGGCTGTGGGGTATCTGTGGTGACTCAGG

TAATTATAGAGGGCTGGCCCCAGCCCTGGTTCTGTACACATGCCCCAACCCCATCCCCATCCACTCCC
 TCGCCAGTCCTAACCTCTTCTGGGTCCCCCCCCCTTCAGCACCTAAGTCCATACCTAGGGCCGTGGAAT
 TCCCGCTCAAGAGCAACAGAAGCCCCCTCTCTGCACCCCCATTTCTGGACTGGATTGTCCACTGAGACGCG
 CAATGTCTGCATCTCTGACATCTAGAGGCTTCTCGGGAAGGGCATGGGGATCTCCGTGAGATGTGGGGA
 5 CTTTCACTGGCCAACCAAGAAATCTACACAGCGTCCGGGGACCTGTGACACACATCCCTCCCGCCTCCTC
 AACCTGATGTCCCTCTCTGAATCTGCAGCTTTCGTGCTGTGAAGGTGTCTTTACATGTGAAACAAACAAA
 CCCAAGTCAAGAGTAAATCATCTCATTTACTAGTGAGAAAATGTTGGAGCTGGAGTCCTTCAGAGAGTCC
 TGGCCAGGCAAGAGGGCCATCAGCTCTTCTGCTCAACAGGGGCTCTCAGCCTCAGGACACTCTCAGGC
 CTGGAATGTCCCCAACACACTCAAGGAGAAACATGTCTGTGCAGACCCACAGGAGGCATCTTTGCCCGG
 10 CACAAGGAAGAGCTGGGGTCAGTGGGACCTGTAGATGTAGACACATCATATGGAGGGTGGGTAGGACCAA
 TGTGGCAGCTTCATGGAGGCCAAGTGTGGCTCTGCACCAGGAAGGGGCTGTGATGGCTGGAGGTGCCCCAG
 CAGTGCAGGCGGGGAGTGCCTGGCAGTGGCGTGGCCAGGTGGAGGCCACCTGTCAAGTTTGCAATAAAGC
 AGTTTCCTGAATTTGGTGAGAA

By “CD45 polypeptide” is meant a polypeptide or fragment thereof having at least
 15 about 95% amino acid sequence identity to an amino acid sequence provided at GenBank
 Accession No. NP_001254727 that binds an anti-CD45 antibody. An exemplary CD45
 polypeptide sequence follows:

>NP_001254727.1 receptor-type tyrosine-protein phosphatase C isoform 5 precursor [Homo sapiens]
 MTMYLWLKLLAFGFALDTEVFVTGQSPTPSPTGHLQAEEQGSQSKSPNLKSSREADSSAFSWWPKA
 20 REPL
 TNHWSKSKSPKAEELGV

By “CD45 polynucleotide” is meant a nucleic acid molecule that encodes a CD45
 polypeptide. An exemplary CD45 polynucleotide sequence follows:

>NM_001267798.2 Homo sapiens protein tyrosine phosphatase receptor type C (PTPRC), transcript
 25 variant 5, mRNA
 GACATCATCACCTAGCAGTTCATGCAGCTAGCAAGTGGTTTGTTCCTTAGGGTAACAGAGGAGAAATTGT
 TCCTCGTCTGATAAGACAACAGTGGAGAAAGGACGCATGCTGTTTCTTAGGGACACGGCTGACTTCAGA
 TATGACCATGTATTTGTGGCTTAACTCTTGGCATTGGCTTTGCCTTTCTGGACACAGAAGTATTTGTG
 ACAGGGCAAAGCCCAACACCTTCCCCACTGGCCATCTGCAAGCTGAGGAGCAAGGAAGCCAATCCAAGT
 30 CACCAAACCTCAAAAGTAGGGAAGCTGACAGTTCAGCCTTCAGTTGGTGGCCAAAGGCCCGAGAGCCCCT
 CACAAACCACTGGAGTAAGTCCAAGAGTCCAAAAGCTGAGGAACCTGGAGTCTGATGTTCAAGAGCAGGA
 AGCAGCCAGCAGAGAGAAAGATGAAGACCAGAAGACTCAGCAAGCTCACTTCTCCTACCTTCTGTGCC
 TGCTTTTTCTAGCCGTGCTGGCAGTTGCTTGGATGATGCCCACTCATATTGGGTGGGGTGGGGGGTTG
 GGGAGGGTCTGCCTCCCCAGTCCACTGACTCAAATGTTAATCTCCCTTGGCAATACGCTCACAGGCACA
 35 CCCAGGAACAATACTTTGCATCCTTCAATCCAATCAAGTTGACACTCAATATTAACCATCAAATACTATT
 ATAAGGAGAAATGTTGCATGATTTTCTTCTAGTCTGTTTGTAAATCACATCTAATGAAAGAGTGAGAGTG
 GACGATAAAGGGAACCTGTTGAAACATTTCTCTCAAAGCAAAAGGGATCATTTGGAAGCAGGCAGACACCA
 GAATTGGTTTAACCTAAAAATAACAAATTAATAATTATCAAGTCTATAATGATGACAGTGACTTAATGTG
 AATAGAAAGAATTTCTAACTCTCTCCTTCCCTCCCTTCTTCTACTTTCTTTCCACTCCCTT

TCTCCACCCCTTTCTTTCTTTCTTTCTCCCACCCTCTCTCCCTCCCTTTCTTTTATTCAATGCA
 TAGTAGTTGAAAAATCTAAAGTTAGACCTGATTTTACTGAAGACTAGAGGTAGTTACTATCCTATTA
 CTGTACTTAGTTGGCTATGCTGGCATGTCATTATGGGTAAAAGTTTGATGGATTATTTGTGAGTTATTT
 GGTATGAAAATCTAGAGATTGAAGTTTTTCATTAGAAAATAACACACATAACAAGTCTATGATCATTTT
 5 GCATTTCTGTAATCACAGAATAGTTCTGCAATATTTTCATGTATATTGGAATTGAAGTTCAATTGAATTTT
 ATCTGTATTTAGTAAAAATTAACTTTAGCTTTGATACTAATGAATAAAGCTGGGTTTTTTATTTA

By “CD34 polypeptide” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_001020280 that binds an anti-CD34 antibody. An exemplary CD34
 10 polypeptide sequence follows:

>NP_001020280.1 hematopoietic progenitor cell antigen CD34 isoform a precursor [Homo sapiens]
 MLVRRGARAGPRMPRGWTALCLLSLLPSGFMSLDNNGTATPELPTQGTFSNVSTNVSYQETTPSTLGSTSLHPV
 SQHGNEATTNITETTVKFTSTSVITSVYNTNSSVQSQTSVISTVFTTPANVSTPETTLKPSLSP
 GNVSDLSTTSTSLATSPTKPYTSSSPIILSDIKAEIKCSGIREVKLTQGICLEQNKTSSEAEFKKDRGEGL
 15 ARVLCGEEQADADAGAQCSSLQAQSEVRPQCLLLVLNRTEISSKLQLMKKHQSDLKKLGILDFTEQDV
 ASHQSYSQKTLIALVTSGALLAVLGITGYFLMNRWSPTGERLGEDPYTENGGGQGYSSGPGTSPEAQ
 GKASVNRGAQENG TGQATSRNGHSARQHVVDTEL

By “CD34 polynucleotide” is meant a nucleic acid molecule that encodes a CD34 polypeptide. An exemplary CD34 polynucleotide sequence follows:

>NM_001025109.2 Homo sapiens CD34 molecule (CD34), transcript variant 1, mRNA
 AGTGTCTTCCACTCGGTGCGTCTCTCTAGGAGCCGCGCGGAAGGATGCTGGTCCGACAGGGCGCGCGCAGGG
 CCCAGGATGCGCGGGGCTGGACCGGCTTTGCTTGTGAGTTTGTGCTTCTGGGTTCATGAGTCTTGACAAC
 AACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTTTCTACAAATGTATCCTACCAAGAA
 ACTACAACACCTAGTACCTTGAAGTACCAGCCTGCACCTGTGTCTCAACATGGCAATGAGGCCACAACAAC
 25 ATCACAGAAACGACAGTCAAATTCACATCTACCTCTGTGATAACCTCAGTTTATGGAAACACAACTCTTCTGTC
 CAGTCACAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAACGTTTCAACTCCAGAGACAACCTTGAAG
 CCTAGCCTGTACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCACTAGCCTTGCAACATCTCCCACTAAACCC
 TATACATCATCTTCTCCTATCCTAAGTGACATCAAGGCAGAAATCAAATGTTTCAGGCATCAGAGAAGTGAAATTG
 ACTCAGGGCATCTGCCTGGAGCAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGGACAGGGGAGAGGGCCTG
 30 GCCCGAGTGCTGTGTTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTCCTTGCCAG
 TCTGAGGTGAGGCCTCAGTGTCTACTGCTGGTCTTGGCCAACAGAACAGAAATTTCCAGCAAACTCCAATTATG
 AAAAAGCACCAATCTGACCTGAAAAGCTGGGGATCTAGATTTCACTGAGCAAGATGTTGCAAGCCACCAGAGC
 TATTTCCCAAAGACCTGATTGCACTGGTCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCT
 CTGATGAATCGCCGAGCTGGAGCCCCACAGGAGAAAGGCTGGGCGAAGACCTTATTACACGAAAACGGTGGA
 35 GGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGGAAAGGCCAGTGTGAACCGAGGGGCTCAG
 GAAAACGGGACCGGCCAGGCCACCTCCAGAAACGGCCATTTCAGCAAGACAACACGTGGTGGCTGATACCGAATTG
 TGACTCGGCTAGGTGGGGCAAGGCTGGGCAGTGTCCGAGAGAGCACCCCTCTCTGCATCTGACCACGTGCTACCC
 CCATGCTGGAGGTGACATCTTTACGCCCAACCTTCCCCACTGCACACACCTCAGAGGCTGTCTTGGGGCCCT
 ACACCTTGAGGAGGGGAGGTAACTCCTGTCTTTACACATTCGGCTCCCTGGAGCCAGACTCTGGTCTTCTTT
 40 GGGTAAACGTGTGACGGGGGAAAGCCAAGGTCTGGAGAAGCTCCAGGAACAATCGATGGCCTTGACAGCACTCAC

ACAGGACCCCTTCCCCCTACCCCTCCTCTCTGCCGCAATACAGGAACCCCAAGGGGAAAGATGAGCTTTTCTAG
 GCTACAATTTTCTCCCAGGAAGCTTTGATTTTTACCGTTTCTTCCCTGTATTTTCTTCTCTACTTTGAGGAAAC
 CAAAGTAACCTTTTGCACCTGCTCTCTTGTAAATGATATAGCCAGAAAAACGTGTTGCCTTGAACCACTTCCCTCA
 TCTCTCCTCCAAGACACTGTGGACTTGGTCACCAGCTCCTCCCTTGTCTCTAAGTTCCACTGAGCTCCATGTGC
 5 CCCCCTCTACCATTTGCAGAGTCCTGCACAGTTTTTCTGGCTGGAGCCTAGAACAGGCCTCCCAAGTTTTAGGACAA
 ACAGCTCAGTTCTAGTCTCTCTGGGGCCACACAGAACTCTTTTTGGGCTCCTTTTTCTCCCTCTGGATCAAAGT
 AGGCAGGACCATGGGACCAGGTCTTGGAGCTGAGCCTCTCACCTGTACTCTTCCGAAAAATCCTCTTCCCTCTGAG
 GCTGGATCCTAGCCTTATCCTCTGATCTCCATGGCTTCTCCTCCCTCCTGCCGACTCCTGGGTGAGCTGTTGC
 CTCAGTCCCCAACAGATGCTTTTCTGTCTCTGCCTCCCTCACCTGAGCCCCCTCCTTGCTCTGCACCCCCATA
 10 TGGTCATAGCCCAGATCAGCTCCTAACCTTATCACAGCTGCCTCTTCTGTGGGTGACCCAGGTCTTGTGTTGC
 TGTGATTTCTTTCCAGAGGGGTGAGCAGGGATCCTGGTTTCAATGACGGTTGAAATAGAAATTTCCAGAGAA
 GAGAGTATTGGGTAGATATTTTTTCTGAATACAAAGTGATGTGTTTAAATACTGCAATTAAAGTGATACTGAAAC
 ACATCTGTTATGTGACTCTGTCTTAGCTGGGTGTGTCTGCATGCAAGAGTGACACCCCTCCATTAGACCTAGCTAG
 ACTGTGCAGTGATGTGGTGGGAGGACCAGCCAGGGAAGAGGGAGCACCTCAGCAGACACAGGCACCAGCCAGGA
 15 TGCTAAGGACCTTTAGCCAAGTCTGCCAACTATTCTCCTCCATGGGGAGAGGAAACATCCATTTCCAGTGGTAGA
 AAGGCAGACCCGAATGTACCAGGGAGCTTCCAAATGGAGGGTGGTATGTTGGGTCTTAGGAGCTGTACCCTTCA
 TGAACACCCCTCTGAGAAGAGGAGCATGCTGATCACTGCTGCAAAATATGCAAAACAAAGGGAAGGGGCAATGTC
 CTGTGCACCCCTTATTATCAGGCCACCCCCCTCCCCAGCCCCCAGGTCAGAGTAGACACAGTGAAGGACTATGT
 GGGGACTGTTGTTCTAGAGACCTGGCAGCCAACCTCAGGGAGGGGGCTGGTTTCCACCCTCAAGATTAAGACAGCA
 20 GCCTAATTAATAAAAAAATCTGTAAGCATGTACCTCCCCCAGCTTCCAAAACAACCCCCACCCACCCCTACCA
 GGCCATAGGAAGTTGGGGAGGGAGTGCTGAGGAGCTCCAGGAAACACTCCCAAGTGTGTGACAGTGGCAGAGGC
 AGTTGGGGCCAAACAAAGGTTGATTCTTCCATTCTTATCTCCATAAAGCCAGACCTTTCCCTTCAGCACTCCTCC
 ACCCCCATCTCCTTCTTGCTTTTCTCCAACCTCCTCTAATCATAGGTTCTTCCCTAGGACAGAGGGGAGGCGAAAT
 GATGAGGTTCCAGAGTCTTCCCTCAAAGGCGATGGCTGCCTTGAGGGTTGGAGCAAAGGATGATGAGCAAAGACG
 25 ATGGTAATCAGTAGGGAAGTCCAGCCCACTTGCATCTAGTTGCACATCTTGCCCTGAGAGTAATCCAGTGAGGGT
 CTGTCCCAGCTAGGACATCAAGTAGGAGGGGTGGGTTGAGGGTTCCAGATTCTTAGGAAATATGGGAGGAGAGGAA
 AAGGCAACTTGGATGCACCTCCAGCTTCAGGCCTAGCAACCTGCAATGCATCTCACCCCTGAGTTTGCTGGAATGT
 GTATGTATGCTTTGGGAGGAAGGGCTGTGTGTGATTGCGGGGTGGGGTGGGCAGCTGGTTCCCTCTGACAGCT
 GGACAGCTTGCCCTGAAGAATTTGCCGTGCTTTCTGGAAAAATCCAACCTTTCCACCGTGGGCCTGAGCGTCCTGG
 30 TACAGCAATGGCGCCACCTGCTGGCCTTATTGAGGTCCTACTGCTCAGCCTCAGCTCAATCGCCTCCATGTTGGG
 CTTCTCTCCCTGGCTGCCCCACCCTCTAGTCCAATTTCTCTTGACACAAAGCTCATATAACTATAGAACGTAC
 TGTTGAAGAGAACTTTAAAGATACATTTAATTAACTCCCTTATGGTATAGTTAAAGACAAACTAAGGCTCAGAG
 AAGGGAGGTGGCTTGCCCAATCACCCAGAATTCCAAAGTCTGAATCTGTAGTTTTCCCTTCCATCATATCATCC
 TACTCTTCTGCCGAGTCTCCGTGTTACTCCAGTTGGATGTCATGAAGCCAGTGTGGCAGTGTGAAGATAGGTTT
 35 GGGACTTCACCTCTGGAGCATTTTCATCAACATAAGCTATCCTAGGCCTGGCCAGCCAAGCAGGTCTTGAGGAGC
 CCCAGGACAAAGATCACAGGAGGCCATGAGGTTGGGCTTCTTCGGCGCCACAGTGAGCCCAGGAAATTAGCTG
 TAGGGTATTACACTGTTGACTATGGAGAGCATATCTGGAATTATCTTCAGCCAGATTTTCATCTGAATGGATAAA
 TGGGAATACCATCTAAGTCCAGATAAATAGATCACTTCCATCTCATCCCTTCTAGGTAGATTAATCCCACACTTC
 CTCTTCACACAAAACAGTAATAGGTCATCGATTTTGTGCAACAGGATGCTGCTTCTCTTCTTAAAGCCCCCATC
 40 GAAGAGGCTTCCAGCCACCATTCAATCATTCATCAAGTCTTATGATGTGCCAGACACTGCGCGAAATGTGCCAGA
 ACATCTGTTATGTGCCAGACACTGTTCTTGAGACTGGGGATACAGCAAACACTCATGAAGCTTATAATTCTAGCA

GAAGAGGACAGTAAACAATGTCATCTCAGTAAGTATATACATGTGTTTTCAGGATTGAGAGCTATGAAAAACATA
AAATATATTGAGAATAATGGTTGGTATTTTACATATGGTGGTACTTTTAGAAAAATAACAGTGGAGAGCACAGC
TTCACCTGAATGAAGTGGAGAAGCAGGTTGTATGCCAAGCTGGGAGAGATTATCCACACAGGGGAAAGGACAAG
TGCAAAGCCCTATGATGAAAAGCTGCCAAGTGCAAGAGCCTCAGATGGCAGGGGGCAAGATGGCCATGAGGTTG
5 TGTCAGTGAGTGGGGGTGGGGAGAGGCAGGAGGTCAGACTACATGGGGCCTTTTGTAGTTGTAGATTGGGAAGCCA
CTGGAGGGTTTTGAGCAGAGAAGTCATATCATCTGCTTTATGTTTTAAAGGATCATGCTGGCTGCTGAGTAGAG
AATAGAGGTTGAGGGATAAGAAAGTAGAAGGAGACCGTAGCAAGAAGAACGATCATGGCTGGGAGCAGGTGATCA
TATTGGCAGTGATGAGATCAAGCAGAATTCAAAAAGTGGTTTCAAAGTAGAGGTAACAGGACTTGCTCAGTCTAT
TTATTTCTTCAAATAATAATCATATTTACAATGATAGTAGCTAACAGTTTTTGTAGTGCTTACTGTATGAAAATTG
10 AGATATGGTGCCAATATTTAAATAGCATATTTTACTTAACATTCACAGAAACCCTGTGAAGTAGGTTCTATTATC
TCAGAAAAAGAACTGAAACTCAGAGAATAACAAGGGACTGTGTTACGTGCACAGTGGCAGAGGCCAAGATGAAT
AGGATGTGAGTTTATTTGAACCCCAAATGTTTAAATCTTGGGGATAATACAACACACATTTAAACAAAGAAGCAA
GAAAAAAATGCACAACAGAAAAGTGAGAAATAACACGAGGAAAGACTAAATGAAGTGCTTTGTATCTAGATGTGG
GCAGGACCCCTTCCAGCTGAGAAGATCTGAGACTGGGTCATGAACAGGTGGTTTCTGAGTGGGTCTGTAAAAAT
15 GAATACGATTTTGATGATAGTAATGAGTAAGGACATTTGAGACTGATAGAAGAGTACATACAATATGTAGTGATG
GGGAAAGATAAGGTAAGTGTCAAAGGACAATGTGTTTTCTGGTATGACAGAGAAGTAGAATGTGTTAAGGGAAGCC
GAGTACCAGAAAGATCCGGGTGTCACAGTTTGTGTAGGGTGTTTAAAGCTAAACCACAGAGTTTAATTTTATCCA
ATAGAAGAGGAGCCACAGAAGAGTTTCCATTTATTCATTAATTTATTCATTTATTCAAAAATATTTGAGTGCTT
ATTATAAGCCAGGTACTATGCCAGGCACCTGGGATAAGACATAGTCCCTTCTGTCAAGTCTTTACATTGGGTGGA
20 TGTGGGAGGGACAGATGACAGAACAATATGCATTGAGTGTAAGTGCTATGGTATAGGAAGCTCTGAGTGGGAGGG
GCATGGAAGCCGTGGAAGACCATGGAAGGCTTCCCAGGAGAAGTGACGTCTGGACTGATCCTTTGGTCAAGCAGG
AGTTAAAGAGGAGAAAAGGAGAGATATGGGTGTTCCCAGAGAGGAAGAAGCCTTGTCCCAGGAGCAAAGTGAGG
GTGATTGTTCCAGAAATGTGAGTGATTCTTTTAAGGCTCAAGCAAAGCATGTGATTCTTCTTTATACCTTCTATT
TCTTTGCTGAGTGTTTCTGTTCTTTTGTTCGATGCAATTGCTCATTAAGCATGTTTATGATGGCTGT
25 CTGTTTTAAATTTCTTGTGATGTTTCAACATCTTTATCATCTCAATGTTGGCATCTGTTAATGGTTTTTCT
CAATCAAATTTGAGATTTTCTGTTCTTGGTATTACCAAGTGATTTTAAATGCATCTGGAAATTTGGGATTTATGT
TGAAAGACTGGATCTTATTGAAAGATTCTGTTTAGACCCCTCCTTTGATACCACACTGGTGGGTCCAGGTTCCC
CATTACAGCTGTTGACACCTTCAGGGCAGAGAGGTGGGATGGGGTGAAGGGGTACCTCATTATTGCTGGCCCAGG
TTAGAAGTTCAGGCTTCCCAGTAGATCTCTGCTGATACCACCTGGTGCCATGTCAATTCCTTGAGTCCAAAAGTC
30 CCTCCCAATTTCTGCCTTCTTCTCTACATATCGGAGTCTCCCTATGTTTGACTTATATATAATGTCCAGGGTTT
TTAGAGTTAGTTAACAGGAGGCATAAGAAAAAGTGTGTCCACTCCATCTTGTCTGGAACTGGAAGTTCAAGTCGA
ATATAAGAGAGAGGAGAGGAAATTACAAGCCATGAGACTGGAGAGTTAGGCAGGTTCTACACCAGCTATTCTCAA
AGCCCTCTTACACTCTTAAAAATTTAGAACTTCAAAGAGCTTTTGATTTTGAAAGTTACATCTATCAATTATTAC
TGTTTCAAATTTAAATTTGAGAAAATTTATTTATTAATTTGTTTAAAAATAACAATAATTATTCAATTACATG
35 ATAATGTAAGTAATGCTTTTCTTAATGAAAAATAATTATATTTCCAAAACAAAAACAATTAGGAAAAAGAGTGT
CATTTGTTTTAGACTTTGGTAAATCTCTCTAATATCTGGCTGAAGAGAAGAATGCTGATTCTTTTTTTTTTTTTT
TTTTTTGAGACGGAGTCTCGCTCTGTCAACCAGGCTGGAGTGTAGTGGTGTGATCTCGGCTCACTGCAAGCTCTG
CCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCACCCGCCACCACGCCCG
GCTAATTTTTTTGTATTTTGTAGTAGAGATGGGGTTTACCCTGTTAGCCAGGCTGGTCTCGATCTCCTGACCTCA
40 TGATCCACCCACCTCAGCCTCCCAAAGCGCTGGGATTACAGGTGTGAGACACCGCGCCAGCCCCGAATGCTGA
TTCTTTTATCTGCTTCTGTATTCAATCTGTTGTGATATGATGGGTAGCCTCTGAAACACTCCACTGTATACTGT

5 GAAAGAATGAATGTGAAAAAGGAAAATAGATTTGTAGTATTATTATTCAAATTGTTTTGACCTCAGAGACCACTT
 GGAAATGTTTTAGGGAACCCCCAGAGGACCTTGGATCATGCTTTGAGAACCGCGGCTCTAGATATGTTACTATTT
 CAGTAGCATCTAAGTACATGTGGCTGCTGAGCACTTGTAATGTGGCTAGTGCAAATGAGAGACAGGACTTCCAGC
 TATATGTAATTTAATAAACTCAAATTTAAAACTGGAACCTCATAAAATGTTTTGTTGTTGTTGTTAAACATGAC
 CTTATAGTTTTGGTAGGAA

By “Stem Cell Factor (SCF) polypeptide” is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to an amino acid sequence provided at GenBank Accession No. NP_000890 that functions in hematopoiesis. In some embodiments, a SCF polypeptide or fragment thereof binds CD117. An exemplary SCF polypeptide sequence follows:

10 >NP_000890.1 kit ligand isoform b precursor [Homo sapiens]
 MKKTQTWILTLCIYLQLLLFPVKTEGICRNRVTNNVKDVTKLVANLPKDYMITLKYVPGMDVLP SHCWI
 SEMVVQLSDSLTDLDDKFSNISEGLSNYSIIDKLVNIVDLVECVKENS SKDLKKSFKSPEPRLEFPPEEF
 FRI FNRSIDAFKDFVVASETSDCVVSSSTLSPEKDSRVSVTKPFMLPPVAASSLRNDSSSSNRKAKNPPGD
 15 SSLHWAAMALPALFSLIIGFAFGALYWKRPQPSLTRAVENIQINEEDNEISMLQEKEREFQEV

By “SCF polynucleotide” is meant a nucleic acid molecule that encodes a SCF polypeptide. An exemplary SCF polynucleotide sequence follows:

>NM_003994.5 Homo sapiens KIT ligand (KITLG), transcript variant a, mRNA
 GGGCTTCGCTCGCCGCCTCGCGCCGAGACTAGAAGCGCTGCGGGAAGCAGGGACAGTGGAGAGGGCGCTG
 20 CGCTCGGGCTACCCAATGCGTGACTATCTGCCGCCGCTGTTTCGTGCAATATGCTGGAGCTCCAGAACAG
 CTAAACGGAGTCGCCACACCACTGTTTGTGCTGGATCGCAGCGCTGCCTTTCTTATGAAGAAGACACAA
 ACTTGGATTCTCACTTGCACTTTATCTTCAGCTGCTCCTATTTAATCCTCTCGTCAAACTGAAGGGATCT
 GCAGGAATCGTGTGACTAATAATGTAAAAGACGTCATAAATTGGTGGCAAATCTTCCAAAAGACTACAT
 GATAACCCCTCAAATATGTCCCGGGATGGATGTTTTGCCAAGTCATTGTTGGATAAGCGAGATGGTAGTA
 25 CAATTGTCAGACAGCTTGACTGATCTTCTGGACAAGTTTTCAAATATTTCTGAAGGCTTGAGTAATTATT
 CCATCATAGACAACTTGTGAATATAGTGGATGACCTTGTGGAGTGCCTGAAAGAAAACATCATTAAGGA
 TCTAAAAAATCATTCAAGAGCCAGAACCCAGGCTCTTTACTCCTGAAGAATCTTTAGAATTTTAAAT
 AGATCCATTGATGCCTTCAAGGACTTTGTAGTGGCATCTGAACTAGTGATTGTGTGGTTTCTTCAACAT
 TAAGTCCTGAGAAAGGGAAGGCCAAAAATCCCCCTGGAGACTCCAGCCTACACTGGGCAGCCATGGCATT
 30 GCCAGCATTGTTTTCTCTTATAATTGGCTTTGCTTTTGGAGCCTTATACTGGAAGAAGAGACAGCCAAGT
 CTTACAAGGGCAGTTGAAAATATACAAATTAATGAAGAGGATAATGAGATAAGTATGTTGCAAGAGAAAG
 AGAGAGAGTTTCAAGAAGTGAATTGTGGCTTGTATCAACACTGTTACTTTTCGTACATTGGCTGGTAACA
 GTTCATGTTTGCTTCATAAATGAAGCAGCTTTAAACAAATTCATATTCTGTCTGGAGTGACAGACCACAT
 CTTTATCTGTTCTTGCTACCCATGACTTTATATGGATGATTAGAAAATTGGAACAGAATGTTTTACTGTG
 35 AAAGTGGCACTGAATTAATCATCTATAAAGAAGAACTTGCATGGAGCAGGACTCTATTTTAAGGACTGCG
 GGACTTGGGTCTCATTTAGAACTTGCAAGCTGATGTTGGAAGAGAAAGCACGTGTCTCAGACTGCATGTAC
 CATTTGCATGGCTCCAGAAATGTCTAAATGCTGAAAAACACCTAGCTTTATTCTTCAGATACAACTGC
 AGCCTGTAGTTATCCTGGTCTCTGCAAGTAGATTTAGCTTGGATAGTGAGGGTAACAATTTTCTCAAA
 GGGATCTGGAATAATGTTTAAACTCAGTAGTGTACGCCACTGTACAGTGTAGAAAGCAGTGGGAAGT

TGATTGGATTTGGCAACATGTCAGCTTTATAGTTGCCGATTAGTGATATGGGTCTGATTTTCGATCTCTTC
CTGATGTAAACCATGCTCACCCATATCCCCTATACAAATGCAAATGGTTGCCTGGTTCCATTTATGCAA
GGGAGCCAGTACTGAATTATGCCTTGGCAGAGGGGAGACTCCAAAAGAGTCATCGCAGGAAGAAGTTAAG
AACACTGAACATCAGAACAGTCTGCCAAGAAGGACATTGGCATCCTGGGAAAGTCCGCCTTTTCCCTTGA
5 CCCTATAGGGTGTATAAATCGTGTTTGCAAATGTGTTATGATGTGTTTATATTCTAAACATTATACAG
AGCTATGTAAAGGGACTTAGGAGAAAATGCTGAATGTAAAGATGGTCCCATTTCATTTCCACCATGGGAG
AGCCTAAAAATAAATTATGACATTTAGTATCTAAGGTTAGAAAACCACGCCACATGCTAATATGGGTGT
TGAAAACCTAGGTTACTTATAATGCAAGGAATCAGGAACTTTAGTTATTTATAGTATAATCACCATTATC
TGTTTAAAGGATCCATTTAGTTAAAATCGGGCACTCTATATTCATTAAGGTTTATGAATTAAGAAAG
10 CTTTATGTAGTTATGCATGTCAGTTTGCTATTTAAAATGTGTGACAGTGTTTGTATATTAAGAGTGAAT
TTGGCAGGAATTCCTAAGATGGACATTGTGCTTTTAACTAGAACTTGTAAGACATTATGTGAATATCCC
TTGCCAATTTTTTTTATAATAAGAAAACATCTGACTAAAGTCAAAGAATGATTTCTATGGTTTATTTTG
ATGAAAGTTCCTTTAACATGTCTTGAATGTACACATAAAGGAATCCAAAGCTTTCCATTCTAACTTAATC
TTTGTGATAACATTATTGCCATGTTCTACAACCGTAAGATGACAGTTTCAATGTAGTGACACAAAAGGG
15 CATGAAAACCTAACTGCTAGCTTTCTTTCATTTCAAAGTCCAAGAATTTCTAGTATATTTGGATTTTA
GCTTCTGTTCAAAGCAAATCCAGATGCAACTCCAGTAAGTGGCCTTTGCTCTTTTTTGTACCAAAGAGCC
CAGATGATTCCTACAGTCCCTTTCTTCTCTAACATGCTGTGGTTCCTTAAATATGAGTAATTTCTCTAAG
ATATAACCCAGGTGCTTTGAGAAGCTGCATTAAGGTGTTCAGGCCCTCAGATATCACATGGTACACTTGA
TTAGTAATAAAACCAGAGATCAATTTAAATTGCTGATAGGTCTGTCTCAGTGTGTGGCATTGACTGTTT
20 TCAGGAAAATAGATACAGATTAATATGAGTTATGCGTGTAGGTTGTGTATAGATTGAGAAGATAGATACT
TCTCAATCTAGTAGTTTGATTTATTTAACCAATGGTTTCAGTTTGCTTGAGCATATGAAAATCCTGCTTA
ATGTGCTTAAGAGTATAATAAATGTGTACTTTTGTCTCAAACCTAGTAGCTGGGTTTTAACTCATGG
ACATGGTCTTAAATCAATGGAGTTAAATAACAAATTCAGCAAGTTATTAAATCTGACATGGTAGGAGAGG
GGAGATGTGCTGCTTATTAAATGTGTGGTCCATTGAAAGTTACATGGATTGCCAATTTTTTAAACAC
25 TAAAGTTGAATAAAATGCATGAACAATAGAAAATGCTGAACATTATTTTGGATGCTAGCTGCTTGGACA
TTAACTGTGTATTTCTGCTTTGAGATGAAAATATATATTTATCTTTGCTTATTTTATCCCAGATGTGTT
CTGAATATCCTTCTTCATAAATCATGGAACCTCACTGCTGAGATAGTAAACCATGAAATCGCCTTTTCA
GTTGGTGCCATGTATCTGACAGTTCATCTTGGAAGGTTTCAAATTTACCTTTTAAATGATCTCAGAAG
TCTGTAGATTCTCAATGATACTGAAAGCTTTGCACCTCTTTGGTAGAAACCAGGTCTATTTAGAAAATGG
30 CTTTATGATAAATGTTGCCCTCCTGAGTGATAATGAAGTGTTCCTGGATATTGTATTGTAATTTAATGTGC
TTACCACACTGCCACATTTTAAATGAGTCAGAGAAAAATTAATTTTCTTCAATACAATAATAGAACAGT
AGCCTATTCTCTAAAAAGTATGTGAAAAGAAAATTATGAAAAATATGCATACCTAATGAAGTATTGGT
TTTAGTAAGAATTAAATACATTTCAATTGAGCTTTAAAGTACTTTGGAGAACTTTGGGGCACGTTTTCT
ACTCTAATTCAACTAAAGTTATAAATAAAGAGAAAACTCATTGAGAAATCATGGATTTTAAAAATATTT
35 TACTGCAGCCAGTTTTTCATTTCAAATGTAATTTTCAGTTTGGAGCTTTTAGGCATTATGTATATTTAAA
AAATATATTTCTTCAAATGCAATTTTGGCATGGTGGGATGGATGTTGCAAAGATATCCGGAGCCTCCAG
TCTGTCAATTAAGTATGGTAAATCACCTCTCTTCTTTGGGTCTCAATTTTTTATTTATCTATATGGTA
AACTCAGAGATCACTCCTTAGGGGTGAGTCCATTGCAATATGACCGACAAAGAAGACAAAATAGCATTG
AACTAACCACATACAAAATATCCAACCTCGGATTCTGTGAATAAGTATCTTGACCATAAAAAGTCATTGC
40 TGTTCTTGTTTCTAATGTAAATAGTGTCCATTAGTAAAAGTGAATTCAGTCTTAAAGTAGGGTGAATTGG
ATCACCATTTACACAAGAGATGGCTTTTCTTTGCTTGAATAAACATTTTGGATCACCTCCAAAGAATG

AAAACCAGTAGTACGTTTTAGTCATATTAGTCAGGATGAGAACTATAAGATGTGTGTAAACATTTGGAAA
 TGCACCAAAGTGAGCGTTTAAATCTTCTCATTTTATTGAAAAC TAAGAGCAGAAAATGTAAATGCTCAT
 GAAGGTTTTGAATGCCAAAAGATATTTTAGAATCAATTTATAAAGGGTAATTCATTAATTACACTTTAA
 AATTGGAAAGTGGGATAAGAAATCTAAAGTAAACCAGCTTATCTTTGAAACAATATTATTTTGAATTTGG
 5 CTTTAAATAAAAACCATTGAGATTGAAATTCTAATTAGCTCATTTGTGGAGTTTGATCACACAATTCATA
 ATGTTGCTGCTTTCCATTAACTAGTCTTGAAATGCCTTTGTTTGTAATAATAATAATGGTACTTTTCAT
 TTTATAACAAGGTGTTTTTTTCAAGAAATAATCCATGCTAAAATGGATATTTGTGATCCTGAAATGTTTA
 CTAAGCATTTGTAATTTTATTATAACTGCCATCTCCAAC TACATCCTTATGATGTTTTTAACAATAAAAT
 TAAAACACTGTAAACTAAAAACCACCGTTTTTCCAGTACTTGATCTCTGAGCTACAATACTCACTAA
 10 ATATAATTTTCCAATCAAAATATTCTATTCTATATTCTAAGGGTTAATATGTGATTATAGTGTCCTTG
 CCACCATTTTTTTTAAATCAATGGACTTGAAAAGTATTAATTTAGATGGATGCGCAGATATACCCCTCAGTT
 CAGTCATAGATTGGAGTTTGCATATAATAATGTAAATGTATGTCGACACTATTCTAAATAGTTCTATTAT
 GACTGAAATTTAATTAATAAAAAAGGTTGTAAATGTGATGTGTATGTGTATATACTGTATGTGTACTT
 TTTAAATAGGTGTATGTCCCAACCCTTTTTTATACAGGTTTGAATTTAAATTTACATGATATATACATA
 15 TACTTTATTGTTCTAAATAAAGAATTTTATGCACTCTCAAAAAAAAAAAAAAAAAAAAA

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
 20 (*e.g.*, dCas9) and a deaminase domain ((*e.g.*, an adenosine deaminase, a cytidine deaminase,
 or an adenosine deaminase and 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
 25 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
 30 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
 35 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.

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.

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, 65, 70, 75, 80, 85, 90, 95, 100, 101, 102, 103, 104, 105, 110, 120, 130, 140, 150, 160, 175, 180, 190, or 200 amino acids in length.

In some embodiments, the domains of a base editor are fused via a linker that

5 comprises the amino acid sequence of SGGSSGSETPGTSESATPSSGGS, SGGSSGGSSGSETPGTSESATPSSGGSSGGS, or GGSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS. In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence

10 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

15 SGGSSGGSSGSETPGTSESATPSSGGSSGGSSGGSSGGS. In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPSSGGSSGGSSGGSSGSETPGTSESATPSSGGSSGGS. In some embodiments, the linker is 92 amino acids in length. In some embodiments, the

20 linker comprises the amino acid sequence PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS.

“Makassar” or “Hb G-Makassar” refers to a human β -hemoglobin variant, the human Hemoglobin (Hb) of G-Makassar variant or mutation (HB Makassar variant), which is an

25 asymptomatic, naturally-occurring variant (E6A) hemoglobin. Hb G-Makassar was first identified in Indonesia. (Mohamad, A.S. et al., 2018, *Hematol. Rep.*, 10(3):7210 (doi:10.4081/hr.2018.7210). The Hb G-Makassar mobility is slower when subjected to electrophoresis. The Makassar β -hemoglobin variant has its anatomical abnormality at the β -6 or A3 location where the glutamyl residue typically is replaced by an alanyl residue. The

30 substitution of single amino acid in the gene encoding the β -globin subunit β -6 glutamyl to valine will result as sickle cell disease. Routine procedures, such as isoelectric focusing, hemoglobin electrophoresis separation by cation-exchange High Performance Liquid Chromatography (HPLC) and cellulose acetate electrophoresis, have been unable to separate

the Hb G-Makassar and HbS globin forms, as they were found to have identical properties when analyzed by these methods. Consequently, Hb G-Makassar and HbS have been incorrectly identified and mistaken for each other by those skilled in the art, thus leading to misdiagnosis of Sickle Cell Disease (SCD). In one embodiment, the valine at amino acid
5 position 6, which causes sickle cell disease, is replaced with an alanine, to thereby generate an Hb variant (Hb Makassar) that does not generate a sickle cell phenotype. In some embodiments, a Val → Ala (GTG → GCG) replacement (*i.e.*, the Hb Makassar variant) can be generated using an A•T to G•C base editor (ABE).

Thus, the present invention includes compositions and methods for base editing a
10 thymidine (T) to a cytidine (C) in the codon of the sixth amino acid of a sickle cell disease variant of the β-globin protein (Sickle HbS; E6V), thereby substituting an alanine for a valine (V6A) at this amino acid position. Substitution of alanine for valine at position 6 of HbS generates a β-globin protein variant that does not have a sickle cell phenotype (*e.g.*, does not have the potential to polymerize as in the case of the pathogenic variant HbS). Accordingly,
15 the compositions and methods of the invention are useful for the treatment of sickle cell disease (SCD).

By “marker” is meant any protein or polynucleotide having an alteration in expression level or activity that is associated with a disease or disorder.

The term “mutation,” as used herein, refers to a substitution of a residue within a
20 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,
25 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
30 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.

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.

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.

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-hydmethylcytosine. 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.

10 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”
15 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
20 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
25 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
30 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).

10 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 or guide polynucleotide (*e.g.*, gRNA), that guides the napDNAbp to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some
15 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
20 is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nC9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ. Non-limiting
25 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/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1,
30 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, 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.

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.*, an adenine deaminase or an adenosine deaminase; or a cytidine deaminase or a cytosine deaminase). In some embodiments, the nucleobase editing domain is more than one deaminase domain (*e.g.*, an adenine deaminase or an adenosine deaminase and a cytidine or a cytosine 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.

As used herein, “obtaining” as in “obtaining an agent” includes synthesizing, purchasing, or otherwise acquiring the agent.

A “patient” or “subject” as used herein refers to a mammalian subject or individual diagnosed with, at risk of having or developing, 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, camels, llamas, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other mammals that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female.

“Patient in need thereof” or “subject in need thereof” is referred to herein as a patient diagnosed with, at risk or having, predetermined to have, or suspected of having a disease or disorder.

The terms “pathogenic mutation”, “pathogenic variant”, “disease causing 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.

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.

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-acetylaminoethyl-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.

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.

By "reduces" is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By "reference" is meant a standard or control condition. In one embodiment, the reference is a wild-type or healthy cell. In other embodiments and without limitation, a reference is an untreated cell that is not subjected to a test condition, or is subjected to

placebo or normal saline, medium, buffer, and/or a control vector that does not harbor a polynucleotide of interest.

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, at least about 20 amino acids, at least about 25 amino acids, 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, at least about 60 nucleotides, at least about 75 nucleotides, about 100 nucleotides or about 300 nucleotides or any integer thereabout or therebetween. In some embodiments, a reference sequence is a wild-type sequence of a protein of interest. In other embodiments, a reference sequence is a polynucleotide sequence encoding a wild-type protein.

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). In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (CsnI) from *Streptococcus pyogenes* (*See, e.g.*, "Complete genome sequence of an M1 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).

As used herein, the term "scFv" refers to a single chain Fv antibody in which the variable domains of the heavy chain and the light chain from an antibody have been joined to form one chain. scFv fragments contain a single polypeptide chain that includes the variable region of an antibody light chain (VL) (*e.g.*, CDR-L1, CDR-L2, and/or CDR-L3) and the variable region of an antibody heavy chain (VH) (*e.g.*, CDR-H1, CDR-H2, and/or CDR-H3) separated by a linker. The linker that joins the VL and VH regions of a scFv fragment can be

a peptide linker composed of proteinogenic amino acids. Alternative linkers can be used to so as to increase the resistance of the scFv fragment to proteolytic degradation (for example, linkers containing D-amino acids), in order to enhance the solubility of the scFv fragment (for example, hydrophilic linkers such as polyethylene glycol-containing linkers or

5 polypeptides containing repeating glycine and serine residues), to improve the biophysical stability of the molecule (for example, a linker containing cysteine residues that form intramolecular or intermolecular disulfide bonds), or to attenuate the immunogenicity of the scFv fragment (for example, linkers containing glycosylation sites). It will also be understood by one of ordinary skill in the art that the variable regions of the scFv molecules described

10 herein can be modified such that they vary in amino acid sequence from the antibody molecule from which they were derived. For example, nucleotide or amino acid substitutions leading to conservative substitutions or changes at amino acid residues can be made (*e.g.*, in CDR and/or framework residues) so as to preserve or enhance the ability of the scFv to bind to the antigen recognized by the corresponding antibody.

15 By “selectively binds” is meant specifically binds a wild-type version of the cell surface protein, but exhibits reduced binding or fails to bind to the cell surface protein comprising a mutation.

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

20 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 said to be alleles for this position. SNPs underlie differences in susceptibility to disease. The severity of illness and

25 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

30 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 can also be called a single-nucleotide alteration.

5 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.

10 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
15 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
20 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).

25 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
30 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.

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 an embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In another 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.

By “split” is meant divided into two or more fragments.

A “split Cas9 protein” or “split Cas9” refers to a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a “reconstituted” Cas9 protein. In particular embodiments, the Cas9 protein is divided into two fragments within a disordered region of the protein, *e.g.*, as described in Nishimasu *et al.*, Cell, Volume 156, Issue 5, pp. 935-949,

2014, or as described in Jiang *et al.* (2016) Science 351: 867-871. PDB file: 5F9R, each of which is incorporated herein by reference. In some embodiments, the protein is divided into two fragments at any C, T, A, or S within a region of SpCas9 between about amino acids A292-G364, F445-K483, or E565-T637, or at corresponding positions in any other Cas9, 5 Cas9 variant (*e.g.*, nCas9, dCas9), or other napDNAbp. In some embodiments, protein is divided into two fragments at SpCas9 T310, T313, A456, S469, or C574. In some embodiments, the process of dividing the protein into two fragments is referred to as “splitting” the protein.

In other embodiments, the N-terminal portion of the Cas9 protein comprises amino 10 acids 1-573 or 1-637 *S. pyogenes* Cas9 wild-type (SpCas9) (NCBI Reference Sequence: NC_002737.2, Uniprot Reference Sequence: Q99ZW2) and the C-terminal portion of the Cas9 protein comprises a portion of amino acids 574-1368 or 638-1368 of SpCas9 wild-type.

The C-terminal portion of the split Cas9 can be joined with the N-terminal portion of the split Cas9 to form a complete Cas9 protein. In some embodiments, the C-terminal portion 15 of the Cas9 protein starts from where the N-terminal portion of the Cas9 protein ends. As such, in some embodiments, the C-terminal portion of the split Cas9 comprises a portion of amino acids (551-651)-1368 of spCas9. “(551-651)-1368” means starting at an amino acid between amino acids 551-651 (inclusive) and ending at amino acid 1368. For example, the C-terminal portion of the split Cas9 may comprise a portion of any one of amino acid 551-1368, 20 552-1368, 553-1368, 554-1368, 555-1368, 556-1368, 557-1368, 558-1368, 559-1368, 560-1368, 561-1368, 562-1368, 563-1368, 564-1368, 565-1368, 566-1368, 567-1368, 568-1368, 569-1368, 570-1368, 571-1368, 572-1368, 573-1368, 574-1368, 575-1368, 576-1368, 577-1368, 578-1368, 579-1368, 580-1368, 581-1368, 582-1368, 583-1368, 584-1368, 585-1368, 586-1368, 587-1368, 588-1368, 589-1368, 590-1368, 591-1368, 592-1368, 593-1368, 594- 25 1368, 595-1368, 596-1368, 597-1368, 598-1368, 599-1368, 600-1368, 601-1368, 602-1368, 603-1368, 604-1368, 605-1368, 606-1368, 607-1368, 608-1368, 609-1368, 610-1368, 611-1368, 612-1368, 613-1368, 614-1368, 615-1368, 616-1368, 617-1368, 618-1368, 619-1368, 620-1368, 621-1368, 622-1368, 623-1368, 624-1368, 625-1368, 626-1368, 627-1368, 628-1368, 629-1368, 630-1368, 631-1368, 632-1368, 633-1368, 634-1368, 635-1368, 636-1368, 30 637-1368, 638-1368, 639-1368, 640-1368, 641-1368, 642-1368, 643-1368, 644-1368, 645-1368, 646-1368, 647-1368, 648-1368, 649-1368, 650-1368, or 651-1368 of spCas9. In some embodiments, the C-terminal portion of the split Cas9 protein comprises a portion of amino acids 574-1368 or 638-1368 of SpCas9.

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). In one embodiment, such a sequence is at least
 5 60%, 80%, 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid level to the sequence used for comparison.

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,
 10 BLAST, BESTFIT, 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
 15 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,
- 20 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:

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

The term "target site" refers to a sequence within a nucleic acid molecule that is deaminated by a deaminase (*e.g.*, cytidine or adenine deaminase) or a fusion protein

comprising a deaminase (*e.g.*, a dCas9-adenosine deaminase fusion protein or a base editor disclosed herein).

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 al, Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013); Mali, P. et al, RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013); Hwang, W.Y. et al, Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013); Jinek, M. et al, RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013); Dicarlo, J.E. et al, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. et al 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).

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.

By “uracil glycosylase inhibitor” or “UGI” 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. In an embodiment, a UGI is a protein, a fragment thereof, or a domain 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 modified version thereof. In some embodiments, a UGI domain comprises a fragment of the exemplary amino acid sequence set forth below. In some embodiments, a UGI fragment comprises an amino acid sequence that

comprises 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 100% of the exemplary UGI sequence provided below. In some embodiments, a UGI comprises an amino acid sequence that is homologous to the exemplary UGI amino acid sequence or fragment thereof, as set forth below. In some embodiments, the UGI, or a portion thereof, is 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%, at least 99.5%, at least 99.9%, or 100% identical to a wild type UGI or a UGI sequence, or portion thereof, as set forth below. An exemplary UGI comprises an amino acid sequence as follows:

10 >splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor
 MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
 APEYKPWALVIQDSNGENKIKML.

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.

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.

All terms are intended to be understood as they would be understood by a person skilled in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains.

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.

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic diagram depicting that anti-CD117 antibody AMG 191 (Humanized SR1) can be used for conditioning in patients prior to hematopoietic stem cell transplantation.

FIG. 2 is a schematic diagram depicting that the anti-CD117 antibody SR1 (AMG191) blocks Stem Cell Factor (SCF) binding to CD117 and depletes human
10 hematopoietic stem progenitor cells (HSPCs) and Myelodysplastic syndromes (MDS) cells in vivo.

FIG. 3 is a schematic diagram depicting a strategy for non-toxic conditioning to enable hematopoietic stem cell transplantation (HSCT) with autologous gene edited cells for
15 patients with hematologic diseases. Stem cell factor (SCF) drives HSC self-renewal and differentiation into progenitor cells. Administering anti-CD117 antibody blocks SCF binding to CD117, thereby depleting HSCs and progenitor cells in the patient (conditioning). Autologous gene edited HSCs are transplanted into the patient. Base editing is used to generate an amino acid substitution in CD117, which prevents the binding of anti-CD117
20 antibodies to the edited cell, but does not interfere with normal SCF signaling. Gene-edited cells compete with residual host HSCs to repopulate bone marrow (BM). Anti-CD117 antibody blocks SCF binding to wildtype (WT) CD117, but cannot bind to HSCs with an edited CD117. Thus, native, wild-type HSCs are targeted by anti-CD117 antibody, but gene edited HSCs are not.

25 FIG. 4 is a schematic diagram depicting SCF binding to wild-type and gene-edited HSCs. Both cells express CD117 that are activated by SCF binding.

FIG. 5 is a schematic diagram depicting the effect of anti-CD117 antibody on wild-type and gene-edited HSCs. Binding of anti-CD117 antibody to wild-type CD117 disrupts SCF binding and results in inhibition of SCF signaling wild-type cells. In contrast, gene-
30 edited HSCs are refractory to anti-CD117 antibody because the amino acid substitution introduced in CD117 prevents the binding of anti-CD117 antibody, but does not interfere with normal SCF binding and signaling.

FIG. 6 are tables depicting data from sequence reads of base edited cells at target sites for Makassar-WT guide (top) and Hereditary Persistence of Fetal Hemoglobin (HPFH) guide (bottom).

FIGS. 7A-7B are graphs depicting the on-target editing (FIG. 7A) and the indel rate (FIG. 7B) of five (5) guides (ABE8.8). HPFH guide was used as a control.

FIGS. 8A-8B are graphs depicting the on-target editing (FIG. 8A) and the indel rate (FIG. 8B) of eleven (11) guides (ABE8.8). HPFH guide was used as a control.

FIGS. 9A-9B are graphs depicting the on-target editing (FIG. 9A) and the indel rate (FIG. 9B) of nine (9) guides (ABE8.8). HPFH guide was used as a control.

FIG. 10 is a graph depicting the on-target editing of four (4) guides for ABE8.8 and thirteen (13) guides for IBE-NGC (Makassar).

FIGS. 11A-11T are schematic drawings depicting the location of guide targeted mutations for anti-CD117 antibody campaign.

FIGS. 12A-12N depict guides that introduce naturally occurring mutations in a CD117 target sequence. FIG. 12A is a schematic drawing depicting a CD117 target nucleotide sequence and corresponding amino acid sequence for the hybridization of guides cc-102 and cc-103. FIGS. 12B and 12C are tables depicting the editing efficiencies at different base pair positions along a CD117 target site for either guide cc-102 (FIG. 12B) or guide cc-103 (FIG. 12C). FIG. 12C is a schematic drawing depicting a CD117 target nucleotide sequence and corresponding amino acid sequence for the hybridization of guide cc-78. FIG. 12E is a table depicting the editing efficiency at different base pair positions along a CD117 target site for guide cc-78. FIG. 12F is a schematic drawing depicting a CD117 target sequence and corresponding amino acid sequence for the hybridization of guides cc-87 and cc-89. FIGS. 12G and 12H are tables depicting the editing efficiencies at different base pair positions along a CD117 target site for either guide cc-87 (FIG. 12G) or guide cc-89 (FIG. 12H). FIG. 12I is a schematic drawing depicting a CD117 target sequence and corresponding amino acid sequence for the hybridization of guide cc-110. FIG. 12J is a table depicting the editing efficiency at different base pair positions along a CD117 target site for guide cc-110. FIG. 12K is a schematic drawing depicting a CD117 target sequence and corresponding amino acid sequence for the hybridization of guide cc-146. FIG. 12L is a table depicting the editing efficiency at different base pair positions along a CD117 target site for guide cc-146. FIG. 12M is a schematic drawing depicting a CD117 target sequence and corresponding amino acid sequence for the hybridization of guide cc-182. FIG. 12N is a

table depicting the editing efficiency at different base pair positions along a CD117 target site for guide cc-182.

FIGS. 13A and 13B are graphs depicting multiplex editing. FIG. 13A is a graph depicting c-KIT editing using dual guides for c-KIT and HPFH (day 3), dual guides for c-KIT and HPFH (day 5), or just the c-KIT guide and no HPFH guide. FIG. 13B is a graph depicting HPFH editing using dual guides for c-KIT and HPFH at either 72 or 120 hours. HPFH only was used as a control.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides a base editing strategy targeting cell surface proteins (*e.g.*, one or more of CD117, CXCR4, CD135, CD90, CD45, CD34) that is useful for non-toxic conditioning.

The invention is based, at least in part, on the discovery that base editing can be used to generate a single base substitution in the CD117 gene resulting in an amino acid substitution, that alters an epitope binding domain in the encoded protein, thereby preventing anti-CD117 antibodies, antibody drug conjugates (ADCs), or chimeric antigen receptor (CAR)-T cells from binding to CD117 on edited cells. Advantageously, this alteration in CD117 does not alter Stem Cell Factor (SCF) binding or CD117 biological activity. Thus, the invention provides for the post-HSCT use of anti-CD117 therapies, as well as therapies targeting other antigens expressed on the surface of hematopoietic stem cells. which do not bind to or deplete CD117-edited cells. Accordingly, gene edited cells may be expanded in vivo. Advantageously, this allows for cancer treatment without hematopoietic toxicity.

Prior art conditioning methods were limited to administering a conditioning regimen prior to HSC transplantation (*i.e.*, cannot re-dose patient to expand edited cells in vivo or continue to treat malignant disease). To avoid affecting donor/edited cells, the half-life of the antibody needed to be short, such that it had been cleared from the body prior to transplantation. This ensured that the antibody did not target transplanted cells. There is currently no method/therapy that allows the selective targeting of patient HSCs. This applies to all current conditioning protocols (*e.g.*, Abs, ADCs, CAR-T cell). Thus, clinical trial design is extremely challenging for conditioning prior to HSCT and consideration must be given to number/concentration of doses vs HSC depletion vs hematopoietic failure. However, research suggests that efficient antibody-based conditioning may only be achieved through multiple doses in immunocompetent patients (unlike SCID trials).

Thus, the present methods for conditioning were developed to overcome the limitations of previous methods. There are several advantages of the conditioning methods of the invention. The methods described herein provide for the selective targeting of endogenous HSCs, while sparing edited HSCs. Accordingly, antibody or ADC treatment can
5 continue to be administered following HSCT to expand gene edited cells in vivo or treat malignant disease with repeated dosing. This minimizes the risk of killing edited cells. Edited cells would allow for the administration of antibody or ADC without Fc modifications to reduce their half-life. This has the potential to enable the use of antibodies with longer half-lives, such as AMG191, and simplify the development of ADCs. Clinical trial design is
10 also simplified – HSCs could be infused prior to or concurrently with conditioning with little or no risk of being depleted (*e.g.*, in MDS or AML patients). The methods provide a benefit for all patients regardless of immune status.

In one embodiment, CD117 is altered (*e.g.*, using base editing) in a cell for transplantation to prevent binding of anti-CD117 antibody, but not interfere with normal SCF
15 signaling. Using base editing, a nucleobase change may be generated to create an amino acid substitution in CD117. Stem cell factor (SCF) drives HSC self-renewal and differentiation into progenitor cells. Administering anti-CD117 antibody blocks SCF binding to CD117, thereby depleting HSCs and progenitor cells in the patient (conditioning). Autologous gene edited HSCs are transplanted into the patient. Gene-edited cells compete with residual host
20 HSCs to repopulate bone marrow (BM). Anti-CD117 antibody blocks SCF binding to wildtype (WT) CD117, but cannot bind to HSCs with an edited CD117. Thus, native, wild-type HSCs are targeted by anti-CD117 antibody, but gene edited HSCs are not. Thus, both cells express CD117 polypeptides that are activated by SCF binding. However, binding of anti-CD117 antibody to wild-type CD117 disrupts SCF binding and results in inhibition of
25 SCF signaling wild-type cells. In contrast, gene-edited HSCs are refractory to anti-CD117 antibody because the amino acid substitution introduced in CD117 prevents the binding of anti-CD117 antibody, but does not interfere with normal SCF binding and signaling.

In various embodiments, the antibodies of the invention are currently used in the clinic (*e.g.* if the epitope edited is required for the binding of AMG 191). In one
30 embodiment, CD117 is altered to resemble murine CD117, which does not cross react with AMG 191 or SR1. Because murine SCF activates murine and human cells, species-to-species variation in the epitope provides potential variants that have SCF activity while having reduced or no anti-CD117 binding. For the development of new antibodies, ADCs, CAR-T cellss etc. that could target a novel epitope in one of the extracellular domains of

CD117 (D1-D5). Thus, methods of identifying a candidate agent for selectively depleting or ablating an endogenous stem cell population are also within the scope of the invention. Such methods may comprise the steps of: (a) contacting a sample comprising the stem cell population with a test agent (*e.g.*, antibody); and (b) detecting whether one or more cells of the stem cell population are depleted or ablated from the sample; wherein the depletion or ablation of one or more cells of the stem cell population following the contacting step identifies the test agent as a candidate agent. In a further step an edited cell is identified as not similarly depleted or ablated by the agent. In some embodiments, the cell is contacted with the test agent for at least about 2-24 hours.

Exemplary, yet nonlimiting, examples of compositions and methods for treating hemoglobinopathies are described in International Publication No. WO2020168133, which is incorporated herein by reference for its entirety.

Methods of Treatment

The methods and compositions disclosed herein may be used to condition a subject's tissues (*e.g.*, bone marrow) for engraftment or transplant and following such conditioning, a stem cell population is administered to the subject. The transplanted cells (*e.g.*, HSCs) can be autologous cells or allogeneic cells. In certain aspects, the stem cell population comprises an exogenous stem cell population. In some embodiments, the stem cell population comprises the subject's endogenous stem cells (*e.g.*, endogenous stem cells that have been genetically modified to correct a disease or genetic defect). Preferably, such methods and compositions are useful for treating such diseases without causing the toxicities that are observed in response to traditional conditioning therapies, such as irradiation.

Hematopoietic stem cell transplant therapy can be administered to a subject in need of treatment so as to populate or re-populate one or more blood cell types. Hematopoietic stem cells generally exhibit multi-potency, and can thus differentiate into multiple different blood lineages including, but not limited to, granulocytes (*e.g.*, promyelocytes, neutrophils, eosinophils, basophils), erythrocytes (*e.g.*, reticulocytes, erythrocytes), thrombocytes (*e.g.*, megakaryoblasts, platelet producing megakaryocytes, platelets), monocytes (*e.g.*, monocytes, macrophages), dendritic cells, microglia, osteoclasts, and lymphocytes (*e.g.*, NK cells, B-cells and T-cells). Hematopoietic stem cells are additionally capable of self-renewal, and can thus give rise to daughter cells that have equivalent potential as the mother cell, and also feature the capacity to be reintroduced into a transplant recipient whereupon they home to the hematopoietic stem cell niche and re-establish productive and sustained hematopoiesis.

Hematopoietic stem cells can thus be administered to a patient defective or deficient in one or more cell types of the hematopoietic lineage in order to reconstitute the defective or deficient population of cells *in vivo*, thereby treating the pathology associated with the defect or depletion in the endogenous blood cell population. The compositions and methods described herein can thus be used to treat a non-malignant hemoglobinopathy (*e.g.*, a hemoglobinopathy selected from the group consisting of sickle cell anemia, thalassemia, Fanconi anemia, aplastic anemia, and Wiskott-Aldrich syndrome). Additionally or alternatively, the compositions and methods described herein can be used to treat a malignancy or proliferative disorder, such as a hematologic cancer, myeloproliferative disease. In the case of cancer treatment, the compositions and methods described herein may be administered to a patient so as to deplete a population of endogenous hematopoietic stem cells prior to hematopoietic stem cell transplantation therapy, in which case the transplanted cells can home to a niche created by the endogenous cell depletion step and establish productive hematopoiesis. This, in turn, can re-constitute a population of cells depleted during cancer cell eradication, such as during systemic chemotherapy. Exemplary hematological cancers that can be treated using the compositions and methods described herein include, without limitation, acute myeloid leukemia, acute lymphoid leukemia, chronic myeloid leukemia, chronic lymphoid leukemia, multiple myeloma, diffuse large B-cell lymphoma, and non-Hodgkin's lymphoma, as well as other cancerous conditions, including neuroblastoma.

Antibodies, antigen-binding fragments thereof, and ligands described herein can be administered to a patient (*e.g.*, a human patient suffering from cancer, an autoimmune disease, or in need of hematopoietic stem cell transplant therapy) in a variety of dosage forms. For instance, antibodies, antigen-binding fragments thereof, and ligands described herein can be administered to a patient suffering from cancer, an autoimmune disease, or in need of hematopoietic stem cell transplant therapy in the form of an aqueous solution, such as an aqueous solution containing one or more pharmaceutically acceptable excipients. Pharmaceutically acceptable excipients for use with the compositions and methods described herein include viscosity-modifying agents. The aqueous solution may be sterilized using techniques known in the art.

The antibodies, antigen-binding fragments, and ligands described herein may be administered by a variety of routes, such as orally, transdermally, subcutaneously, intranasally, intravenously, intramuscularly, intraocularly, or parenterally. The most suitable route for administration in any given case will depend on the particular antibody, antigen-

binding fragment, or ligand administered, the patient, pharmaceutical formulation methods, administration methods (*e.g.*, administration time and administration route), the patient's age, body weight, sex, severity of the diseases being treated, the patient's diet, and the patient's excretion rate.

5

NUCLEOBASE EDITOR

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 (*e.g.*, adenosine deaminase, cytidine deaminase). 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

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.

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
 5 nucleotide binding domain can be a ribonuclease.

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
 10 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
 15 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
 20 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
 25 nickase domain can comprise a deletion of all or a portion of the RuvC domain or the HNH domain.

The amino acid sequence of an exemplary catalytically active Cas9 is as follows:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 30 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSGL
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK

DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIILDF
 5 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 10 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRMLASAGELQKGNELALPSKYVNFYLLAS
 HYEKLKGSPEQNEQQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 15 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD.

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
 20 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-
 25 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.

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
 30 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.

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). 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).

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.

5 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-
 10 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
 15 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
 20 repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self.

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
 25 (or polynucleotide, e.g., DNA or RNA) target to be modified. Thus, a skilled artisan can change the genomic or polynucleotide target of the Cas protein by changing the target sequence present in the gRNA. The specificity of the Cas protein is partially determined by how specific the gRNA targeting sequence is for the genomic polynucleotide target sequence compared to the rest of the genome.

30 In some embodiments, the gRNA scaffold sequence is as follows: GUUUUAGAGC
 UAGAAUAGC AAGUUAAAUA AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU
 GGCACCGAGU CGGUGCUUUU.

In an embodiment, the RNA scaffold comprises a stem loop. In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUUGUACUCUCAAGAUUUUAGUAACUGUACAACGAAACUUACACAGUUACUAAAUCUU
GCAGAAGCUACAAAGAUUAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUUAUGGCAGGGU

5 G. In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU.

In an embodiment, an *S. pyrogenes* sgRNA scaffold polynucleotide sequence is as follows:

10 GUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGC.

In an embodiment, an *S. aureus* sgRNA scaffold polynucleotide sequence is as follows:

GUUUUAGUACUCUGUAAUGAAAAUACAGAAUCUACUAAAACAAGGCAAAUAGCCGUGUUUA
15 UCUCGUCAACUUGUUGGCGAGA.

In an embodiment, a BhCas12b sgRNA scaffold has the following polynucleotide sequence:

GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCU
AUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC.

20 In an embodiment, a BvCas12b sgRNA scaffold has the following polynucleotide sequence:

GACCUAUAGGGUCAUGAAUCUGUGCGUGGCCAUUAGUAAUUAAAAUUACCCACCACAGG
AGCACCUGAAAAACAGGUGCUUGGCAC.

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.

25

30

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, Cas12i, and Cas12j/CasΦ, 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.

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.

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 syphilidicola* (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*.

5 **Cas9 domains of Nucleobase Editors**

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 *et al.*, 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 C.M., 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 will 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 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.

10 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. 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, 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.

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, Cas12c/C2C3, and Cas12j/CasΦ.

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).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAAAAATTTGGCAGATTCT
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTAGAGTAGA
TGCTAAAGCGATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG
ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTTGGCGCAAATTTGGAGATCAATATGCTGATTTGT
TTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTACAGATATCCTAAGAGTAAATAGT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT

TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
5 GGCGCGTGGCAATAGTCGTTTTGTCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAAATCTTCCAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAA
10 GTAACCGTTAAGCAATTAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
15 TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT
GATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA
20 AAATCAGACAACCTAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG
GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA
AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT
AGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAG
ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAAC
25 GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTCTAAACGCCAA
GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG
GCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA
GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
30 ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTGTT
GGAAGTCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAAATAGGCAAAGCAACCGCAA
AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAACAGAAATTACACTTGCAAATGGA
GAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA

AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
 AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
 AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAAC
 GGTAGCTTATTGAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAAT
 5 CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCGATT
 GACTTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
 ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
 AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
 TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
 10 TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTATG
 CAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGT
 GAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
 TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
 CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
 15 GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLADSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVKLFI
 20 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDYYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYVGPLARGNSRFWMTRKSEETITPWNFEVVVDKGASQSFIERMT
 25 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLKIIKDKDFLDNEENEDILEDIV
 LTTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMILHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLO
 30 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKROLVETROITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD

KLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENPI
 DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIR
 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQL
 5 GGD

(single underline: HNH domain; double underline: RuvC domain)

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

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCGGTTGGATGGGCTGTCAT
 10 AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGAACACAGACCGTCATT
 CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACT
 CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
 AGAAATTTTGTAGCAATGAGATGGCCAAAGTTGACGATTCTTCTTTCACCGTTTGGAAGAGT
 CCTTCCTTGTGGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
 15 GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
 AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
 GGCACCTTCTCATTGAGGGTGATCTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATC
 CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA
 TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC
 20 AATTACCCGGAGAGAAGAAAAATGGGTTGTTCCGTAACCTTATAGCGCTCTCACTAGGCCTG
 ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA
 CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT
 TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT
 GAGATTACCAAGGCGCGTTATCCGCTCAATGATCAAAGGTACGATGAACATACCAAGA
 25 CTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATATTCT
 TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC
 TACAAGTTTATCAAACCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAACCT
 CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA
 TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTATCCGTTCTCAAA
 30 GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT
 GGCCCGAGGGAACCTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT
 GGAATTTTGTAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC
 AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA
 TTTCACAGTGTACAATGAACCTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG

CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTGCA
GATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA
TAATTAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
5 TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA
CCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACCTATTCTCGATTTT
CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA
10 TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG
GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACG
CGAAAATCAAACGACTCAGAAGGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG
AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
15 ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA
AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGGAAAAGTGAC
AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC
GAACTGATAACGCAAAGAAAGTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAACCCGCCAAATCACAAAGCAT
20 GTTGACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCTG
GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTGCGACTTCAGAAAGGATTTTCAAT
TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
25 CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA
TAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAA
AGAAAACCTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCAAAAAGGAATAGT
GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCC
30 TACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAACTGA
AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAACTACC
AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC
TTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTCCTGTATTTAGCGTCC

CATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCA
 GCACAAACATTATCTCGACGAAATCATAGAGCAAATTTGGAATTCAGTAAGAGAGTCATCC
 TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCATA
 CGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
 5 ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
 ACGCGACACTGATTACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTACAG
 CTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGA
 CGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

 10 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 15 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIHQIHLGELHAILRRQEDFYFPFLK
DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIV
 20 LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFQM LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETROITKH
 25 VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNP
IDFLEAKGYKEVKKDIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 30 HYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

(single underline: HNH domain; double underline: RuvC domain).

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):

```

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
5 GTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAAAAATTGGTAGATTC
10 TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGTATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAACTATTTATC
CAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCTATTAACGCAAGTGGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTG
15 ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGA
TACTTACGATGATGATTTAGATAATTTATTTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTCTCAGATATCCTAAGAGTAAATACT
GAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTT
20 TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCATCAAA
TTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTTCGAATTCCTTATTATGTTGGTCCATT
25 GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAAATCTTCAAATGAAAAAGTACTACCAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAA
30 GTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT

```

TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGACATTCAAAAAGCACAAAGTGTCTGGACAAGGCGATAGTTTACATGAACATA
TTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTT
5 GATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACG
TGAAATCAGACAACCTCAAAAGGGCCAGAAAAATTTCGCGAGAGCGTATGAAACGAATCGAAG
AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTG
CAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGA
ATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTA
10 AAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGAT
AACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAACTATTGGAGACAACCTTCTAAACGC
CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT
GTGGCACAAATTTTGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCTG
15 AGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAAT
TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTC
GTTGGAAGCTGCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTA
TAAAGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCG
CAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAAT
20 GGAGAGATTTCGAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGA
TAAAGGGCGAGATTTTGCCACAGTGCACAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA
AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAGAAAAGAAATTCG
GACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCC
AACGGTAGCTTATTGAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGAAATCGAAGAAGTTAA
25 AATCCGTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAATCCG
ATTGACTTTTTAGAAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAACTACC
TAAATATAGTCTTTTGGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT
TACAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT
CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCA
30 GCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTT
TAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATA
CGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG

ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG
CTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
5 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
10 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
DNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDL LKI IKDKDFLDNEENEDILEDIV
LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
15 LKSDGFANRNFQM LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
DELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
20 VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLLAS
HYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
25 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQ
LGGD (single underline: HNH domain; double underline: RuvC domain)

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 syrophidicola* (NCBI Ref: NC_021284.1);
30 *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) or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

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.

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).

The amino acid sequence of an exemplary catalytically inactive Cas9 (dCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKS NFDAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFY PFLK
DNREKIEKILTFRIPIYVGPLARGNSRFWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLIKDKDFLDNEENEDILEDIV
LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL

QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIASEQEIGKATAKYFFYSNIMNFFKTEITLAN
 5 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEFKNP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEQKQKQFLVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 10 LGGD

(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).

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA
 15 cleavage domain, that is, the Cas9 is a nickase, referred to as an “nCas9” protein (for
 “nickase” Cas9). 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
 20 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
 25 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, the dCas9 domain comprises an amino acid sequence that is at
 30 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.

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):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 15 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 20 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGGASQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIV
 LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 25 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKROLVETROITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 30 VGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTETVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPNP
 IDFLEAGKYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLAS
 HYEKLKGSPEQKQFLVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPI

REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD (single underline: HNH domain; double underline: RuvC domain).

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.

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.

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.

The amino acid sequence of an exemplary catalytically Cas9 nickase (nCas9) is as follows:

5 MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
QLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
10 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
DNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKGASAQSFIERMT
NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIV
15 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
20 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNP
IDFLEAKGYKEVKKDIIKLPKYSLFELENGRKRMLASAGELQGNELALPSKYVNFYLYLAS
25 HYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
LGGD

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
30 embodiments, the programmable nucleotide binding protein may be a CasX or CasY protein, 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.

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 ease 99.5% identical to a naturally-occurring CasX or CasY protein. In some embodiments, the programmable nucleotide binding protein is a naturally-occurring CasX or CasY protein. In some embodiments, the programmable nucleotide binding protein 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.

An exemplary CasX ((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIHCRISPR-associatedCasx protein OS = *Sulfolobus islandicus* (strain HVE10/4) GN = SiH_0402 PE=4 SV=1) amino acid sequence is as follows:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLI IAAAGWVLTRLGKAKVSEG DYVGVNVFTP
TRGILYSLIQNVNGIVPGIKPETAFLGIARKVSVTNPNVSVVRIYITISDAVGQNPTTIN
GGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANIYIYEYLTG SKRLEDLLY
FANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

An exemplary CasX (>tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = *Sulfolobus islandicus* (strain REY15A) GN=SiRe_0771 PE=4 SV=1) amino acid sequence is as follows:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
 AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQVKECEEVSAP
 SFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLMIAAAGWVLTRLGKAKVSEG DYVGVNVFT P
 TRGILYSLIQNVNGIVPGIKPETAFLWIARKVSVSVTNPNVSVVSIYITISDAVGQNPTTIN
 5 GGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANIYIYEYLTGSKRLEDLLYF
 ANRDLIMNLSDDGKVRDLKLISAYVNGELIRGEG.

Deltaproteobacteria CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLVVRVMTDDLKKRLEKRRKKPEVMPQVISNNA
 NNLRLMLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLPKPEMDEKGNLTTA
 10 GFACSQCGQPLFVYKLEQVSEK GKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLG
 KFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIII
 EHQKVVGKNQKRLESRELAGKENLEYPSVTLPPQPHTKEGVDFAYNEVIARVRMWNVLNLW
 QKLKLSRDDAKPLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAE
 KRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERI
 15 DKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRKASFVLERLKEMDEKEFYACEIQLOKWY
 GDLRGNPF AVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIR
 FTDGTDIKKS GKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLET
 GLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVI
 ALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRK FASKSRNL
 20 ADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGL
 TSKTYLSKTLAQYTSKTC SNCGFTITYADMVMLVRLKKTSDGWATTLNNKELKAEYQITYY
 NRYKRQTVEKELSAELDR LSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCG
 HEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPN

An exemplary CasY ((ncbi.nlm.nih.gov/protein/APG80656.1) >APG80656.1

25 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]) amino acid
 sequence is as follows:

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGL
 YGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTA PGLLNVAEVRGGSYELTKTL
 KGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCENKLADDIKN
 30 AKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFN
 KLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLENKITEKKAMMDITDAW
 RGQEQEEELEKRLRILAALTIKLREP KFDNHGGYRSDINGKLSSWLQNYINQTVKIKEDLK
 GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIYRRFLSD
 GRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKSDAEDEKETIDFKELFPHLAKPLKL

VPNFYGD SKRELYKKYKNAAIYTDALWKA VEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQK
 IFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSA AIDKNRVRLPSTEN
 IAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLAVTETQLDISALDFVE
 NGTVKDFMKT RDGNLVLEGRFLEMFSQSIVFSELRLAGLMSRKEFITRS AIQTMNGKQAE L
 5 LYIPHEFQSAKITTPKEMSR AFLDLAPAEFATSLEPESLSEKSL LKLKQMRYYPHYFGYELT
 RTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHR
 PKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQ
 RYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGT FAMPSTK
 IARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDAD
 10 KNLQTTVWGKLA VASEISASYTSQFCGACKKLWRAEMQVDETIT TQELIGTVRVIKGGT LID
 AIKDFMRPPIFDENDT PFPKYRDFCDKHHISKMRGNSCLFICPFCRANADADIQASQTIAL
 LRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI .

In some embodiments, the nucleic acid programmable DNA binding protein
 (napDNAbp) is a single effector of a microbial CRISPR-Cas system. Single effectors of
 15 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
 20 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
 25 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.

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
 30 *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.

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.

A Cas12b/C2c1 ((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-associated endonuclease C2c1 OS = *Alicyclobacillus acido-terrestris* (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) amino acid sequence is as follows:

MAVKS I KVKLR LDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECD
KTAECKAELLERLRARQVENGHARGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKF
LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEKEKAETRKSADRTADVLRALADFG
LKPLMRVYTDSEMSVVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQ
KNRFEQKNFVGQEHVLVNLQQLQDMKEASPGLESKEQTAHYVTGRALRGSDKVFKEKGKLA
PDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGLHGYTFLFNEFGERRHAIRFHKLLKVENGVADEV
DDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRG
ARDVYLVNSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSE

GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNKGRVPFFFPIKGNDNLVAVHERSQLL
 KLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAAN
 HMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPK
 IRGYAKDVVGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRAITLREHIDHAKE
 5 DRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLM
 QWSHRGVFQELINQAQVHDLVGTMYAAFSSRFDARTGAPGIRCRVPARCTQEHNPFPFPW
 WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQORLWSDF
 DISQIRLRCDWGEVDGELVLIPLRTGKRTADSYSNKVFTNTGVTYERERGGKRRKVFAQE
 KLSEEEAELLVEADEAREKSVVLMRDPGSIINRGNWTRQKEFWSMV NQRIEGLVKQIRSR
 10 VPLQDSACENTGDI .

BhCas12b (*Bacillus hisashii*) NCBI Reference Sequence: WP_095142515

MAPKKKKRVGIHGVPAATRSLFKIEPNEEVKKGLWKTHEVLNHGIAAYMNLKLIHQEAI
 YEHHEQDPKNPKKVSKEAIQAELWDFVLKMQKCNSFTEHVDKDEVFNILRELYEELVPSSVE
 KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKKWEEDKKKDP
 15 LAKILGKLAIEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
 WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYKERQEQLLRDTLNTNEYRLSKRGLR
 GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKL
 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
 20 LGGARVQFDRDHLRRYPHKVESGNVGRIFYNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
 KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLQGRQAAAASIFEVVDQKPDIEGKLFFPIK
 GTELYAVHRASFNIKLPGETLVKSREVLKAREDNLKL MNQKLNFLRNVLHFQQFEDITERE
 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFKQLHKRLEVEIGKEVKHWRKS
 LSDGRKGGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFADQLNHLNALKED
 25 RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYEERSRFENSKLMKWSR
 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFKKNLQREGR
 LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK
 AYQVDGQTVYIPESKDQKQKIIIEFGEGYFILKDGVEWVNAGKLKIKKGSSKQSSSELVDS
 DILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQYSISTIE
 30 DDSSKQSMKRPAATKKAGQAKKKK

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

MAIRS IKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYMNLTLTYRQEAIGDKTKEAYQAE
 LINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNS
 QSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPFLPLFT
 NIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLT
 5 GGEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWK
 VVAEQQNKMSSEFGDPKVSFLANRENNDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTL
 PDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIWPSEEKWIENIEIPLAPSI
 QFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFF
 NLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLL
 10 EGMRVMSIDMGQRTSASVSI FEVVKELPKDQEQLFYSINDTELFAIHKRSFLLNLPGEVVT
 KNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKT PDERKKA I HKLMEIVQSYDSWTASQKE
 VWEKELNLLTNMAAFNDEIWKESLVELHHRIEYPYVGQIVSKWRKGLSEGRKNLAGISMWNID
 ELEDTRLLIISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLLKQMANLIIMTALGFK
 YDKEEKDRYKRWKETY PACQIILFENLNRYLFNLDRSRRENSRLMKWAHRSI PRTVSMQGEM
 15 FGLQVGDVRSSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKG
 DIIPSQGGELFVTLSKRYKKDSNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMG
 EDKLYIPKSQTETIKKYFGKGSFVKNNTQEVEYKWEKSEKMKIKTDTT FDLQDLGDFEDISK
 TIELAQEQQKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSCCLKKKILSNKVEL

In some embodiments, the nucleic acid programmable DNA binding protein
 20 (napDNAbp) of any of the fusion proteins provided herein may be a Cas12j/CasΦ protein.
 Cas12j/CasΦ is described in Pausch *et al.*, “CRISPR-CasΦ from huge phages is a
 hypercompact genome editor,” *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337,
 which is incorporated herein by reference in its entirety. In some embodiments, the
 napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%,
 25 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 Cas12j/CasΦ protein. In
 some embodiments, the napDNAbp is a naturally-occurring Cas12j/CasΦ protein. In some
 embodiments, the napDNAbp is a nuclease inactive (“dead”) Cas12j/CasΦ protein. It should
 be appreciated that Cas12j/CasΦ from other species may also be used in accordance with the
 30 present disclosure.

Exemplary Cas12j/CasΦ amino acid sequences follow:

>CasΦ-1

MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFLRGKSEESPPDFQPPVKC
 PIIACSRPLTEWPIYQASVAIQGYVYGQSLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSVQ

GLNLI FQNARKRYIGVQTKVTNRNEKRHKLLKRINAKRIAEGLPELTSDEPESALDETGHLLI
 DPPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAGYSTSGDDPIQPMVTKDRLSISKGQPGYI
 PEHQRALLSQQKHHRRMRGYGLKARALLVIVRIQDDWAVIDLRSLLRNAYWRRIVQTKEPSTI
 TKLLKLVTGDPVLDATRMVATFTYKPGIVQVRSKCLKNKQGSKLFSERYLNETVSVTSIDL
 5 GSNNLVAVATYRLVNGNTPPELLQRFTLPSHLVKDFERYKQAHDTLEDSIQKTAVASLPQGQQ
 TEIRMWSMYGFREAQERVQCQELGLADGSI PWNVMTATSTILTDLFLARGGDPKKCMFTSEPK
 KKKNSKQVLYKIRDRAWAKMYRTLSSKETREAWNKA LWGLKRGSPDYARLSKRKEELARRCV
 NYTISTA EKRAQCGR TIVALEDLNIGFFHGRGKQEPGWVGLFTRKKENRWLMQALHKAFLLEL
 AHRGYHVIEVNPAYTSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVATHNIAMV
 10 AITGESLKRARGSVASKTPQPLAAE*

>CasΦ-2

MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQ GKSEEEPPNFQPPA
 KCHVVTKS R D FAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHA AWFAATGVSNHGYS
 HVQGLNLI F D H T L G R Y D G V L K K V Q L R N E K A R A R L E S I N A S R A D E G L P E I K A E E E E V A T N E T G
 15 HLLQPPGINPSFYVYQTISPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGVVRNRCDIQKGCP
 GYIPEWQREAGTAISPKTGKAVTVPLSPKKNKMRRYWRSEKEKAQDALLVTVRIGTDWVV
 IDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDACGTYARKWTLKG
 KQTKATLDKLTATQTVALVAIDLGTNPISAGISRVTQENGALQCEPLDRFTLPDDLKDIS
 AYRIAWDRNEEELRARSVEALPEAQQA EVRALDGVSKETARTQLCADFG LDPKRLPWDKMSS
 20 NTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVEVMRKDRTWARAYKPRLSVEAQKLKNE
 ALWALKRTSPEYLLKLSRRKEELCRRSINYVIEKTRRRTQCQIVIPVIEDLNVRFFHGS GKRL
 PGWDNFFTAKKENRWFIQGLHKA FSDLRTHRSFYVFEVRPERTSITCPKCGHCEVGNRDGEA
 FQCLSCGKTCNADLDVATHNLTQVALTGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAER
 EDQTPAQEPSQTS

25 >CasΦ-3

MEKEITELTKIRREFPNKKFSSTDMKKAGKLLKAEGPDAVRDFLNSCQEIIGDFKPPVKTN I
 VSISRPFE EW P V S M V G R A I Q E Y Y F S L T K E E L E S V H P G T S S E D H K S F F N I T G L S N Y N Y T S V Q G
 LNLIFKNAKAIYDGTLVKANNKNKKLEKKFNEINH KRSLEGLPIITPD FE E P F D E N G H L N N P
 PGINRNIYGYQGCAAKVFVPSKHKMVSLPKEYEGYNRDPNLSLAGFRNRLEIPEGE PGHVPW
 30 FQRM DIPEGQIGHVNKIQRNFVHGKNSGKVKFS DKTGRVKRYHHSKYKD ATKPYKFLEESK
 KVSALDSILAIITIGDDWVVFDIRGLYRNVFYRELAQKGLTAVQLLDLFTGDPVIDPKKGVV
 TFSYKEGVVPVFSQKIVPRFKSRDTLEKLT SQGPVALLSVDLGQNEPVAARVCSLKNINDKI
 TLDNSCRISFLDDYKKQIKDYRDSLDELEIKIRLEAINSLETNQQVEIRDLDVFSADRAKAN
 TVDMFDIDPNLISWDSMSDARVSTQISDLYLKNGGDES RVYFEINNKR I K R S D Y N I S Q L V R P

KLSDSTRKNLNDISIWKLKRTSEEYLKLSKRKLELSRAVVNYTIRQSKLLSGINDIVIIILEDL
DVKKKFNGRGIRDIGWDNFFSSRKENRWFIPAFHKAFSELSSNRGLCVIEVNPAAWTSATCPD
CGFCSKENRDGINFTCRKCGVSYHADIDVATLNIARVAVLGKPMSPADRERLGDTKKPRVA
RSRKTMRKRDISNSTVEAMVTA*

5 >CasΦ-4

MYSLEMADLKSEPSLLAKLLRDRFPKGYWLPKYWKLAIEKKRLTGEEAAACEYMADKQLDSPP
PNFRPPARCVILAKSRPFEDWPVHRVASKAQSFVIGLSEQGFAALRAAPPSTADARRDWLRS
HGASEDDLMALEAQLLETIMGNAISLHGGVLKKIDNANVKAARKLSGRNEARLNKGLQELPP
EQEGSAYGADGLLVNPPGLNLNIYCRKSCCPKPVKNTARFVGHYPGYLRDSDSILISGTMDR
10 LTIIIEGMPGHIPAWQREQGLVKPGGRRRLSGSESNMQRKQVDPSTGPRRSTRSGTVNRSNQR
TGRNGDPLLVEIRMKEDWVLLDARGLLRNLRWRESKRGLSCDHEDLSLSGLLALFSGDPVID
PVRNEVVFLYEGEIIIPVRSTKPVGTQSKKLLERQASMGPLTLISCDLGQTNLIAGRASAI
LTHGSLGVRSSVRIELDPEIIKSFERLRKADARLETEILTAAKETLSDEQRGEVNSHEKDSP
QTAKASLCRELGLHPPSLPWGQMGPSTTFIADMLISHGRDDDAFLSHGEFPTLEKRKKFDKR
15 FCLESRPLLSSETRKALNESLWEVKRTSSEYARLSQRKKEMARRAVNFVVEISRRKTGLSNV
IVNIEDLNVRIFHGGGKQAPGWDGFFRPKSENRFIQAIHKAFSDLAAHHGIPVIESDPQRT
SMTCEPCGHCDSKNRNGVRFLCKGCGASMDADFDAACRNLERVALTGKMPKPKSTSCERLLS
ATTGKVCSDHSLSHDAIEKAS*

>CasΦ-5

MSSLPTPLELLKQKHADLFKGLQFSSKDNKMAGKVLKKDGEEAALAFLSERGVSRGELPNFR
PPAKTLVVAQSRPFEEFPIYRVSEAIQLYVYSLSVKELETVPSPGSSTKKEHQRFQDSSVPD
FGYTSVQGLNKIFGLARGIYLGVITRGENQLQKAKSKHEALNKKRRASGEAETFDPTPYEY
MTPERKLAKPPGVNHSIMCYVDISVDEFDFRNPDGIVLPSEYAGYCREINTAIEKGTVDRLG
HLKGGPGYIPGHQRKESTTEGPKINFRKGRIRRSYALYAKRDSRRVRQGLALPSYRHHMM
25 RLNSNAESAILAVIFFGKDWVFDLRGLLRNVWRNLFVDGSTPSTLLGMFGDPVIDPKRGV
VAFCYKEQIVPVVSKSITKMKVKAPELLNKLYLKSEDPLVLVAIDLQGTNPVGVGVYRVMNAS
LDYEVVTRFALESELLREIESYRQRTNAFEAQIRAETFDAMTSEEQEEITRVRAFSASKAKE
NVCHRFMPVDAVDWATMGSNTHIAKWVMRHGDPSLVEVLEYRKDNEIKLDKNGVPKVKVL
TDKRIANLTSIRLRFSQETSKHYNDTMWELRRKHPVYQKLSKSKADFSRRVVNSIIRRVNHL
30 VPRARIVFIIEDLKNLGKVFHSGKRELGWDSYFEPKSENRFIQVLHKAFFSETGKHKGYYI
IECWPNWTSCTCPKSCCDSNRHGEVFRCLACGYTCNTDFGTAPDNLVKIATTGKGLPGPK
KRCKGSSKGNPKIARSSETGVSVTESGAPKVKSSPTQTSQSSSQSAP*

>CasΦ-6

MNKIEKEKTPLAKLMNENFAGLRFPFAIIKQAGKKLLKEGELKTIEYMTGKGSIEPLPNFKP
 PVKCLIVAKRRDLKYFPICKASCEIQSYVYSLNYKDFMDYFSTPMTSQKHHEEFFKKSGLNI
 EYQNVAGLNLIFFNNVKNTYNGVILKVKNRNEKLKKKAIKNNYEFEEIKTFNDDGCLINKPGI
 NNVIYCFQSSISPKILKNITHLPKEYNDYDCSVDRNIIQKYVSRLDIPESQPGHVPEWQRKLP
 5 EFNNTNNPRRRRKWYSNGRNISKGYSVDQVNQAKIEDSLLAQIKIGEDWIILDIRGLLRDLN
 RRELISYKNKLTIKDVLGFFSDYPIIDIKKNLVTFCYKEGVIQVVSQKSIGNKKSQLEKL
 IENKPIALVSIDLQGTNPVSVKISKLNKINNKISIESFTYRFLNEEILKEIEKYRKDYDKLE
 LKLINEA

>CasΦ-7

10 MSNTAVSTREHMSNKTTPPSPLSLLLRAHFPGKLFESQDYKIAGKKLRDGGPEAVISYLTGK
 GQAKLKDVKPPAKAFVIAQSRPFIEWDLVRVSRQIQEKIFGIPATKGRPKQDGLSETAFNEA
 VASLEVDGKSKLNEETRAAFYEVGLDAPSLHAQAQNALIKSAISIREGLVKKVENRNEKNL
 SKTKRRKEAGEEATFVEEKHAHERGYLIHPPGVNQTIPGYQAVVIKSCPSDFIGLPSGCLAK
 ESAEALTDYLP HDRMTIPKGQPGYVPEWQHPLNRRKNRRRRDWYSASLNKPKATCSKRS GT
 15 PNRKNSRTDQIQSGRFKGAIPVLMRFQDEWVIIDIRGLLRNARYRKLLKEKSTIPDLLSLFT
 GDPSIDMRQGVCTFIYKAGQACSAKMVKTKNAP EILSELTKSGPVVLVSIDLQGTNP IAAKV
 SRVTQLSDGQLSHETLLRELLSNDSSDGKEIARYRVASDRLRDKLANLAVERLSPEHKSEIL
 RAKNDTPALCKARVCAALGLNPEMIAWDKMTPYTEFLATAYLEKGGDRKVATLKPKNRPEML
 RRDIFKFGTEGVRIEVSPEAAEAYREAQWDLQRTSPEYLRLLSTWKQELTKRIILNQLRHKAAK
 20 SSQCEVVVMAFEDLNKMMHGNGKWADGGWDAFFIKKRENRFWMQAFHKS TELGAHKGVPT
 IEVTPHRTSITCTKCGHCDKANRDGERFACQKCGFVAHADLEIATDNI ERVALTGKPM PKPE
 SERSGDAKKSVGARKAAFKPEEDAEAAE*

>CasΦ-8

MIKPTVSQFLTPGFKLIRNHSRTAGLKLKNEGEEACKKFVRENEIPKDECPNFQGGPAIANI
 25 IAKSREFTWEIYQSSLAIQEVI FTLPKDKLPEPILKEEWRAQWLSEHGLDTPYKEAAGLN
 LIIKNAVNTYKGVQVKVDNKNKNNLAKINRKNEIAKLNGEQEISFEEIKAFDDKGYLLQKPS
 PNKSIYCYQSVSPKPFITSKYHNVNLP EEYIGYYRKSNEPIVSPYQFDR LRPIGEPGYVPK
 WQYTFLSKKENKRRKLSKRIKNVSPILGIICIKKDWCVFDMRGLLR TNHWKKYHKPTDSIND
 LFDYFTGDPVIDTKANVVRFRYKMENGIVNYKPVREKKGKELLENICDQNGSCKLATVDVGQ
 30 NNPVAIGLFELKKVNGELTKTLISRHPTPIDFCNKITAYRERYDKLESSIKLDAIKQLTSEQ
 KIEVDNYNNTFTPQNTKQIVCSKLNINPNDLPWDKMISGTHFISEKAQVSNKSEIYFTSTDK
 GKTKDVMKSDYKWFQDYKPKLSKEVRDALSDIEWRLRRESLEFNKLSKSREQDARQLANWIS
 SMCDVIGIENLVKKNNFFGGSGKREPGWDNFYKPKKENRWWINAIHKALTELSQNKGKRVI L

LPAMRTSITCPKCKYCDKSNRNGEKFNCLKCGIELNADIDVATENLATVAITAQSMPKPTCE
 RSGDAKKPVRARKAKAPEFHDKLAPSYTVVLREAV*

>CasΦ-9

MRSSREIGDKILMRQPAEKTAQVFRQEVIGTQKLSSGDAKTAGRLYKQGKMEAAREWLLKG
 5 ARDDVPPNFQPPAKCLVVAVSHPFEEWDISKTNHDVQAYIYAQPLQAEGHLNGLSEKWEDTS
 ADQHKLWFEKTGVPDRGLPVQAINKIAKAAVNRAFGVVRKVENRNEKRRSRDNRIA EHNREN
 GLTEVVREAPEVATNADGFLHPPGIDPSILSYASVSPVPYNSSKHSFVRLPEEYQAYNVEP
 DAPIPQFVVEDRFAIPPGQPGYVPEWQRLKCS TNKHRMRQWSNQDYKPKAGRRAKPLEFQA
 HLTRERAKGALLVVMRIKEDWVVF DVRGLLRNVEWRKVLSEEAREKLT LKGLLDLFTGDPVI
 10 DTKRGIVTFLYKAEITKILSKRTVKTKNARDLLLRLTEPGEDGLRREVGLVAVDLGQTHPIA
 AAIYRIGRTSAGALESTVLHRQGLREDQKEKLKEYRKRHTALDSRLRKEAFETLSVEQQKEI
 VTVSGSGAQITKDKVCNYLGVD PSTLPWEKMGSYTHFISDDFLRRGGDPNIVHFD RQPKKGK
 VSKKSQRIKRSDSQWVGRMRPRLSQETAKARMEADWAAQNENEEYKRLARSKQELARWC VNT
 LLQNTRCITQCDEIVVIEDLNVKSLHGKGAREPGWDNFFTPKTENRWF IQILHKT FSELPK
 15 HRGEHVIEGCPLRTSITCPACSYCDKNSRNGEKFVCVACGATFHADFEVATYNLVR LATTGM
 PMPKSLERQGGGEKAGGARKARKKAKQVEKIVQANANVTMNGASLHSP*

>CasΦ-10

MDMLDTETNYATETPAQQQDYSPPKPKKAQRAPKGFSSKKARPEKKPPKPITLFTQKHFS GVR
 FLKRVIDASKILKLSESRITITFLEQAIERDGSAPPDVTPPVHNTIMAVTRPFEEWPEVILS
 20 KALQKHCYALTKKIKIKITWPKKGGPKKCLAAWSARTKIPLIPGQVQATNGLFDRIGSIYDGV
 EKKVTNRNANKKLEYDEAIKEGRNPAVPEYETAYNIDGT LINKPGYNPNLYITQSRTPR LIT
 EADRPLVEKILWQMVEKKTQSRNQARRARLEKAAHLQGLPVPKPFVPEKVDRSQKIEIRI IDP
 LDKIEPYMPQDRMAIKASQDGHVPYWQRPFLSKRRNRVRAGWGKQVSSIQAWLTGALLVIV
 RLGN EAF LADIRGALRNAQWRKLLKPDATYQSLFNLFTGDPVVNTRTNH LTMAYREGVVNIV
 25 KRSRFFKGRQTREHLLTLLGQGKT VAGVSFDLGQKHAAGLLAAHFGLGEDGNPVFTPIQACFL
 PQRYLDSL TNRYRNRYDALTLDMRRQSL LALTPAQQQEFADAQRDPGGQAKRACCLKLNLNPD
 EIRWDLVSGISTMISDLYIERGGDPRDVHQQVETKPKGKRKSEIRILKIRDGKWAYDFRPKI
 ADETRKAQREQLWKLQKASSEFERLSRYKINIARA IANWALQWGRELSGCDIVIPVLEDLNV
 GSKFFDGKGKWWLGWDNRFTP KKENRWFIVLHKAVAE LAPHRGVPVYEVMPHRTSMTCPAC
 30 HYCHPTNREGDRFECQSCHVVKN TDRDVAPYNI LRVAVEGKTLDRWQAEKKPQAE PDRPMIL
 IDNQES*

The asterisk (*) in the sequences above denotes a STOP codon. Alternatively, CasΦ-1 is also termed Cas12j ortholog 1. Thus, CasΦ-1- CasΦ-10 may also be referred to as Cas12j orthologs 1-10, respectively.

The Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9 undergoes a 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.

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 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).

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).

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.

While NHEJ-mediated DSB repair often disrupts the open reading frame of the gene, 5
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. 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 10
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 15
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.

In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence 20
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 25
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.

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. 30
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.”

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%,
5 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.

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
10 (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
15 (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).

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
20 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
25 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).

In some cases, a variant Cas9 protein has a reduced ability to cleave both the
30 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).

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).

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).

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).

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.

10 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.

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.

20 In some embodiments, a modified SpCas9 including amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (SpCas9-MQKFRAER) and having specificity for the altered PAM 5'-NGC-3' is used.

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 doesn't need 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.

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/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ. 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.

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 alpha-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.

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 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, 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.

The amino acid sequence of wild type *Francisella novicida* Cpf1 follows. D917, E1006, and D1255 are bolded and underlined.

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEIIKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFNYYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS

KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 5 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI DRGERHLAYYTLD
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 10 AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVPYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGT ELDYLISPVADVNGNF
 FDSRQAPKNMPQDA DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

15 The amino acid sequence of *Francisella novicida* CpfI D917A follows. (A917,
 E1006, and D1255 are bolded and underlined).

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 20 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD LAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 25 DNLAQISIKYQNQGKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYS SKGRPNLHTL
 30 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI ARGERHLAYYTLD
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC

YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTLDYLISPVADVNGNF
 FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

The amino acid sequence of *Francisella novicida* Cpf1 E1006A follows. (D917,

5 A1006, and D1255 are bolded and underlined).

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKSSDDNLQKDFKSADTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 10 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS FYEQIAAFKTVEEKSIKE
 TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNGKDLLQASAEDDVKA IKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 15 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKY YLGVMNKKNNKIFDDKA IENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSI SKHPEWKDFGFRFSDT
 QRYNSID E FYREVENQGYKLT FENISESYIDS VVNQGLYLFQIYNKDFSAYS KGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAE LFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
 20 KDKRFTEDKFFFHC PITINFKSSGANKFNDEINLLLKEKANDVHILSI DRGERHLAYYTLVD
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVF ADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYVYPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 25 KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTLDYLISPVADVNGNF
 FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

The amino acid sequence of *Francisella novicida* Cpf1 D1255A follows. (D917,

E1006, and A1255 mutation positions are bolded and underlined).

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 30 EEILSSVCISEDLLQNYSDVYFKLKSSDDNLQKDFKSADTIKKQISEYIKDSEKFKNLFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQS FYEQIAAFKTVEEKSIKE

TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 5 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI DRGERHLAYYTIVD
 10 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVPPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLT SVLNTILQMRNSKTGTELDYLISPVADVNGNF
 15 FDSRQAPKNMPQDA AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

The amino acid sequence of *Francisella novicida* Cpf1 D917A/E1006A follows.

(A917, A1006, and D1255 are bolded and underlined).

MSIIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSADTIKKQISEYIKDSEKFKNLFN
 20 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 25 KKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 30 QRYNSIDEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI ARGERHLAYYTIVD
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVF ADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA

YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

- 5 The amino acid sequence of *Francisella novicida* CpfI D917A/D1255A follows.
(A917, E1006, and A1255 are bolded and underlined).

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDLQKDFKSADTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTFYFKGFHENR
10 KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
15 DNLAQISIKYQNQGKKDLLQASAEDDVKA IKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYLLGVMNKKNNKIFDDKA IENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSIKHPWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYS KGRPNLHTL
20 YWKALFDERNLQDVVYKLNGEAE LFYRKQSI PKKITHPAKEAIANKNDNPKKESVF EYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI ARGERH LAYYTLDV
GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF EDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNILVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
25 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

- The amino acid sequence of *Francisella novicida* CpfI E1006A/D1255A follows.
(D917, A1006, and A1255 are bolded and underlined).
- 30 MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDLQKDFKSADTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTFYFKGFHENR
KNVYSSNDIPTSI IYRIVDDNL PKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ

INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 5 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIAKNKDNPKKESVF EYDLI
 10 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI DRGERHLAYYTLDV
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
 AKLVIEYNAIVVF ADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLR
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVPYPTKELEKLL
 15 KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
 FDSRQAPKNMPQDA AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

The amino acid sequence of *Francisella novicida* CpfI D917A/E1006A/D1255A
 follows. (A917, A1006, and A1255 are bolded and underlined).

MSIIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
 20 EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSADTIKKQISEYIKDSEKFKLNFN
 QNLIDAKKGQESDLILWLKQSKDNGIELFKANS DITDIDEALEI IKSFKGWTTYFKGFHENR
 KNVYSSNDIPTSI IYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKD LAEELTFDIDY
 KTSEVNQRVFSLDEVFEIANFN NYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
 INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKSIKE
 25 TSLLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
 KKEQELIAKKTEKAKYLSLETIKLALAEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
 DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
 EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
 LFIKDDKYLLGVMNKKNNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFY
 30 NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
 QRYNSIDEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYS SKGRPNLHTL
 YWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEAIAKNKDNPKKESVF EYDLI
 KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI ARGERHLAYYTLDV
 GKGNI IKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI

AKLVIEYNAlVVFA¹DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNILVFKDNEFDKTGGVLRA
 YQLTAPFETFKKMGKQTGI IYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
 YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
 KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
 5 FDSRQAPKNMPQDA²ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN.

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.

In some embodiments, the Cas 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 domain comprises a N579A mutation, or a corresponding mutation in any of the amino acid sequences provided herein.

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 NNGRRT 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.

The amino acid sequence of an exemplary SaCas9 is as follows:

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH
 RIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV
 EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKV
 QKAYHQLDQSFIDTYIDLLETRRTYYEGPGEPSFGWKDIKEWYEMLMGHCTYFPEELRSVK
 30 YAYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIENVFKQKKKPTLKQIAKEILVNEEDI
 KGYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLN
 SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPPKVDLSQQK
 EIPTTLVDDFILSPVVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQK
 RNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEV DHI

IPRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAAGKGRISK
 TKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTS
 FLRRKWKFKKERNKGYKHHAEDALI IANADFIKKEWKKLDKAKKVMENQMFEKQAESMPEI
 ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLN
 5 GLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKY
 SKKDNGPVIKKIKYGNKLNHLDDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKN
 LDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
 EVNMIDITYREYLENMNDKRPPRI IKTASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.
 In this sequence, residue N579, which is underlined and in bold, may be mutated (*e.g.*, to a
 10 A579) to yield a SaCas9 nickase.

The amino acid sequence of an exemplary SaCas9n is as follows:

KRNYILGLDIGITSVGYGIIDYETRVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
 IQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
 EDTGNESTKEQISRNKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVQ
 15 KAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
 AYNADLYNALNDLNNLVITRDENEKLEYEYEFQIIENVFQKQKKPTLKQIAKEILVNEEDIK
 GYRVTSTGKPEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNS
 ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
 IPTTLVDDFILSPVVKRSFIQSIVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKR
 20 NRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
 PRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAAGKGRISK
 KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTS
 LRRKWKFKKERNKGYKHHAEDALI IANADFIKKEWKKLDKAKKVMENQMFEKQAESMPEIE
 TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLN
 25 LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYS
 KKDNGPVIKKIKYGNKLNHLDDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
 DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
 VNMIDITYREYLENMNDKRPPRI IKTASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

In this sequence, residue A579, which can be mutated from N579 to yield a SaCas9
 30 nickase, is underlined and in bold.

The amino acid sequences of an exemplary SaKKH Cas9 is as follows:

KRNYILGLDIGITSVGYGIIDYETRVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
 IQRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
 EDTGNESTKEQISRNKALEEKYVAELQLERLKKDGEVRGSINRFKTS DYVKEAKQLLKVQ

KAYHQLDQSFIDTYIDLLETRRTYYEGPGEQSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
 AYNADLYNALNDLNNLVITRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
 GYRVTSTGKPEFTNLKVYHDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNS
 ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
 5 IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
 NRQTNERIEEIIIRTTGKENAKYLIIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHII
 PRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISK
 KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSF
 LRRKWKFKKERNKGYKHAEDALIIANADFIKFWKKLDAKAKVMENQMFEKQAESMPEIE
 10 TEQEYKEIFITPHQIKHIKDFKDYKYSRVDKKPNR**K**LINDTLYSTRKDDKGNTLIVNNLNG
 LYDKDNDKLKKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDEKNPLYKYEETGNLYTKYS
 KKDNGPVIKKIKYYGNKLNALDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
 DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY**K**NDLIKINGELYRVIGVNNDDLNRIE
 VNMDITYREYLENMNDKRPP**H**IKTIASKTQSIKKYSTDILGNLYEVKSKKHPPQIIKKG.

15 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.

High fidelity Cas9 domains

20 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 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-
 25 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
 30 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%.

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
 5 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
 10 engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); the entire contents of each are incorporated herein by reference.

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
 15 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
 20 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 underline
 MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 25 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 30 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASAQSFIERMT
AFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTTTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDF

LKSDGFANRNFM^ALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGGLSELDKAGFIKRQLVETRA^AITKH
 5 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLLAS
 10 HYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL SQ
 LGGD.

Guide Polynucleotides

15 In an embodiment, the guide polynucleotide is a guide RNA. 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
 20 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
 25 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
 30 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.

5 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.

10 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
15 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
20 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.

In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via for example complementary base
25 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).

In type II CRISPR systems, targeting of a nucleic acid by a CRISPR protein (*e.g.*,
30 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.

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
 5 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.

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
 10 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.

Typically, a guide polynucleotide (*e.g.*, crRNA/trRNA complex or a gRNA)
 15 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
 20 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
 25 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
 30 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.

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
5 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.

As discussed above, a guide RNA or a guide polynucleotide can be an expression
10 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
15 gene delivery.

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
20 of plasmid comprising encoding sequence for a guide RNA.

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
25 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.

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
30 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.

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
 5 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
 10 about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

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
 15 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.

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
 20 exon or in some cases, multiple guide RNAs that can target different exons. An exon and an intron of a gene can be targeted.

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,
 25 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,
 30 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, or 1-100 nucleotides.

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.

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.

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 publicly 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.

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.

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

5 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

10 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.

The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the guide RNA can be synthesized using standard

15 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

20 tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

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,

25 at least 30 g RNA, at least 50 gRNA) comprised in a base editor system. The multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

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

30 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.

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).

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.

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.

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.

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 quencher 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.

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.

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.

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 3'-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

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).

The PAM sequence is essential for target binding, but the exact sequence depends on a type of Cas protein.

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

5 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

10 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. 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
spCas9-MQKFRAER	NGC
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)

15

In some embodiments, the PAM is NGC. In some embodiments, the NGC PAM is recognized by a Cas9 variant. In some embodiments, the NGC PAM variant includes one or

more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed “MQKFRAER”).

In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the NGT PAM variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or 1219. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1219, 1335, 1337, 1218. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335. In some embodiments, the NGT PAM variant is selected from the set of targeted mutations provided in **Tables 2 and 3** below.

Table 2: 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 3: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

Variant	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			

In some embodiments, the NGT PAM variant is selected from variant 5, 7, 28, 31, or 36 in Tables 2 and 3. In some embodiments, the variants have improved NGT PAM recognition.

- 5 In some embodiments, the NGT PAM variants have mutations at residues 1219, 1335, 1337, and/or 1218. In some embodiments, the NGT PAM variant is selected with mutations for improved recognition from the variants provided in **Table 4** below.

Table 4: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 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
---	---	---	---	---

In some embodiments, the NGT PAM is selected from the variants provided in **Table 5** below.

Table 5. NGT PAM variants

	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

- 5 In some embodiments the NGTN variant is variant 1. In some embodiments, the NGTN variant is variant 2. In some embodiments, the NGTN variant is variant 3. In some embodiments, the NGTN variant is variant 4. In some embodiments, the NGTN variant is variant 5. In some embodiments, the NGTN variant is variant 6.
- 10 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 any amino acid except for D. In
- 15 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.
- 20

- 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
- 25 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

T1337T1337X 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

5 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

10 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 domain comprises a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some embodiments, the Cas9 domains of any of the fusion proteins provided herein

15 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

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

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 embodiments, providing PAM on a

25 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.

In an embodiment, *S. pyogenes* Cas9 (SpCas9) can be used as a CRISPR endonuclease for genome engineering. However, others can be used. In some embodiments,

30 a different endonuclease can be used to target certain genomic targets. In some embodiments, 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

5 modifying target genes in mammalian cells *in vitro* and in mice *in vivo*. In some embodiments, a Cas protein can target a different PAM sequence. In some embodiments, a target gene can be adjacent to a Cas9 PAM, 5'-NGG, for example. In other embodiments, 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

10 *Neisseria meningitidis* (5'-NNNGATT) can also be found adjacent to a target gene.

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 embodiments, an adjacent cut can be or can be about 3 base pairs upstream of a PAM. In some embodiments,

15 an adjacent cut can be or can be about 10 base pairs upstream of a PAM. In some embodiments, 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. The sequences of

20 exemplary SpCas9 proteins capable of binding a PAM sequence follow:

The amino acid sequence of an exemplary PAM-binding SpCas9 is as follows:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 25 QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPIYVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 30 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLIKDKDFLDNEENEDILEDIV
 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV
 DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL

QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 5 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 10 LGGD.

The amino acid sequence of an exemplary PAM-binding SpCas9n is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 15 QLVQTYNQLFEE NPINASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYFPFLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 20 NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDL LKIIKDKDFLDNEENEDILEDIV
 LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
 DELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHPVENTQL
 25 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNS
 30 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLAS
 HYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGD.

The amino acid sequence of an exemplary PAM-binding SpEQR Cas9 is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFVEEDKKHERHPIFGNIVDE
 VAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQ
 5 LVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLSGLT
 PNFKSNFDLAEDAQLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
 ITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
 KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKD
 NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS FIERMTN
 10 FDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
 TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDI EDIVL
 TLT LFEDEMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
 KSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQD SLHEHIANLAGSPA IKKGILQTVKVVD
 ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
 15 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
 VPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHV
 AQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV V
 GTALIKKYPKLESEFVYG DYKVYDVRKMI AKSEQEIGKAT AKYFFYSNIMNFFKTEITLANG
 EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSD
 20 KLIARKKDWD PKKYGGF **E**SPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPI
 DFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
 YEKLKGS PEDNEQQLFVEQH KHYLDEII EQISEFSKR VILADANLDKVL SAYNKH RDKPIR
 EQAENIIHLFTLTNLGAPAAFKYFDTTIDRK **QYR**STKEVL DATLIHQ SITGLYETRIDLSQL
 GGD. In this sequence, residues E1135, Q1335 and R1337, which can be mutated from
 25 D1135, R1335, and T1337 to yield a SpEQR Cas9, are underlined and in bold.

The amino acid sequence of an exemplary PAM-binding SpVQR Cas9 is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF I
 30 QLVQTYNQ LFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLSGL
 TPNFKSNFDLAEDAQLQ LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQ LPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLK
 DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS FIERMT

NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLEFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 5 DELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 10 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSSFENP
 IDFLEAGKYKEVKKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLAS
 HYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RD KPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI TGLYETRIDL SQ
 15 LGGD. In this sequence, residues V1135, Q1335, and R1337, which can be mutated from
 D1135, R1335, and T1337 to yield a SpVQR Cas9, are underlined and in bold.

The amino acid sequence of an exemplary PAM-binding SpVRER Cas9 is as follows:

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVD
 20 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGL
 TPNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
 25 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 LTLTLEFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
 30 DELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN

GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSSFENP
 IDFLKAGYKEVKKDLIIKLPKYSLFELENGRKRMLASARRELQKGNELALPSKYVNFYLLAS
 HYEKLKGSPEQKQQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 5 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVL DATLIHQSI TGLYETRIDLSQ
 LGGD. In the above sequence, residues V1135, R1218, Q1335, and R1337, which can be
 mutated from D1134, G1217, R1335, and T1337 to yield a SpVRER Cas9, are underlined
 and in bold.

In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some
 10 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
 15 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence
 having a NAA PAM sequence.

Exemplary SpyMacCas9

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAET
 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 20 EVAYHEKYPTIYHLRKKLADSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALLSLGL
 TPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 YKFIKPILEKMDGTEELLVVLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYFPFLK
 25 DNREKIEKILTFRIPIYYVGPLARGNSRFWMTRKSEETITPWNFEVVVDKGASQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDL LKIIKDKDFLDNEENEDILEDIV
 LTTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIIKKGILQTVKIV
 30 DELVKVMGHKPENIVIEARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
 NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
 VPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
 AQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANG

5 EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEIQTVGQNGGLFDDNPKS P
 LEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRD
 RGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSCKSQILLYHAHHLSDSL
 SNDYLNHNQQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGF
 10 TQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

In some cases, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A 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
 10 Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A 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).
 15 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 D1218A 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
 20 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,
 25 H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

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
 30 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); Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition"

Nature Biotechnology 33, 1293-1298 (2015); R.T. Walton *et al.* “Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants” *Science* 10.1126/science.aba8853 (2020); Hu *et al.* “Evolved Cas9 variants with broad PAM compatibility and high DNA specificity,” *Nature*, 2018 Apr. 5, 556(7699), 57-63; Miller *et al.*, “Continuous evolution of SpCas9 variants compatible with non-G PAMs” *Nat. Biotechnol.*, 2020 Apr;38(4):471-481; the entire contents of each are hereby incorporated by reference.

Fusion proteins comprising a Cas9 domain and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain or other nucleic acid programmable DNA binding protein and one or more adenosine deaminase domain, cytidine deaminase domain, and/or DNA glycosylase domains. It should be appreciated that the Cas9 domain may be any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein. In some embodiments, any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein may be fused with any of the cytidine deaminases and adenosine deaminases provided herein. The domains of the base editors disclosed herein can be arranged in any order.

For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH₂-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[cytidine deaminase]-[Cas9 domain]-COOH;
 NH₂-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-COOH;
 NH₂-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or
 NH₂-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*8 and a cytidine deaminase. In some embodiments, the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

Exemplary fusion protein structures include the following:

NH₂-[adenosine deaminase]-[Cas9]-[cytidine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas9]-[adenosine deaminase]-COOH;

NH₂-[TadA*8]-[Cas9]-[cytidine deaminase]-COOH; or

NH₂-[cytidine deaminase]-[Cas9]-[TadA*8]-COOH.

In some embodiments, the fusion proteins comprising a cytidine deaminase, abasic editor, and adenosine deaminase and a napDNAbp (e.g., Cas9 domain) do not include a linker sequence. In some embodiments, a linker is present between the cytidine deaminase and adenosine deaminase domains and the napDNAbp. In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker. In some embodiments, the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided below in the section entitled “Linkers”.

In some embodiments, the general architecture of exemplary Cas9 or Cas12 fusion proteins with a cytidine deaminase, adenosine deaminase and a Cas9 or Cas12 domain comprises any one of the following structures, where NLS is a nuclear localization sequence (e.g., any NLS provided herein), NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein.

NH₂-NLS-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-COOH;

NH₂-NLS-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-COOH;

NH₂-NLS-[adenosine deaminase] [cytidine deaminase]-[Cas9 domain]-COOH;

NH₂-NLS-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-COOH;

NH₂-NLS-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH;

NH₂-NLS-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-NLS-COOH;

NH₂-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-NLS-COOH;

NH₂-[adenosine deaminase] [cytidine deaminase]-[Cas9 domain]-NLS-COOH;

NH₂-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-NLS-COOH;

NH₂-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-NLS-COOH; or

NH₂-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-NLS-COOH.

In some embodiments, the NLS is present in a linker or the NLS is flanked by linkers, for example 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:

PKKKRKVEGADKRTADGSEFESPKKKRKV.

In some embodiments, the fusion proteins comprising a cytidine deaminase, adenosine deaminase, a Cas9 domain and an NLS do not comprise a linker sequence. In some embodiments, linker sequences between one or more of the domains or proteins (*e.g.*, cytidine deaminase, adenosine deaminase, Cas9 domain or NLS) are present.

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.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/2017/044935 and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

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

5 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

10 comprises the amino acid sequence PKKKRKVEGADKRTADGSEFESPKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGKTR, RKSGKIAAIVVKRPRKPKKKRKV, or MDSLLMNRKFLYQFKNVRWAKGRRETYLC. 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

15 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

20 sequence of an exemplary bipartite NLS follows:
PKKKRKVEGADKRTADGSEFESPKKKRKV.

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. In some embodiments, the general architecture of exemplary Cas9

25 fusion proteins with an adenosine deaminase or a cytidine deaminase and a Cas9 domain comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.*, any NLS provided herein), NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:

NH₂-NLS-[adenosine deaminase]-[Cas9 domain]-COOH;
 30 NH₂-NLS [Cas9 domain]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas9 domain]-NLS-COOH;
 NH₂-[Cas9 domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-NLS-[cytidine deaminase]-[Cas9 domain]-COOH;

NH₂-NLS [Cas9 domain]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9 domain]-NLS-COOH; or
 NH₂-[Cas9 domain]-[cytidine deaminase]-NLS-COOH.

It should be appreciated that the fusion proteins of the present disclosure may
 5 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)
 10 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
 15 some embodiments, the fusion protein comprises one or more His tags.

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
 20 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.

CRISPR enzymes used in the methods can comprise about 6 NLSs. An NLS is
 25 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.

Fusion proteins with Internal Insertions

30 Provided herein are fusion proteins comprising a heterologous polypeptide fused to a
 nucleic acid programmable nucleic acid binding protein, for example, a napDNAbp. A
 heterologous polypeptide can be a polypeptide that is not found in the native or wild-type
 napDNAbp polypeptide sequence. The heterologous polypeptide can be fused to the

napDNAbp at a C-terminal end of the napDNAbp, an N-terminal end of the napDNAbp, or inserted at an internal location of the napDNAbp. In some embodiments, the heterologous polypeptide is inserted at an internal location of the napDNAbp.

In some embodiments, the heterologous polypeptide is a deaminase or a functional
 5 fragment thereof. For example, a fusion protein can comprise a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 (*e.g.*, Cas12b/C2c1), polypeptide. The deaminase in a fusion protein can be an adenosine deaminase. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA7.10 or TadA*8). In some
 10 embodiments, the TadA is a TadA*8. TadA sequences (*e.g.*, TadA7.10 or TadA*8) as described herein are suitable deaminases for the above-described fusion proteins.

The deaminase can be a circular permutant deaminase. For example, the deaminase can be a circular permutant adenosine deaminase. In some embodiments, the deaminase is a circular permutant TadA, circularly permuted at amino acid residue 116 as numbered in the
 15 TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permuted at amino acid residue 136 as numbered in the TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permuted at amino acid residue 65 as numbered in the TadA reference sequence.

The fusion protein can comprise more than one deaminase. The fusion protein can comprise, for example, 1, 2, 3, 4, 5 or more deaminases. In some embodiments, the fusion
 20 protein comprises one deaminase. In some embodiments, the fusion protein comprises two deaminases. The two or more deaminases in a fusion protein can be an adenosine deaminase, cytidine deaminase, or a combination thereof. The two or more deaminases can be homodimers. The two or more deaminases can be heterodimers. The two or more deaminases can be inserted in tandem in the napDNAbp. In some embodiments, the two or
 25 more deaminases may not be in tandem in the napDNAbp.

In some embodiments, the napDNAbp in the fusion protein is a Cas9 polypeptide or a fragment thereof. The Cas9 polypeptide can be a variant Cas9 polypeptide. In some
 30 embodiments, the Cas9 polypeptide is a Cas9 nickase (nCas9) polypeptide or a fragment thereof. In some embodiments, the Cas9 polypeptide is a nuclease dead Cas9 (dCas9) polypeptide or a fragment thereof. The Cas9 polypeptide in a fusion protein can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein may not be a full length Cas9 polypeptide. The Cas9 polypeptide can be truncated, for example, at a N-terminal or C-terminal end relative to a naturally-occurring Cas9 protein. The Cas9 polypeptide can be a circularly permuted Cas9 protein. The Cas9 polypeptide can be a

fragment, a portion, or a domain of a Cas9 polypeptide, that is still capable of binding the target polynucleotide and a guide nucleic acid sequence.

In some embodiments, the Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), or fragments or variants thereof.

The Cas9 polypeptide of a fusion protein can comprise 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 Cas9 polypeptide.

The Cas9 polypeptide of a fusion protein can comprise 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 the Cas9 amino acid sequence set forth below (called the “Cas9 reference sequence” below):

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEAT
 15 RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGL
 TPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
 20 YKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLK
 DNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGGASQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIV
 LTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 25 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV
DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKROLVETROITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
 30 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTETVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPNP
 IDFLEAGKYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLAS
 HYEKLKGSPEQKQKQFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI

REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI^{TGLYETRIDLSQ}
 LGGD (single underline: HNH domain; double underline: RuvC domain).

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas9 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas9 and one or more deaminase domains, *e.g.*, adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas9 sequences are also useful for highly specific and efficient base editing of target sequences. In an embodiment, a chimeric Cas9 fusion protein contains a heterologous catalytic domain (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas9 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas9. In some embodiments, an adenosine deaminase is fused within a Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase fused to the N-terminus.

Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas9 are provided as follows:

NH₂-[Cas9(adenosine deaminase)]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas9(adenosine deaminase)]-COOH;
 NH₂-[Cas9(cytidine deaminase)]-[adenosine deaminase]-COOH; or
 NH₂-[adenosine deaminase]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA*8 fused to the N-terminus. Exemplary

structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas9 are provided as follows:

- NH₂-[Cas9(TadA*8)]-[cytidine deaminase]-COOH;
- NH₂-[cytidine deaminase]-[Cas9(TadA*8)]-COOH;
- 5 NH₂-[Cas9(cytidine deaminase)]-[TadA*8]-COOH; or
- NH₂-[TadA*8]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

- The heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp (*e.g.*, Cas9 or Cas12 (*e.g.*, Cas12b/C2c1)) at a suitable location, for example, such that the napDNAbp retains its ability to bind the target polynucleotide and a guide nucleic acid. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted into a napDNAbp without compromising function of the deaminase (*e.g.*, base editing activity) or the napDNAbp (*e.g.*, ability to bind to target nucleic acid and guide nucleic acid). A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp at, for example, a disordered region or a region comprising a high temperature factor or B-factor as shown by crystallographic studies. Regions of a protein that are less ordered, disordered, or unstructured, for example solvent exposed regions and loops, can be used for insertion without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp in a flexible loop region or a solvent-exposed region. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in a flexible loop of the Cas9 or the Cas12b/C2c1 polypeptide.

- In some embodiments, the insertion location of a deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is determined by B-factor analysis of the crystal structure of Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in regions of the Cas9 polypeptide comprising higher than average B-factors (*e.g.*, higher B factors compared to the total protein or the protein domain comprising the disordered region). B-factor or temperature factor can indicate the fluctuation of atoms from their average position (for example, as a result of temperature-dependent atomic vibrations or static disorder in a crystal lattice). A high B-

factor (*e.g.*, higher than average B-factor) for backbone atoms can be indicative of a region with relatively high local mobility. Such a region can be used for inserting a deaminase without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a C α atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, or greater than 200% more than the average B-factor for the total protein. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a C α atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% or greater than 200% more than the average B-factor for a Cas9 protein domain comprising the residue. Cas9 polypeptide positions comprising a higher than average B-factor can include, for example, residues 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence. Cas9 polypeptide regions comprising a higher than average B-factor can include, for example, residues 792-872, 792-906, and 2-791 as numbered in the above Cas9 reference sequence.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, or 1249-1250 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. It should be understood that the reference to the above Cas9 reference sequence with respect to insertion positions is for illustrative purposes. The insertions as discussed herein are not

limited to the Cas9 polypeptide sequence of the above Cas9 reference sequence, but include insertion at corresponding locations in variant Cas9 polypeptides, for example a Cas9 nickase (nCas9), nuclease dead Cas9 (dCas9), a Cas9 variant lacking a nuclease domain, a truncated Cas9, or a Cas9 domain lacking partial or complete HNH domain.

5 A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030,
10 1040-1041, 1068-1069, or 1247-1248 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the
15 heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue as described herein, or a corresponding amino acid residue in another
20 Cas9 polypeptide. In an embodiment, a heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-691, 569-578, 530-539, and 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The deaminase (*e.g.*, adenosine deaminase,
25 cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at the N-terminus or the C-terminus of the residue or replace the residue. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of the residue.

In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted at an amino
30 acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted in place of residues 792-872, 792-906, or 2-791 as numbered in the above Cas9 reference sequence, or a corresponding

amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the N-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted to replace an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a CBE (*e.g.*, APOBEC1) is inserted at an amino acid residue selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the ABE is inserted at the N-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the ABE is inserted at the C-terminus of an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the ABE is inserted to replace an amino acid selected from the group consisting of: 1016, 1023, 1029, 1040, 1069, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 768 as numbered in

the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 791 or is inserted at amino acid residue 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid 791, or is inserted to replace amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid

residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1022, or is
5 inserted at amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1022 or is inserted at the N-terminus of amino acid residue 1023, as numbered in the above Cas9 reference
10 sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1022 or is inserted at the C-terminus of amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In
15 some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1022, or is inserted to replace amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1026, or is
20 inserted at amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1026 or is inserted at the N-terminus of amino acid residue 1029, as numbered in the above Cas9 reference
25 sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1026 or is inserted at the C-terminus of amino acid residue 1029, as numbered in the above Cas9
30 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1026, or is inserted to replace amino acid residue 1029, as numbered in the above Cas9 reference sequence, or corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1052, or is inserted at amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1052 or is inserted at the N-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1052 or is inserted at the C-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1052, or is inserted to replace amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1067, or is inserted at amino acid residue 1068, or is inserted at amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9

polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1067 or is inserted at the N-terminus of amino acid residue 1068 or is inserted at the N-terminus of amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1067 or is inserted at the C-terminus of amino acid residue 1068 or is inserted at the C-terminus of amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1067, or is inserted to replace amino acid residue 1068, or is inserted to replace amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1246, or is inserted at amino acid residue 1247, or is inserted at amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1246 or is inserted at the N-terminus of amino acid residue 1247 or is inserted at the N-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 1246 or is inserted at the C-terminus of amino acid residue 1247 or is inserted at the C-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1246, or is inserted to replace amino acid residue 1247, or is inserted to replace amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a heterologous polypeptide (*e.g.*, deaminase) is inserted in a flexible loop of a Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, or 1298-1300 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of: 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, or 1248-1297 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted into a Cas9 polypeptide region corresponding to amino acid residues: 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002 – 1003, 943-947, 530-537, 568-579, 686-691, 1242-1247, 1298 – 1300, 1066-1077, 1052-1056, or 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted in place of a deleted region of a Cas9 polypeptide. The deleted region can correspond to an N-terminal or C-terminal portion of the Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-872 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-906 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 2-791 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 1017-1069 as numbered in the above Cas9 reference sequence, or corresponding amino acid residues thereof.

Exemplary internal fusions base editors are provided in **Table 6** below:

Table 6: Insertion loci in Cas9 proteins

BE ID	Modification	Other ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09

BE ID	Modification	Other ID
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term fusion single TadA helix truncated 165-end	ISLAY21
IBE029	TadA-Circular Permutant 116 ins1067	ISLAY29
IBE031	TadA- Circular Permutant 136 ins1248	ISLAY31
IBE032	TadA- Circular Permutant 136ins 1052	ISLAY32
IBE035	delta 792-872 TadA ins	ISLAY35
IBE036	delta 792-906 TadA ins	ISLAY36
IBE043	TadA-Circular Permutant 65 ins1246	ISLAY43
IBE044	TadA ins C-term truncate2 791	ISLAY44

A heterologous polypeptide (*e.g.*, deaminase) can be inserted within a structural or functional domain of a Cas9 polypeptide. A heterologous polypeptide (*e.g.*, deaminase) can be inserted between two structural or functional domains of a Cas9 polypeptide. A

5 heterologous polypeptide (*e.g.*, deaminase) can be inserted in place of a structural or functional domain of a Cas9 polypeptide, for example, after deleting the domain from the Cas9 polypeptide. The structural or functional domains of a Cas9 polypeptide can include, for example, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH.

In some embodiments, the Cas9 polypeptide lacks one or more domains selected from

10 the group consisting of: RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH domain. In some embodiments, the Cas9 polypeptide lacks a nuclease domain. In some embodiments, the Cas9 polypeptide lacks an HNH domain. In some embodiments, the Cas9 polypeptide lacks a portion of the HNH domain such that the Cas9 polypeptide has reduced or abolished HNH activity. In some embodiments, the Cas9 polypeptide comprises a deletion of the

15 nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some embodiments, the HNH domain is deleted and the deaminase is inserted in its place. In some embodiments, one or more of the RuvC domains is deleted and the deaminase is inserted in its place.

A fusion protein comprising a heterologous polypeptide can be flanked by a N-

20 terminal and a C-terminal fragment of a Cas9 polypeptide. In some embodiments, the fusion protein comprises a deaminase flanked by a N- terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The N terminal fragment or the C terminal fragment can bind the target polynucleotide sequence. The C-terminus of the N terminal fragment or the N-

25 terminus of the C terminal fragment can comprise a part of a flexible loop of a Cas9 polypeptide. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of an alpha-helix structure of the Cas9 polypeptide. The N-terminal fragment or the C-terminal fragment can comprise a DNA binding domain. The N-

terminal fragment or the C-terminal fragment can comprise a RuvC domain. The N-terminal fragment or the C-terminal fragment can comprise an HNH domain. In some embodiments, neither of the N-terminal fragment and the C-terminal fragment comprises an HNH domain.

In some embodiments, the C-terminus of the N terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. In some embodiments, the N-terminus of the C terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. The insertion location of different deaminases can be different in order to have proximity between the target nucleobase and an amino acid in the C-terminus of the N terminal Cas9 fragment or the N-terminus of the C terminal Cas9 fragment. For example, the insertion position of an ABE can be at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment of a fusion protein (*i.e.* the N-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the N-terminus of a Cas9 polypeptide. The N-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The N-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising 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% sequence identity to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The C-terminal Cas9 fragment of a fusion protein (*i.e.* the C-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the C-terminus of a Cas9 polypeptide. The C-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The C-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368,

718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising 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% sequence identity to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment and C-terminal Cas9 fragment of a fusion protein taken together may not correspond to a full-length naturally occurring Cas9 polypeptide sequence, for example, as set forth in the above Cas9 reference sequence.

The fusion protein described herein can effect targeted deamination with reduced deamination at non-target sites (*e.g.*, off-target sites), such as reduced genome wide spurious deamination. The fusion protein described herein can effect targeted deamination with reduced bystander deamination at non-target sites. The undesired deamination or off-target deamination can be reduced by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide. The undesired deamination or off-target deamination can be reduced by at least one-fold, at least two-fold, at least three-fold, at least four-fold, at least five-fold, at least tenfold, at least fifteen fold, at least twenty fold, at least thirty fold, at least forty fold, at least fifty fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, or at least hundred fold, compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) of the fusion protein deaminates no more than two nucleobases within the range of an R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than three nucleobases within the range of the R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleobases within the range of the R-loop. An R-loop is a three-stranded nucleic acid structure including a DNA:RNA hybrid, a DNA:DNA or an RNA: RNA complementary structure and the associated with single-stranded DNA. As used herein, an R-loop may be formed when a target polynucleotide is contacted with a CRISPR complex or a base editing complex, wherein a portion of a guide polynucleotide, *e.g.* a guide RNA, hybridizes with and displaces with a portion of a target polynucleotide, *e.g.* a target DNA. In

some embodiments, an R-loop comprises a hybridized region of a spacer sequence and a target DNA complementary sequence. An R-loop region may be of about 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 nucleobase pairs in length. In some
 5 embodiments, the R-loop region is about 20 nucleobase pairs in length. It should be understood that, as used herein, an R-loop region is not limited to the target DNA strand that hybridizes with the guide polynucleotide. For example, editing of a target nucleobase within an R-loop region may be to a DNA strand that comprises the complementary strand to a guide RNA, or may be to a DNA strand that is the opposing strand of the strand
 10 complementary to the guide RNA. In some embodiments, editing in the region of the R-loop comprises editing a nucleobase on non-complementary strand (protospacer strand) to a guide RNA in a target DNA sequence.

The fusion protein described herein can effect target deamination in an editing window different from canonical base editing. In some embodiments, a target nucleobase is
 15 from about 1 to about 20 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 2 to about 12 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 1 to 9 base pairs, about 2 to 10 base pairs, about 3 to 11 base pairs, about 4 to 12 base pairs, about 5 to 13 base pairs, about 6 to 14 base pairs, about 7 to 15
 20 base pairs, about 8 to 16 base pairs, about 9 to 17 base pairs, about 10 to 18 base pairs, about 11 to 19 base pairs, about 12 to 20 base pairs, about 1 to 7 base pairs, about 2 to 8 base pairs, about 3 to 9 base pairs, about 4 to 10 base pairs, about 5 to 11 base pairs, about 6 to 12 base pairs, about 7 to 13 base pairs, about 8 to 14 base pairs, about 9 to 15 base pairs, about 10 to 16 base pairs, about 11 to 17 base pairs, about 12 to 18 base pairs, about 13 to 19 base pairs,
 25 about 14 to 20 base pairs, about 1 to 5 base pairs, about 2 to 6 base pairs, about 3 to 7 base pairs, about 4 to 8 base pairs, about 5 to 9 base pairs, about 6 to 10 base pairs, about 7 to 11 base pairs, about 8 to 12 base pairs, about 9 to 13 base pairs, about 10 to 14 base pairs, about 11 to 15 base pairs, about 12 to 16 base pairs, about 13 to 17 base pairs, about 14 to 18 base pairs, about 15 to 19 base pairs, about 16 to 20 base pairs, about 1 to 3 base pairs, about 2 to 4
 30 base pairs, about 3 to 5 base pairs, about 4 to 6 base pairs, about 5 to 7 base pairs, about 6 to 8 base pairs, about 7 to 9 base pairs, about 8 to 10 base pairs, about 9 to 11 base pairs, about 10 to 12 base pairs, about 11 to 13 base pairs, about 12 to 14 base pairs, about 13 to 15 base pairs, about 14 to 16 base pairs, about 15 to 17 base pairs, about 16 to 18 base pairs, about 17 to 19 base pairs, about 18 to 20 base pairs away or upstream of the PAM sequence. In some

embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs away from or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, or 9 base pairs upstream of the PAM sequence. In some embodiments, a target nucleobase is about 2, 3, 4, or 6 base pairs
 5 upstream of the PAM sequence.

The fusion protein can comprise more than one heterologous polypeptide. For example, the fusion protein can additionally comprise one or more UGI domains and/or one or more nuclear localization signals. The two or more heterologous domains can be inserted in tandem. The two or more heterologous domains can be inserted at locations such that they
 10 are not in tandem in the NapDNABp.

A fusion protein can comprise a linker between the deaminase and the napDNABp polypeptide. The linker can be a peptide or a non-peptide linker. For example, the linker can be an XTEN, (GGGS)_n, (GGGGGS)_n, (G)_n, (EAAAK)_n, (GGS)_n, SGSETPGTSESATPES. In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9
 15 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the N-terminal and C-terminal fragments of napDNABp are connected to the deaminase with a linker. In some embodiments, the N-terminal and C-terminal fragments are joined to the deaminase domain without a linker. In some embodiments, the fusion protein comprises a
 20 linker between the N-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the N-terminal Cas9 fragment and the deaminase.

In some embodiments, the napDNABp in the fusion protein is a Cas12 polypeptide,
 25 *e.g.*, Cas12b/C2c1, or a fragment thereof. The Cas12 polypeptide can be a variant Cas12 polypeptide. In other embodiments, the N- or C-terminal fragments of the Cas12 polypeptide comprise a nucleic acid programmable DNA binding domain or a RuvC domain. In other embodiments, the fusion protein contains a linker between the Cas12 polypeptide and the catalytic domain. In other embodiments, the amino acid sequence of the linker is GSGSGS or
 30 GSSGSETPGTSESATPESSG. In other embodiments, the linker is a rigid linker. In other embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC.

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas12 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas12 and one or more deaminase domains, *e.g.*, adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas12 sequences are also useful for highly specific and efficient base editing of target sequences. In an embodiment, a chimeric Cas12 fusion protein contains a heterologous catalytic domain (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas12. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase fused to the N-terminus. Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas12 are provided as follows:

NH₂-[Cas12(adenosine deaminase)]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[Cas12(adenosine deaminase)]-COOH;
 NH₂-[Cas12(cytidine deaminase)]-[adenosine deaminase]-COOH; or
 NH₂-[adenosine deaminase]-[Cas12(cytidine deaminase)]-COOH;

In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas12 are provided as follows:

N-[Cas12(TadA*8)]-[cytidine deaminase]-C;

N-[cytidine deaminase]-[Cas12(TadA*8)]-C;
 N-[Cas12(cytidine deaminase)]-[TadA*8]-C; or
 N-[TadA*8]-[Cas12(cytidine deaminase)]-C.

In some embodiments, the “-” used in the general architecture above indicates the presence of
 5 an optional linker.

In other embodiments, the fusion protein contains one or more catalytic domains. In other embodiments, at least one of the one or more catalytic domains is inserted within the Cas12 polypeptide or is fused at the Cas12 N- terminus or C-terminus. In other
 10 embodiments, at least one of the one or more catalytic domains is inserted within a loop, an alpha helix region, an unstructured portion, or a solvent accessible portion of the Cas12 polypeptide. In other embodiments, the Cas12 polypeptide is Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the Cas12 polypeptide has at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus sp. V3-13* Cas12b, or *Alicyclobacillus*
 15 *acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus sp. V3-13* Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide has at least about 95% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus sp. V3-13* Cas12b,
 20 or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide contains or consists essentially of a fragment of *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus sp. V3-13* Cas12b, or *Alicyclobacillus acidiphilus* Cas12b.

In other embodiments, the catalytic domain is inserted between amino acid positions
 25 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605, or 344-345 of BhCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P153 and S154 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K255 and E256 of BhCas12b. In other
 30 embodiments, the catalytic domain is inserted between amino acids D980 and G981 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1019 and L1020 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids F534 and P535 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K604 and G605 of BhCas12b. In other

embodiments, the catalytic domain is inserted between amino acids H344 and F345 of BhCas12b. In other embodiments, catalytic domain is inserted between amino acid positions 147 and 148, 248 and 249, 299 and 300, 991 and 992, or 1031 and 1032 of BvCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 5 Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P147 and D148 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G248 and G249 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids P299 and E300 of BvCas12b. In other 10 embodiments, the catalytic domain is inserted between amino acids G991 and E992 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1031 and M1032 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acid positions 157 and 158, 258 and 259, 310 and 311, 1008 and 1009, or 1044 and 1045 of AaCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the 15 catalytic domain is inserted between amino acids P157 and G158 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids V258 and G259 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids D310 and P311 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1008 and E1009 of AaCas12b. In other embodiments, the catalytic 20 domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein contains a nuclear localization signal (*e.g.*, a bipartite nuclear localization signal). In other embodiments, the amino acid sequence of the nuclear localization signal is MAPKKRKVGIHGVPAA. In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence: 25 ATGGCCCCAAAGAAGAAGCGGAAGGTCTGGTATCCACGGAGTCCCAGCAGCC. In other embodiments, the Cas12b polypeptide contains a mutation that silences the catalytic activity of a RuvC domain. In other embodiments, the Cas12b polypeptide contains D574A, D829A and/or D952A mutations. In other embodiments, the fusion protein further contains a tag (*e.g.*, an influenza hemagglutinin tag).

30 In some embodiments, the fusion protein comprises a napDNAbp domain (*e.g.*, Cas12-derived domain) with an internally fused nucleobase editing domain (*e.g.*, all or a portion of a deaminase domain, *e.g.*, an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b. In some embodiments, the base editor comprises

a BhCas12b domain with an internally fused TadA*8 domain inserted at the loci provided in Table 7 below.

Table 7: Insertion loci in Cas12b proteins

BhCas12b	Insertion site	Inserted between aa
position 1	153	PS
position 2	255	KE
position 3	306	DE
position 4	980	DG
position 5	1019	KL
position 6	534	FP
position 7	604	KG
position 8	344	HF
BvCas12b	Insertion site	Inserted between aa
position 1	147	PD
position 2	248	GG
position 3	299	PE
position 4	991	GE
position 5	1031	KM
AaCas12b	Insertion site	Inserted between aa
position 1	157	PG
position 2	258	VG
position 3	310	DP
position 4	1008	GE
position 5	1044	GK

5 By way of nonlimiting example, an adenosine deaminase (*e.g.*, ABE8.13) may be inserted into a BhCas12b to produce a fusion protein (*e.g.*, ABE8.13-BhCas12b) that effectively edits a nucleic acid sequence.

In some embodiments, the base editing system described herein comprises an ABE with TadA inserted into a Cas9. Sequences of relevant ABEs with TadA inserted into a Cas9
10 are provided.

101 Cas9 TadAins 1015

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
15 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHAL
 TLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQG
 20 GLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS
 LMDVLHYPGMNRHVEITEGILADECAALCYFFRMPRQVFNAQKKAQSST
 DYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 25 GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGD

30

102 Cas9 TadAins 1022

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESGSEVEFSHE
 YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
 20 IMALRQGGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNA
 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKLKS VKELL
 25 GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

103 Cas9 TadAins 1029

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKSEKQEIIGSSGSETPGTSESATPESGGS
 EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLH
 20 DPTAHAEIMALRQGGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRV
 VFGVRNAKTGAAGSLMDVLHYPGMNRHVEITEGILADECAALLCYFFRMP
 RQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 25 GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

103 Cas9 TadAins 1040

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKAKSEQEIGKATAKYFFYSGSSGSETPGT
 SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLN RVI
 20 GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
 GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNRHVEITEGILADEC
 AALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKLKS VKELL
 25 GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

105 Cas9 TadAins 1068

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANG EIRKRPLIETNGEGSSGSETPGTSESATPESSGSEVEFSHEYWMR
 20 HALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMAL
 RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
 AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ
 SSTDTGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKLKS VKELL
 25 GITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENKRKRL
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

106 Cas9 TadAins 1247

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
 20 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSS
 GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
 VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF
 25 EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
 GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQ
 LGGD

30

107 Cas9 TadAins 1054

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKAKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGSSGSETPGTSESATPESGSEVEFSHEYWMRHALTLAKRARDERE
 20 VPGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
 ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
 HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 25 GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

108 Cas9 TadAins 1026

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 15 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIASEGSSGSETPGTSESATPESGSEVE
 FSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPT
 20 AHAEIMALRQGGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRVVFG
 VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQV
 FNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKLKS VKELL
 25 GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
 LGGD

30

109 Cas9 TadAins 768

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 5 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 10 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQSSGSETPGTSESATPESGSEVEFSHEYWMR
 HALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMAL
 15 RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
 AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRTTQKGQKNSR
 ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
 MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
 20 KHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREI
 NNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQE
 IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
 DFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD P
 KKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERS SFEKNP
 25 IDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIEQISEFSK
 RVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFD
 TTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

30 110.1 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD SFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP

INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 5 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 10 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLT RSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 15 TKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITL KSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLVAKVE
 20 KGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKD LIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPG
 SSGSETPGTSESATPESGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
 VLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV
 TFEPVCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
 25 TEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIEQISEFSK
 RVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFD
 TTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

110.2 Cas9 TadAins 1250

30 MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGLTP

NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 5 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHDD
 10 SLTFKEDIQKAQVSGQGD SLHEHIANLAGSPA I KKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 15 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDIIKLPK
 20 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PG
 SSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
 VGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRLIDAT
 LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
 VEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISE
 25 FSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFK
 YFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

110.3 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 30 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDK
 5 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKV
 10 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIK
 15 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPG
 20 SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDER
 EVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
 DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
 NHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEII EQ
 ISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPA
 25 AFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

110.4 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 30 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEI

FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 5 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA I KKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 10 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEI
 15 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKD LIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PG
 SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
 20 EVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
 DATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
 NHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQ
 ISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPA
 AFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

25

110.5 Cas9 TadAins 1249

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 30 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 5 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 10 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 15 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSGS
 SGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
 VPGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
 20 ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
 HRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEII EQ
 ISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPA
 AFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

25 110.5 Cas9 TadAins delta 59-66 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 30 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ
 5 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 10 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKAKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 15 KGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGKSPG
 SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDERE
 VPGGAVLVNNRVIGEGWNRHAHEIMALRQGGGLVMQNYRLIDATLYVTFE
 PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG
 20 ILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLD
 EIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTN
 LGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGG
 D

25 110.6 Cas9 TadAins 1251

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 30 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKS NFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 5 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNF MQLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 10 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLVAKVE
 15 KGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKD LIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
 GSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
 REVVPVGA VLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGG LVMQNYRL
 IDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
 20 MNHRVEITEGILADECAALLCYFFRMRDNEQKQLFVEQHKHYLDEII EQ
 ISEFSKR VILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPA
 AFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

110.7 Cas9 TadAins 1252

25 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDD SFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 30 NFKS NFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDD
 5 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLT RSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 10 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDIIKLPK
 15 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
 DGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARD
 EREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
 LIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEII EQ
 20 ISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPA
 AFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

110.8 Cas9 TadAins delta 59-66 C-truncate 1250

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 25 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKS NFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 30 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 5 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQEELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 10 YPKLESEFVYGDYKVYDVRKMIKSEKQEI GKATKAYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPG
 15 SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
 VLVLNRRVIGEGWNRHAHEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
 AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
 CAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADA
 NLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
 20 YTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

111.1 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHHR
 25 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 30 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 5 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAISH
 YWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAE
 10 IMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVR
 KMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
 GEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKL
 15 IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM
 ERSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE
 LQKGNELALPSKYVNFYLYLASHYEKLKGSPEQNEQQLFVEQHKHYLDEI
 IEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLG
 APAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
 20

111.2 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 25 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 30 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQ

LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 5 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAISH
 YWMRHALLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAE
 IMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
 10 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDGSSGSSGSETPGTSESATPESGGSSIKKYPKLESEFVYGDY
 KVDYVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
 ETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPK
 RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKEL
 15 LGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKR
 LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQNEQQLFVEQHK
 HYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLF
 TLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL
 QLGGD

20

112 delta HNH TadA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 25 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 30 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTFLFEDREMIEERLKYAHLFDDKVMKQ

LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAICKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
 YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
 5 IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
 KLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
 LKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFK
 10 TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKK
 TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
 KVEKGKSKKLKSVKELLGITIMERSSEKKNPIDFLEAKGYKEVKKDLIIK
 LPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
 SPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLKVLSAYNKH
 15 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTL
 IHQSITGLYETRIDLSQLGGD

113 N-term single TadA helix trunc 165-end

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
 20 LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
 RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
 MPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSV
 GWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
 TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
 25 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRG
 HFLIEGDLNPDNSVDKLFQQLVQTYNQLFEENPINASGVDAKAILSARL
 SKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQL
 SKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
 LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
 30 ASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHL
 GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
 RKSEETITPWNFEVVDKGASQSFIERMNFDKNLPNEKVLPHKSLLYE
 YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
 DYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDIL

EDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKL
 INGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQ
 GDSLHEHIANLAGSPAIAKKGILQTVKVDELVKVMGRHKPENIVIAMARE
 NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
 5 QNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGK
 SDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
 IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDF
 RKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVY
 DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
 10 GETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
 TIMERSSSFENPIDFLEAKGYKEVKKDIIKLPKYSLFELNGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKHYL
 DEIEEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLT
 15 NLGAPAAFYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLG
 GD

114 N-term single TadA helix trunc 165-end delta 59-65

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRTAH
 20 AEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
 NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPSGGS
 SGGSSGSETPGTSESATPSSGGSSGSDKKYSIGLAIGTNSVGWAVITD
 EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
 RRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIV
 25 DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
 LNPDNSTDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLE
 NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDD
 DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
 RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
 30 KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
 RRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETI
 TPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNE
 LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
 CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL

TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
 QSGKTILDFLKS DGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
 IANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
 QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
 5 VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
 EVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
 TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFY
 KVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMIA
 KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
 10 WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
 KDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERS
 FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
 NELALPSKYVNFYLYLASHYEKLKGS PEDNEQKQLFVEQH KHYLDEIIEQI
 SEFSKRVI LADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAA
 15 FKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

115.1 Cas9 TadAins1004

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHHR
 20 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 25 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 30 IKDKDFLDNEENEDILEDIVLTLT LTFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFQMQLIHDD
 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVRREINNYHHAHDAYLNAVVGTAIIK
 YPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
 5 PVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDA
 TLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
 RVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 10 PKKYGGFDSPTVAYSVLVVAKEKGSKKLKSVELLGITIMERSSEFEKN
 PIDFLEAKGYKEVKKDLIIKLPKYSLEFELENGRKRMLASAGELQKGNELA
 LPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS
 KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYF
 DTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD

15

115.2 Cas9 TadAins1005

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
 20 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 25 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 30 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNL

TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKK
 YPKLGSSGSETPGTSESATPESGSEVEFSHEYWMRHALTLAKRARDERE
 VPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
 5 ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
 HRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 PKKYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEFEKN
 10 PIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELA
 LPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS
 KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYF
 DTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD

15 115.3 Cas9 TadAins1006

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 20 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 25 YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDD
 30 SLTFKEDIQKAQVSGQDLSHEHIANLAGSPAIIKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER
 EVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
 DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
 5 NHRVEITEGILADECAALLCYFFRMPRQSEFVYGDKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 PKKYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEFEKN
 PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
 10 LPSKYVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS
 KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYF
 DTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD

115.4 Cas9 TadAins1007

15 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTP
 20 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 25 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKV
 30 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK

YPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
 REVPGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
 IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
 MNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIKSEQ
 5 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 PKKYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEFEKN
 PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
 LPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFS
 10 KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAKFYF
 DTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

116.1 Cas9 TadAins C-term truncate2 792

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 15 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 20 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 25 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDLVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETP
 30 GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
 VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
 CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
 ECAALLCYFFRMPRQS QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
 LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
 TKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRE
 INNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 5 RDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 PKKYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEFEKN
 PIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKRMLASAGELQKGNELA
 LPSKYVNFYLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIEQISEFS
 KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYF
 10 DTTIDRKRYTSTKEVLDTLHQSIITGLYETRIDLSQLGGD

116.2 Cas9 TadAins C-term truncate2 791

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHR
 15 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 20 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 25 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGSSGSETPG
 TSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRV
 30 IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
 AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
 CAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
 LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
 KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

TKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRE
 INNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 5 PKKYGGFDSPTVAYSVLVVAKEKKGKSKKLKSVKELLGITIMERSSEFEKN
 PIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELA
 LPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS
 KRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAKFYF
 DTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

10

116.3 Cas9 TadAins C-term truncate2 790

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSEFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 15 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 20 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKYAHLFDDKVMKQ
 25 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGIKEGSSGSETPGT
 SESATPESGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVI
 GEGWNRAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCA
 30 GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC
 AALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
 LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
 KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
 TKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRE

INNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIKSEQ
 EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
 RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
 PKKYGGFDSPTVAYSVLVVAKEKKGSKKLKSVKELLGITIMERSSEFEKN
 5 PIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELA
 LPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFS
 KRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYF
 DTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD

10 117 Cas9 delta 1017-1069

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 15 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIY
 20 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 25 SLTFKEDIQKAQVSGQGD SLHEHIANLAGSPA IKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLT RSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLI
 30 REVKVITL KSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKK
 YPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
 VLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV
 TFEP CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
 TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVR

KVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
 DSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEA
 KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN
 FLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLAD
 5 ANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK
 RYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD

118 Cas9 TadA-CP116ins 1067

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 25 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLT RSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSEL DKA GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITL KSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEI
 30 TLANGEIRKRPLIETNMNRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR
 DEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
 RLIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
 PGGETGEI VWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR

NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
 GITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKH
 YLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ
 LGGD

119 Cas9 TadAins 701

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKS NFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
 GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL
 25 YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
 EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD LTFKEDIQKAQVS
 GQGDSLHEHIANLAGSPAIKKGILQTVKVDELVKVMGRHKPENIVIEMA
 RENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHPVENTQLQNEKLYLY
 YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
 30 GKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
 GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVS
 DFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYK
 VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR

NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
 GITIMERSSSFENPIDFLEAKGYKEVKKDIIKLPKYSLFELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDL SQ
 LGGD

120 Cas9 TadACP136ins 1248

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA IKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGikelGSQILKEHP
 25 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 30 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
 KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS MN
 HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT

SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
 GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
 GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENI IH LFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ
 LGGD

121 Cas9 TadACP136ins 1052

MDKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKYAH LFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQD SLHEHIANLAGSPA I KKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHP
 25 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSEL D KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEI
 30 TLAMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
 ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
 NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEP
 CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR

NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
 GITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKH
 YLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ
 LGGD

122 Cas9 TadACP136ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS DGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQD SLHEHIANLAGSPA I KKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHP
 25 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSEL D KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKS KLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYS MNHRVEITEG
 30 ILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
 SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAI
 GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRI
 GRVVFGVRNAKTGAAGSLMDVLHYPGNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKR

NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
 GITIMERSSSFENPIDFLEAKGYKEVKKDIIKLPKYSLSFELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGD

123 Cas9 TadACP139ins 1299

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVV DKGASAQS FIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA I KKGILQTVKVVDELVKV
 MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHP
 25 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVG TALIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 30 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
 KGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
 DNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RMN

HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
 SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
 GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
 GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFT
 5 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLQ
 LGGD

124 Cas9 delta 792-872 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 10 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 15 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 20 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSEVEFSHE
 25 YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
 IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDDEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
 GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVS
 30 DFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYK
 VYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML

ASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGD

5

125 Cas9 delta 792-906 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFH R
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
 10 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAIL SARLSKSRRL ENLIAQLPGEKKNGLFGNLI ALSGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 15 KQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI PY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
 NLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ
 20 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDD
 SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGikelGSEVEFSHE
 YWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAE
 IMALRQGGLVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNA
 25 KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
 KKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
 LIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALI
 KKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKT
 EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
 30 EVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVAYSVLVVAK
 VEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKL
 PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
 PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH R
 DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLI

HQSITGLYETRIDLSQLGGD

126 TadA CP65ins 1003

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR
 25 VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
 PRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHA
 LTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPLESEFVYGDYK
 VYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 GITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQL

LGGD

127 TadA CP65ins 1016

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
 25 CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
 ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSE
 VEFSEHYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
 PYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGIEIRKPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQ

LGGD

128 TadA CP65ins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTP
NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
YPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYV
25 TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNLRVIGEGWN
RAIGLHDPKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKH
YLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQ

LGGD

129 TadA CP65ins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKSEQEITAHAEIMALRQGGLVMQNYRL
 25 IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
 MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP
 GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNR
 VIGEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELL
 GITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQ

LGGD

130 TadA CP65ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKSEKQKATAKYFFYSTAHAEIMALR
 25 QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA
 GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQS
 STDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
 PVGAVLVLNNRVIGEGWNRAIGLHDPNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 GITIMERSSEKKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQKQFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRIDLSQ

LGGD

131 TadA CP65ins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTLTLEFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKAKSEQEI GKATAKYFFYSNIMNFFKTEI
 25 TLANTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMHSRIG
 RVVFGVRNAKTGAAGSLMDVLHYPGMNRVEITEGILADECAALLCYFFR
 MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH
 ALTAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKR
 30 NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELL
 GITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRML
 ASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDL SQ

LGGD

132 TadA CP65ins 1246

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENP
 INASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTP
 NFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 10 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDK
 NLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
 15 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
 MGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 20 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
 SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
 REVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK
 YPKLESEFVYGDYKVYDVRKMIKAKSEQEI GKATAKYFFYSNIMNFFKTEI
 25 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVE
 KGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAH
 AEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
 30 NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN
 AQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKR
 ARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKH
 YLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ

LGGD

In some embodiments, adenosine deaminase base editors were generated to insert TadA or variants thereof into the Cas9 polypeptide at the identified positions.

- Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/US2020/016285 and U.S. Provisional Application Nos. 62/852,228 and 62/852,224, the contents of which are incorporated by reference herein in their entireties.

Cas9 Domains with Reduced PAM Specificity

- 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.

30

Nucleobase Editing Domain

Described herein are base editors comprising a fusion protein that includes a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a 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 components of the base editor can then edit a target base.

5 In some embodiments, the nucleobase editing domain includes a deaminase domain. As particularly described herein, the deaminase domain includes a cytosine deaminase or an adenosine deaminase. In some embodiments, the terms “cytosine deaminase” and “cytidine deaminase” can be used interchangeably. 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
10 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), 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
15 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:eaa04774 (2017), the entire contents of which are hereby incorporated by reference.

20

A to G Editing

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)
25 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).

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
30 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

5 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

10 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.

15 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

20 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

25 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, I156F, or

30 a corresponding mutation in another adenosine deaminase.

The adenosine deaminase can be derived from any suitable organism (*e.g.*, *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.

5

Adenosine deaminases

In some embodiments, fusion proteins 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).

In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine. In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine in a deoxyadenosine residue of DNA. 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). 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 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. 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*.

The disclosure provides adenosine deaminase variants that have increased efficiency (>50-60%) and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (*i.e.*, “bystanders”).

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.

In some embodiments, the nucleobase editors of the disclosure are adenosine deaminase variants comprising an alteration in the following sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
5 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (also termed TadA*7.10).

In particular embodiments, the fusion proteins comprise a single (*e.g.*, provided as a monomer) TadA*8 variant. In some embodiments, the TadA*8 is linked to a Cas9 nickase. In some embodiments, the fusion proteins of the disclosure comprise as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA*8 variant. In other embodiments, the fusion
10 proteins of the disclosure comprise as a heterodimer of a TadA*7.10 linked to a TadA*8 variant. In some embodiments, the base editor is ABE8 comprising a TadA*8 variant monomer. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 variant and a TadA(wt). In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 variant and TadA*7.10. In some embodiments, the base editor is
15 ABE8 comprising a heterodimer of a TadA*8 variant. In some embodiments, the TadA*8 variant is selected from **Table 8, 10, 11, or 13**. In some embodiments, the ABE8 is selected from **Table 10, 11 or 13**. The relevant sequences follow:

Wild-type TadA (TadA(wt)) or “the TadA reference sequence”

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
20 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTD

TadA*7.10:

MSEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNRRVI GEGWNRAIGL
HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR
25 VVFGVRNAKT GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM
PRQVFNAQKK AQSSD

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
30 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, 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, 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.

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:

MRRAFITGVFFLSEVEFSHEYWMRHALLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
 HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGA
 AGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAKKKAQSSTD.

It should 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 (ADAT). Without limitation, the amino acid sequences of exemplary AD AT homologs include the following:

Staphylococcus aureus TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIAHNLRETLQQPTAHAEHIA
 IERAAKVLGSRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS
 LMNLLQQSNFNHRAIVDKGVLKEACSTLLTFFKNLRANKKSTN

Bacillus subtilis TadA:

MTQDELYMKEAIKEAKKAEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEA
 CKALGTWRLEGATLYVTLEPCPMCAVAVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNH
 QAEVVSGLVEEECGMLSAFFRELKKKKAARKNLSE

Salmonella typhimurium (*S. typhimurium*) TadA:

MPPAFITGVTSLSDELVDHEYWMRHALLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGR
 HDPTAHAEIMALRQGGVLVQNYRLDITLYVTLEPCVMCAGAMVHSRIGRVVFVGARDAKTGA
 AGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRREQEIKALKKADRAEGAGPAV

Shewanella putrefaciens (*S. putrefaciens*) TadA:

5 MDEYWMQVAMQMAEKAEAGEVPVGAVLVKDGQGIATGYNLSISQHDPTAHAEILCLRSAGK
 KLENYRLLDATLYITILEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQV
 EVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHA
 10 EIIALRNGAKNIQNYRLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHF
 FDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

Caulobacter crescentus (*C. crescentus*) TadA:

MRTDESEDQDHRMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHA
 EIAAMRAAAAKLGNRYRLTDLTLVVTTLEPCAMCAGAI SHARIGRVVFGADDPKGGAVVHGPKF
 15 FAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGNLREGSNDPSAHA
 EMIAIRQAARRSANWRLTGATLYVTLEPCIMCMGAIILARLERVVFGCYDPKGAAGSLYDL
 SADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKAKATPALFIDERKVPPEP

20 An embodiment of *E. Coli* TadA (ecTadA) includes the following:

MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGVLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFVGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

25 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*.

30 In one embodiment, a fusion protein of the disclosure comprises a wild-type TadA
 linked to TadA*7.10, which is linked to Cas9 nickase. In particular embodiments, the fusion
 proteins comprise a single TadA*7.10 domain (e.g., provided as a monomer). In other

embodiments, the ABE7.10 editor comprises TadA*7.10 and TadA(wt), which are capable of forming heterodimers.

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, 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, 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.

It should be appreciated that any of the mutations provided herein (*e.g.*, based on the TadA reference 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). It would be apparent to the skilled artisan that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein. Thus, any of the mutations identified in the TadA reference sequence can be made in other adenosine deaminases (*e.g.*, ecTadA) 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 in the TadA reference sequence or another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises a D108X mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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.

In some embodiments, the adenosine deaminase comprises an A106X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, wild-type TadA or ecTadA).

In some embodiments, the adenosine deaminase comprises a E155X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a D147X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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.

For example, an adenosine deaminase can contain a D108N, a A106V, a E155V, and/or a D147Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA). 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 (*e.g.*, ecTadA): D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V; A106V and D147Y; E155V and D147Y; D108N, A106V, and E155V; D108N, A106V, and D147Y; D108N, E155V, and D147Y; A106V, E155V, and D147Y; and D108N, A106V,

E155V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein can be made in an adenosine deaminase (*e.g.*, ecTadA).

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 mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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, I95L, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H8X, D108X, and/or N127X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

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 mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

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 in TadA reference sequence, or a corresponding mutation or mutations in

another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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, R26X, 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 H8Y, D108N, N127S, D147Y, R152C, and Q154H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R26W, L68Q, D108N, N127S, D147Y, and E155V in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA).

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

Details of A to G nucleobase editing proteins are described in International PCT Application No. PCT/2017/045381 (WO2018/027078) 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.

In some embodiments, the adenosine deaminase comprises one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In

some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a S2X, H8X, I49X, L84X, H123X, N127X, I156X and/or K160X mutation in 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 in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H123X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an I156X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 I156F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA), 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, or seven mutations selected from the group consisting of L84F, A106V, D108N, H123Y, D147Y, E155V, and I156F in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 in TadA reference sequence.

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 in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q and/or A143R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA). 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 (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an E25X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R26X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R107X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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, R107K, R107A, R107N, R107W, R107H, or R107S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A143X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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 in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H36X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an N37X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R51X mutation in 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an S146X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an K157X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an W23X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), 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 in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R152X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*,

ecTadA), 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 R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

- 5 In one embodiment, the adenosine deaminase may comprise the mutations H36L, R51L, L84F, A106V, D108N, H123Y, S146C, 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:
- 10 (A106V_D108N),
 (R107C_D108N),
 (H8Y_D108N_N127S_D147Y_Q154H),
 (H8Y_D108N_N127S_D147Y_E155V),
 (D108N_D147Y_E155V),
 15 (H8Y_D108N_N127S),
 (H8Y_D108N_N127S_D147Y_Q154H),
 (A106V_D108N_D147Y_E155V),
 (D108Q_D147Y_E155V),
 (D108M_D147Y_E155V),
 20 (D108L_D147Y_E155V),
 (D108K_D147Y_E155V),
 (D108I_D147Y_E155V),
 (D108F_D147Y_E155V),
 (A106V_D108N_D147Y),
 25 (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),
 30 (D103A_D104N),
 (G22P_D103A_D104N),
 (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
 35 _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),
 5 (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),
 10 (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),
 15 (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),
 20 (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),
 25 (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),
 30 (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),
 35 (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),
5 (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),
10 (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),
15 (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),
20 (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),
25 (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
30 _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),
35 (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),
5 (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),
10 (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).

15 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
20 a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9).

In some embodiments, the adenosine deaminase is TadA*7.10. In some embodiments, TadA*7.10 comprises at least one alteration. In particular embodiments, TadA*7.10 comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and Q154R. The alteration Y123H is also referred to herein as
25 H123H (the alteration H123Y in TadA*7.10 reverted back to Y123H (wt)). In other embodiments, the TadA*7.10 comprises a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y;
30 Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R. In particular embodiments, an adenosine deaminase variant comprises a deletion of the C terminus beginning at residue 149, 150, 151, 152, 153, 154, 155, 156, and 157, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure is a monomer comprising an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) is a monomer comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H;

Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In one embodiment, an adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

In some embodiments, the TadA*8 is a truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In some embodiments the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

In other embodiments, a base editor of the disclosure is a monomer comprising an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) is a monomer comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a

corresponding mutation in another TadA. In other embodiments, the base editor is a heterodimer comprising a wild-type adenosine deaminase and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in **Table 8**. **Table 8** shows certain amino acid position numbers in the TadA amino acid sequence and the amino acids present in those positions in the TadA-7.10 adenosine deaminase. **Table 8** also shows amino acid changes in TadA variants relative to TadA-7.10 following phage-assisted non-continuous evolution (PANCE) and phage-assisted continuous evolution (PACE), as described in M. Richter *et al.*, 2020, *Nature Biotechnology*, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein. In some embodiments, the TadA*8 is TadA*8a, TadA*8b, TadA*8c, TadA*8d, or TadA*8e. In some embodiments, the TadA*8 is TadA*8e.

Table 8. Additional TadA*8 Variants

TadA	TadA amino acid number									
	26	88	109	111	119	122	147	149	166	167

	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

In one embodiment, a fusion protein of the disclosure comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers. Exemplary sequences follow:

TadA(wt) or “the TadA reference sequence”:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 10 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTD

TadA*7.10:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 15 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

TadA*8:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

20 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
 25 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.

In particular embodiments, a TadA*8 comprises one or more mutations at any of the following positions shown in bold. In other embodiments, a TadA*8 comprises one or more mutations at any of the positions shown with underlining:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG ⁵⁰
 LHDPTAHAEI MALRQGGLVM QNYRLIDATL Y**V**TFEPCVMC AGAMIHSRIG ¹⁰⁰
 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR ¹⁵⁰
 MPRQVFNAQK KAQS**S**T**D**

For example, the TadA*8 comprises alterations at amino acid position 82 and/or 166 (e.g., V82S, T166R) alone or in combination with any one or more of the following Y147T, Y147R, Q154S, Y123H, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In particular embodiments, a combination of alterations is selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the adenosine deaminase is TadA*8, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
 LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR
 MPRQVFNAQK KAQSSTD

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In one embodiment, a fusion protein of the disclosure comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

C to T Editing

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.

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.

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).

10 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 15 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 20 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).

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. 25 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, 30 APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D (“APOBEC3E” now refers to this), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and Activation-induced (cytidine) deaminase. 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). In some embodiments a deaminase incorporated into a base editor comprises all or a portion of cytidine deaminase 1 (CDA1). It should 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.

The amino acid and nucleic acid sequences of PmCDA1 are shown herein below.

>tr|A5H718|A5H718_PETMA Cytosine deaminase OS=Petromyzon marinus OX=7757
PE=2 SV=1 amino acid sequence:
MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSGTE
RGIHAEIIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRNGHTLKIWAC
KLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEK
RRSELSIMIQVKILHTTKSPAV

Nucleic acid sequence: >EF094822.1 Petromyzon marinus isolate PmCDA.21 cytosine deaminase mRNA, complete cds:

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGAATACGTTT
AGAGAGGACATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACC

GACGCTGAGTACGTGAGAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTT
 CAACAACAAAAAATCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTG
 AACGTAGAGCGTGTTTTTGGGGCTATGCTGTGAATAAACCACAGAGCGGGACAGAACGTGGA
 ATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAATACCTGCGCGACAACCCCGGACA
 5 ATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGCTGAAAAGATCTTAG
 AATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAATC
 TATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGG
 GTTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGC
 ACAATCAATTGAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGG
 10 AGCGAGTTGTCCATTATGATTAGGTAATACTCCACACCACTAAGAGTCCTGCTGTTTA
 AGAGGCTATGCGGATGGTTTTTC

The amino acid and nucleic acid sequences of the coding sequence (CDS) of human activation-induced cytidine deaminase (AID) are shown below.

>tr|Q6QJ80|Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens
 15 OX=9606 GN=AICDA PE=2 SV=1 amino acid sequence:
 MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELL
 FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
 AEPEGLRRLHRAGVQIAIMTFKAPV

The amino acid and nucleic acid sequences of the coding sequence (CDS) of human activation-induced cytidine deaminase (AID) are shown below.

>tr|Q6QJ80|Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens
 OX=9606 GN=AICDA PE=2 SV=1 amino acid sequence:
 MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKGCHVELL
 FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK
 25 AEPEGLRRLHRAGVQIAIMTFKAPV

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

AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAG
 AACTCTGGACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGC
 30 CTTCTCTCAGAGCAAATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAAGT
 TCTGACTGATAAGATCAGCTTGATCAATATGCATATATATTTTTTGATCTGTCTCCTTTTCT
 TCTATTAGATCTTATACGCTGTCAGCCCAATTCTTTCTGTTTCAGACTTCTCTTGATTTC
 CTCTTTTTTCATGTGGCAAAGAAGTAGTGCGTACAATGTACTGATTCGTCCTGAGATTTGTA
 CCATGGTTGAACTAATTTATGGTAATAATATTAACATAGCAAATCTTTAGAGACTCAAATC

ATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAATTTTGT
AAATATTCAACAGTAAAACAACCTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAAT
TTAGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGA
AAGTCACTATGATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTT
5 AAAGGCCCAATTGTTAGGCAGTTAATGGCACTTTTACTATTAATAATCTTTCATTTGTTT
AGACGTAGCTTAACTTACCTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATG
TGCAGTTTTTGTAGGTTATTGTCATAGAACTTATTCTATTCCTACATTTATGATTACTATG
GATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAATTTAACTCCTTTATAAAGAAC
TTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACAGGGTCTTAGCCC
10 AGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAAGTGC
TGGAATTATAGACATGAGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATT
TAATGTTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATGTCTCCTCAGTTTACACT
GAGATTTTGAAAACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTGT
CAAAGTAAAATGGAAAGCAAAGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAGG
15 AGAAAAGATGAAATTC AACAGGACAGAAGGGAAATATATTATCATTAAGGAGGACAGTATCT
GTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCAGGATTATTTTTAACCCGCTTGTTTCT
GGTTTGCACGGCTGGGGATGCAGCTAGGGTCTGCCTCAGGGAGCACAGCTGTCCAGAGCAG
CTGTCTAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTACTCAGGACAGAAATG
ACGAGAACAGGGAGCTGGAAACAGGCCCTAACCAGAGAAGGGAAGTAATGGATCAACAAAG
20 TTAAGTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAAC
AGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTAT
CTATGCCACATCCTTCTTATCTATACTTCCAGGACACTTTTTCTCCTTATGATAAGGCTCT
CTCTCTCTCCACACACACACACACACACACACACACACACACACACACACACAAACACAC
ACCCCGCCAACCAAGGTGCATGTAAAAAGATGTAGATTCTCTGCCTTTCTCATCTACACAG
25 CCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGTAAGAGATGATGCTTAATCTGTTTA
ACACTGGGCCTCAAAGAGAGAATTTCTTTCTTCTGTACTTATTAAGCACCTATTATGTGTT
GAGCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGGTTGGTACT
ATGGTAATTACCATAAAAATTATTATCCTTTTAAATAAAGCTAATTATTATTGGATCTTTT
TTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCTGTAC
30 CCAGGCTGGAGTGCAGTGGTGCAATCATAGCTTCTGCAGTCTTGAACCTCTGGGCTCAAGC
AATCCTCCTGCCTTGGCCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCT
AGGATCCATTTAGATTAAAATATGCATTTTAAATTTTAAAATAATATGGCTAATTTTTACCT
TATGTAATGTGTATACTGGCAATAAATCTAGTTTGCTGCCTAAAGTTTAAAGTGCTTTCCAG
TAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAACAGACAGCCAGGTGTGGTGGCTC

ACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTC
AAGACCAGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGGCATGGT
GGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGA
GGTCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGA
5 CCTTGCCTCAAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACACAAAGAGCTGT
TGTCCCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTC
TGTACCTGCACTACATTATTAAATATCAATTCTCAATGTATATCCACACAAAGACTGGTA
CGTGAATGTTCATAGTACCTTTATTACAAAACCCCAAAGTAGAGACTATCCAAATATCCAT
CAACAAGTGAACAAATAAACAAAATGTGCTATATCCATGCAATGGAATACCACCCTGCAGTA
10 CAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAAGAGTCAGACATGA
AGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTACAGAA
AGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGGCCACGTG
GGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCT
CCACAAAATGGGAAAAAAGAAAGCAAATCAGTGGTTGTCTGTGGGGAGGGGAAGGACTG
15 CAAAGAGGGAAGAAGCTCTGGTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTG
GTAGCAGTTTGGGGTGTTTACATCCAAAATATTCGTAGAATTATGCATCTTAAATGGGTGG
AGTTTACTGTATGTAAATTATACCTCAATGTAAGAAAAATAATGTGTAAGAAAACCTTCAA
TTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCAAATTCTCTGCACT
TCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGCATTTCTGGAA
20 AAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTACAGCTTG
TGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAACTTGGGTTACCAGA
GTATTTCCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCC
AGAAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGAT
CATTTTGACTAGTTAAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAA
25 ATCATGTAAATATGTGCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACAGTCAT
TGTCATTAGACACTAAGTCTAATTATTATTATTAGACACTATGATATTTGAGATTTAAAAAA
TCTTTAATATTTTAAATTTAGAGCTCTTCTATTTTTTCCATAGTATTCAAGTTTGACAATGA
TCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTGGTCTTG
TTGCCCATGCTGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTC
30 AAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCACCACCACACTC
GGCTAATGTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAA
CTCCTGACCTCAGAGGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGG
CCACTGCGCCCGGCCAAGTATTGCTCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCC
AGCCAGGTATTGCTCTTATACATTAAAAAATAGGCCGGTGCAAGTGGCTCACGCCTGTAATCC

CAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGGTCAGGAGTCCAAGGCCAGCCTGG
CCAAGATGGTGAACCCCGTCTCTATTAAAAATACAAACATTACCTGGGCATGATGGTGGGC
GCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCAGATCTG
CCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGG
5 CGACAAAGTGAGACCGTAACAAAAAAAAAAAAATTTAAAAAAGAAATTTAGATCAAGATCC
AACTGTAAAAAGTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAGGCAGAAG
AGAACCATCAGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGGTTTTTGTGAGATCAT
GGTGGTGACAGTGTGGGGAATGTTATTTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAG
GCCAGCACAAACAGATAAGGAGGAAGAAGATGAGGGCTTGGACCGAAGCAGAGAAGAGCAAAC
10 AGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCAACACATTTAGATGATTAATT
AAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTTCCAGGCTGCTAGGCTGCTTA
CCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTTGAGGAATATTGT
TTTGATCATTTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAGACTGGAGGGGAAATCT
GAATATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTCGTTTTCTTGTTG
15 AAGAACAAATTTAATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGAC
TGTCTTGTGGGTAAGGGTTTGGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAG
CAGGAAAAGGAGTTTATGATGGATTCCAGGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCA
GCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCCAAGTAATGACTTCCTTAAAAAGCTGA
AGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATTTTCTCTCCCTCCTCTCAC
20 CCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGGGCTA
AGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCTTT
TCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTT
AATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTG
GCATTTGTGTCTCTATCACATTCCCTCAATCCTTTTTTTTTATTTCTTTTTCCATGTCCATGC
25 ACCCATATTAGACATGGCCCAAATATGTGATTTAATTCTCCCCAGTAATGCTGGGCACCC
TAATACCACTCCTTCCTCAGTGCCAAGAACAAGTCTCCCAAAGTGTACCAGCTTTCCT
CAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAAGCATTTTTATATGGGAGAATATTA
TCAGCTTGTCCAAGCAAAAATTTAAATGTGAAAAACAAATTGTGTCTTAAGCATTTTTTGAA
AATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCCAAATGATT
30 TCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCA
GAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAA
TTACATTTCCAAGAGACTTCTTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACG
GGTATCTCCTCTCTCCTAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATC
CGTGGGGTGGAAGGTCATCGTCTGGCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCT

TTGCCTACATTTGTATTGAATACATCCCAATCTCCTTCCTATTTCGGTGACATGACACATTCT
ATTTTCAGAAGGCTTTGATTTTATCAAGCACTTTCATTTACTTCTCATGGCAGTGCCTATTAC
TTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTATTTCCCCTTTTCAGATCCTCCC
AAATGGTCCTCATAAACTGTCTGCCTCCACCTAGTGGTCCAGGTATATTTCCACAATGTTA
5 CATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACA
CAAATTAAATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCT
TCCTCATTTCCACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTT
CAGCTCTACCTACTGGTGTGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGAC
AATAGCTGCAAGCATCCCCAAAGATCATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGC
10 AATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTCTGTCTCTCCAGAACGGCTGCCACGTGGA
ATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCCGCTGCTACCGCGTCA
CCTGGTTACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCTGCGA
GGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAA
GGCTGAGCCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCT
15 TCAAAGGTGCGAAAGGGCCTTCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGGGATTGCGA
TGCGGAATGAATGAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCA
CCTCTGGAGCCGAAATTAAAGATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGC
CCCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTTCTTTCCCCTCGATTTGGAACCTGAACT
GTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTTTTTTTTTTTTTTGAAGATTATTTTTACT
20 GCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCCTGGGAAGGGCTGCATGAA
AATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGTAAGGGGCTTCCTCGCTTT
TTAAATTTTCTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTCTT
ATTGTTCAATCACTCTCAGTTTTTCATCTGATGAAAACCTTATTTCTCCTCCACATCAGCTTT
TTCTTCTGCTGTTTACCATTTCAGAGCCCTCTGCTAAGGTTCCCTTTCCCTCCCTTTTCTTT
25 CTTTTGTTGTTTCACATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTC
AGAATTCTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAACC
CAAAAAACTCTTCCCAATTTACTTTCTTCCAACATGTTACAAAGCCATCCACTCAGTTTA
GAAGACTCTCCGGCCCCACCGACCCCCAACCTCGTTTTGAAGCCATTCACTCAATTTGCTTC
TCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGACTTTGGG
30 ACTTTGATAGCAACTTCAGGAATGTCACACACGATGAAATATCTCTGCTGAAGACAGTGGA
TAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTT
ACAGAAAAATATTTATATACGACTCTTTAAAAGATCTATGTCTTGAAAATAGAGAAGGAA
CACAGGTCTGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTAC
TGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGCAACGTTTTTCTATGACTT

TTAGGTAGGATGAGAGCAGAAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAATGTTTTT
 ATATCAACATCCTTTATTATTTGATTCATTTGAGTTAACAGTGGTGTTAGTGATAGATTTTT
 CTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAACTCTTCCATCAGGCCATGATCT
 ATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAACCATCTCTCCAAAGCATT
 5 AATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAA
 GAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCTACTTTA
 ATAAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTGATGT
 CTGAAGTAGCAAATCTTCTGGAAACGCAAACCTTTTTAAGGAAGTCCCTAATTTAGAAACAC
 CCACAACTTTCACATATCATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGA
 10 GAGGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTT
 TGCTACATTTTGTATGTGTGTGATGCTTCTCCCAAAGGTATATTAAGTATATAAGAGAGTTG
 TGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGG
 AGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAA
 GATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTGGCTC
 15 ACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGTTT
 GAGACCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATTAGCCAGG
 CGTGGTAGCAGGCACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAA
 CCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAA
 GAGCAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAAAGAGAACAATAT
 20 TTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAATTGTGCTTTATCCAACAAAATGTAAGG
 AGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGTCCCTAACAACTGTCTTTG
 ACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGC
 AATGAAGATGAGCAGATCCACAGGAAAACCTGAATGCACAACTGTCTTATTTTAATCTTATT
 GTACATAAGTTTGTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATT
 25 ATTTTGCCTCTAATGATTTTTTATTAACATGATTTCTTTTCTGATATATTGAAATGGAGTC
 TCAAAGCTTCATAAATTTATACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAAC
 ATTGCAGTAATGGTGCTACGAAGCCATTTCTCTTGATTTTTAGTAACTTTTATGACAGCAA
 ATTTGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAA
 GATAAATAACCAAATAAATAATATAAAAAGTGATTTATATGAAGTTAAATAAAAAATCAGT
 30 ATGATGGAATAAACTTG

Other exemplary deaminases that can be fused to Cas9 according to aspects of this disclosure are provided below. In embodiments, the deaminases are activation-induced deaminases (AID). 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:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFL
 5 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPE
 GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEV
DDLRLDAFRTLGL (underline: nuclear localization sequence; double underline: nuclear
 export signal)

Mouse AID:

10 MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFL
 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNSLRIFTARLYFCEDRKAEPE
 GLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEV
DDLRLDAFRMLGF (underline: nuclear localization sequence; double underline: nuclear
 export signal)

Canine AID:

15 MDSLLMKQRKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFL
 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPE
 GLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEV
DDLRLDAFRTLGL (underline: nuclear localization sequence; double underline: nuclear
 20 export signal)

Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETLYCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFL
 RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEP
 EGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYE
 25 VDDLRLDAFRTLGL (underline: nuclear localization sequence; double underline: nuclear
 export signal)

Rat AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQR
 KFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGYLRNKGCHVELLFLRYISDWDL
 30 PGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLTGWGALPAGLMSPARPSDYF
 YCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRLDAFRTLGL
 (underline: nuclear localization sequence; double underline: nuclear export signal)

clAID (Canis lupus familiaris):

MDSLLMKQRKFLYHFKNVRWAKGRHETYL CYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDW
 DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEP EGLRRLHRAGVQI
 AIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDL RDAFRTLGL

btAID (*Bos Taurus*):

5 MDSLLKKQRQFLYQFKNVRWAKGRHETYL CYVVKRRDSPTSFS LDFGHLRNKAGCHVELLFLRYISDW
 DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC DKERKAEP EGLRRLHRAGVQ
 IAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDL RDAFRTLGL

mAID (*Mus musculus*):

MDSLLMNRRKFLYQFKNVRWAKGRRETYL CYVVKRRDSATSFSLDFGYLRNKN GCHVELLFLRYISDW
 10 DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCEDRKAEP EGLRRLHRAGVQI
 AIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDL RDAFRTLGL

rAPOBEC-1 (*Rattus norvegicus*):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
 KFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS
 15 SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP RYPHLWVRLYVLELYCII LGLPPCLNILRRKQPQLT
 FFTIALQSCHYQRLPPHILWATGLK

maAPOBEC-1 (*Mesocricetus auratus*):

MSSETGPVVVDPTLRRRIEPHEFDAFFDQGE LRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIE
 KFTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVT LFIYAARLYHHTDQRNRQGLRDLIS
 20 RGV TIRIMTEQEYCYCWRNFVNYPSPNEVYWP RYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLT
 FFRLNLQSCHYQRIPPHILWATGFI

ppAPOBEC-1 (*Pongo pygmaeus*):

MTSEKGPSTGDPTLRRRIESWEFDVIFYDPREL RKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIE
 KFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVT LVIYVARLFWHMDQRNRQGLRDLVN
 25 SGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQ YPPLWMLYALELHCCI LSLPPCLKISRWRQNH LA
 FFRLHLQNCHYQTIPPHILLATGLIHPSVTWR

ocAPOBEC1 (*Oryctolagus cuniculus*):

MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELR KEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFLE
 KLTSEGRGLPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVT LIIFVARLFQHMDRRNRQGLKDLVT
 30 SGVTVRVMSVSEYCYCWENFVNYPGKAAQWP RYPPRWMLYALELYCII LGLPPCLKISR RHQKQLT
 FFSLTPQYCHYKMIPPYILLATGLLQPSVPWR

mdAPOBEC-1 (*Monodelphis domestica*):

MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELR KETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEK
 FTAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVT LAIFISRLYWHMDQQHRQGLKELVHS
 35 GVTIQIMSYSEYHYCWRNFVDYPQGEEDYWP KYPYLWIMLYVLELHCCI LGLPPCLKISGSHSNQLAL
 FSLDLQDCHYQKIPYNVLVATGLVQPFVTWR

ppAPOBEC-2 (*Pongo pygmaeus*):

MAQKEEAAAAATEAASQNGEDLENLDDPEKLELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT
 FLCYVVEAQKGGQVQASRGYLEDEHAAAHAEAEFFNTILPAFDPALRYNVTWYVSSSPCAACADRII
 KTLSTKTNLRLLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQP
 5 WEDIQENFLYYEEKLADILK

btAPOBEC-2 (*Bos Taurus*):

MAQKEEAAAAAEPAASQNGEEVENLEDPEKLELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT
 FLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIV
 KTLNKTKNLRLLILVGRLFMWEEPEIQAAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEP
 10 WEDIQENFLYYEEKLADILK

mAPOBEC-3-(1) (*Mus musculus*):

MQPQRLGPRAGMGPFCGLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPV
 SLHHGVFKNKDNIAEICFLYWFDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSD
 IFSSRLYNVQDPETQQNLRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKL
 15 QEILRPCYISVPSSSSSTLSNICLTGKLPETRFWVEGRMDPLSEEEFYQFYNQVRVKHLCYYHRMKP
 YLCYQLEQFNGQAPLKGCLLSEKKGQHAELFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRD
 RPDILILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFEVNPKRPFWPWKGLEIISR
 RTQRRRLRIKESWGLQDLVNDGFLNLQLGPPMS

Mouse APOBEC-3-(2):

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNK
 NIAEICFLYWFDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSDIFSSRLYNVQD
 PETQQNLRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPV
 PSSSSSTLSNICLTGKLPETRFVVEGRMDPLSEEEFYQFYNQVRVKHLCYYHRMKPYLCYQLEQFNG
 QAPLKGCLLSEKKGQHAELFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDILILHIYTS
 25 RLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFEVNPKRPFWPWKGLEIISRRTQRRRLRIKE
 SWGLQDLVNDGFLNLQLGPPMS (*italic: nucleic acid editing domain*)

Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNK
 DNIHAEICFLYWFDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSDIFSSRLYNIR
 30 DPENQQNLRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIP
 VPSSSSSTLSNICLTGKLPETRFVVERRRVHLLSEEEFYQFYNQVRVKHLCYYHGVKPYLCYQLEQFN
 GQAPLKGCLLSEKKGQHAELFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDILILHIYT
 SRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFEVNPKRPFWPWKGLEIISRRTQRRLRHRIK
 ESWGLQDLVNDGFLNLQLGPPMS (*italic: nucleic acid editing domain*)

35 **hAPOBEC-3A (*Homo sapiens*):**

MEASPASGPRHLMDPHI FT SNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
 RHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLY
 KEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFPQWDGLDEHSQALSGRLRAILQNQGN

hAPOBEC-3F (*Homo sapiens*):

5 MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEM
 CFLSWFCGNQLPAYKCFQITWVFSWTPCPDCVAKLAEFLAHPNVTLTISAARLYYYWERDYRRALCR
 LSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPEAMYPHIFYHFH
 KNLRKAYGRNESWLCFTMEVVKHHSVPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVT
 WYTSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYCW
 10 ENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE

Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSFNFNRPILSGLNTVWLCCEVKT KDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFH
 KWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRG
 GPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYYTLQATLGELLRHLMDPGTFTSNFNKPKW
 15 VSGQHETLYCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPGKGRHAELCFDLIPFWKLDGQQYRVT
 CFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGRLALHRDGAKIAMMNYSEFEYCWD
 TFVDRQGRPFQPWDGLDEHSQALSGRLRAI (*italic: nucleic acid editing domain; underline:*

cytoplasmic localization signal)

Chimpanzee APOBEC-3G:

20 MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPRPLDAKIFRGQVYSKLKYHPEM
RFFHWFSKWRQLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
 SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPPWNNLPKYYILLHIMLGEILRHSMDPPTFTS
 NFNNELWVRGRHETLYCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFDLVIPFWKLD
 LHQDYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRQCQEGRLTLAKAGAKISIMTY
 25 SEFKHCWDTFVDHQGCPFPQWDGLEEHSQALSGRLRAILQNQGN

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

Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQ GKLYPEAKDHPEM
KFLHWFFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALR
 30 ILCQERGGPHATMKIMNYNEFQHCWNEFVDGQKPFKPRKNLPKHYYTLHATLGELLRHVMDPGTFTS
 NFNKPKWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFDLIPFWKLD
 DQQYRVTCFTSWSPCFSCAQKMAKFISSNNKHVSLCIFAARIYDDQGRQCQEGRLTLHRDGAKIAVMNYS
 EFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

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

35 **Human APOBEC-3G:**

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEM
 RFFHWF~~SKWRKLHRDQEYEV~~TWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
 SLCQKRDGPRATMKIMNYDEFQHCWSKFKVYSQRELFEFWNNLPKYIILLHIMLGEILRHSM DPPTFTF
 NFNNEPWVRGRHETYLCEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAE~~LCFLD~~VIPFWKLD
 5 ~~LDQDYRVTCFTSWSPCFSCA~~QEMAKFISK~~NKHVSLCI~~FTARIYDDQGR~~CQEGLRT~~LAEAGAKISIMTY
 SEFKHCWDTFVDHQGCPFPQPDGLDEHSQDL~~SGRLRAILQN~~QEN

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

Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEM
 10 ~~CFLSWFCGNQLPAYKCFQITWFVSWTPC~~PDCVAKLAEFLAHPNVTLTISAARLYYYWERDYRRALCR
 LSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPM~~EAMYPHIF~~YFHF
 KNLRKAYGRNESWLCFTMEVVKHHSVPVSWKRGVFRNQVDPETH~~CHAERCFLSWFCDDILSPNTNYEVT~~
 WYTSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYCW
 ENFVYNDDPEFPKPKGLKYNFLFLDSKLQEILE
 15 (italic: nucleic acid editing domain)

Human APOBEC-3B:

MNPQIRNP~~MYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAE~~
 MCFLSWFCGNQLPAYKCFQITWFVSWTPCPCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALC
 RLSQAGARVTIMDYEEFAYCWENFVYN~~EGQQFMPWYKF~~DENYAFLHRTLKEILRYLMDPDTFTFNFN
 20 DPLVLRRTQTYLCYEVERLDNGTWVLMQHMGLCNEAKNLLCGFYGRHAE~~LRFLDLVPSLQLDPAQI~~
 YRVTFWISWSPCF~~SWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQS~~IMTYDE
 FEYCWDTFVYRQGC~~FPQPDGLEEHSQALS~~SGRLRAILQNQGN

(italic: nucleic acid editing domain)

Rat APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNP~~LKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCA~~
 LPVPLRQGVFRKQGHIAELCFIYWFHDKVLRVLSPEEFKVTWYMSWSPCSKCAEQVARFLAAHRNL
 SLAIFSSRLYYL~~RNPNYQQKLCRLIQEGVHVAAMDLP~~EFKKCW~~NKFVDNDGQF~~FRP~~WMRLRINFSFY~~
 DCKLQEIFSRM~~NLLREDVFYLQFNNSHRVKPVQNR~~YRRKSYLCYQLERANGQEPLKGYLLYKKGEQH
 VEILFLEKM~~SMELSQVRITCYLTWSPC~~PNCARQLA~~AFKKDHPDLILRIYTSR~~LYFWRKKFQKGLCTL
 30 WRS~~GIHVDVMDLPQFAD~~CW~~TNFVN~~PQRPFRPWNELEKNSWRIQRRLRRIKESWGL

Bovine APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQACVWTPGTRNTM~~NLLREVL~~FKQQFGNQPRVPAP
 YYRRKTYLCYQLKQRNDLTLD~~RGCFRNKKQRHAERFIDKINS~~LDLNP~~SQSYKIICYITWSPC~~PNCANE
 LVNFITRNHLKLEIFASRLYFHWIKSFKMGLQDLQ~~AGISVAVMTHTEFEDCWEQFVDNQSR~~FPQPW
 35 DKLEQYSASIRRRRLQRILTAPI

Chimpanzee APOBEC-3B:

MNPQIRNPMEMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAE
 MCFLSWFCGNQLSAYKCFQITWVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALC
 RLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWKFDNNYAFLHRTLKEIIRHLMDDPTFTFNENN
 DPLVLRRHQTYLCYEVERLDNGTWVLMQHMGLFCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQI
 5 YRVTFWFISSPCFSWGCAGQVRAFLQENTHVRRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
 FEYCWDTFVYRQGCFFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPQSPGPCLPLCSEP
 PLGSLPTGRPAPSLPFLLTASFSPPPASLPPLPSLSLSPGHLVPVPSFHSLSLSCSIQPPCSSRIRET
 EGWASVSKEGRDLG

Human APOBEC-3C:

10 MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAE
RCFLSWFCDDILSPNTKYQVTWYTSWSPCPCAGEVAEFLARHSNVNLTIFTARLYYFYQPCYQEGLR
 SLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPKWGLKTNFRLLKRRLRESLQ

(italic: nucleic acid editing domain)

Gorilla APOBEC-3C

15 MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAE
RCFLSWECDDILSPNTNYQVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFQDQDYQEGLR
 SLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPKWGLKYNFRFLKRRLQEILE

(italic: nucleic acid editing domain)

Human APOBEC-3A:

20 MEASPASGRPHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
RHAELRFLDLVPSLQLDPAQIYRVTFWFISSPCFSWGCAGEVRAFLQENTHVRRLRIFAARIYDYDPLY
 KEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFPQWDGLDEHSQALSGRLRAILQNQGN

(italic: nucleic acid editing domain)

Rhesus macaque APOBEC-3A:

25 MDGSPASRPRHLMPTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCG
DYGCHVELRFLCEVPSWQLDPAQTYRVTFWFISSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYD
 PLYQEALRTLRLDAGAQVSIMTYEEFKHCWDTFVDQRGRPFQPWDGLDEHSQALSGRLRAILQNQGN

(italic: nucleic acid editing domain)

Bovine APOBEC-3A:

30 MDEYTFTEFNFNQGWPSKTYLCYEMERLDGDATIPLEDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSW
NLDRNQHYRLTCFISWSPCYDCAQKLTTFLEKHHISLHILASRIYTHNRFCHQSGLCCELQAAGARI
 TIMTFEDFKHCWETFVDHKGKPFQPWEGLVKSQALCTELQAILKTQQN

(italic: nucleic acid editing domain)

Human APOBEC-3H:

MALLTAETFRLQFNNKRRRLRRPYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGL
 DETQCYQVTCYLTWSPCSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVM
 GFPKFADCWENFVDHEKPLSFNPYKMLEELDKNRAIKRRLERIKIPGVRAQGRYMDILCDAEV

(italic: nucleic acid editing domain)

5 Rhesus macaque APOBEC-3H:

MALLTAKTFSLQFNNKRRVKNPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGL
 DETQCYQVTCYLTWSPCPCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVM
 GLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNQAIRRRLERIKSRSDVLENGLRSLQLGPVTPSS
 SIRNSR

10 Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHR
 QEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWVSVNPNCLPCVVKVTKFLAEHPNVTTLTISAARLY
 YYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDNYASLHRTLKEILRNP
 MEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHSAVFRKRGVFRNQVDPETHCHAERCFLSWFC
 15 DDILSPNTNYEVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGAS
 VKIMGYKDFVSCWKNFVYSDDPEFPKPKGLQTNFRLLKRRRLREILQ

(italic: nucleic acid editing domain)

Human APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFDYDPRELKREACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIE
 20 KFTSERDFHPSMSCSITWFLSWSPCECSQAIREFLSRHPGVTLVIIYVARLFWHMDQQNRQGLRDLVN
 SGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMMLYALELHCCIILSLPPCLKISRWRQNHLT
 FFRLLHLQNCHYQTIPPHILLATGLIHPSVAWR

Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSSVWRHTSQNTSNHVEVNFLE
 25 KFTTERRYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLIS
 SGVTIQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHLWVKLYVLELYCIIILGLPPCLKILRRKQPQLT
 FFTITLQTCHYQRIPPHLLWATGLK

Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSSVWRHTSQNTNKHVEVNFIE
 30 KFTTERRYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS
 SGVTIQIMTEQESGYCWRNFVNYPSPNEAHWPYPHLLWVRLYVLELYCIIILGLPPCLNILRRKQPQLT
 FFTIALQSCHYQRLPPHILWATGLK

Human APOBEC-2:

MAQKEEA AVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT
 35 FLCYVVEAQGGQVQASRGYLEDEHAAAHAEAAFFNTILPAFDPALRYNVTWYVSSSPCAACADRII

KTLSKTKNLRLLILVGRLFMWEEPEIQAAKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQP
WEDIQENFLYYEEKLADILK

Mouse APOBEC-2:

MAQKEEAAEAAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
5 FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK

Rat APOBEC-2:

MAQKEEAAEAAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
10 FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK

Bovine APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT
15 FLCYVVEAQSKGGQVQASRGYLEDEHATNHAAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIV
KTLNKTKNLRLLILVGRLFMWEEPEIQAAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK

***Petromyzon marinus* CDA1 (pmCDA1):**

MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
20 IFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQI
GLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTK
SPAV

Human APOBEC3G D316R D317R:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEM
25 RFFHWFWSKWRKLHRDQEYEV TWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYIILLHFMLGEILRHSMDPPTFTFN
FNNEPWVRGRHETYLCEYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDL
DQDYRVTCFTSWSPCFSCAQEMAKFISKHVS LCIFTARIYRRQGRQCQGLRTLAEAGAKISFTYSEF
KHCWDTFVDHQGCPFPQPDGLDEHSQDLSGRLRAILQNQEN

Human APOBEC3G chain A:

MDPPTFTFNFNNEPWVRGRHETYLCEYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKHVS LCIFTARIYDDQGRQCQGLRTLAEAGA
KISFTYSEFKHCWDTFVDHQGCPFPQPDGLDEHSQDLSGRLRAILQ

Human APOBEC3G chain A D120R D121R:

MDPPTFTFNENPEWVRGRHETYLCEVERMHNDTWLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD
 VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNHVSLCIFTARIYRRQGRCQEGRLTLAEAG
 AKISFMTYSEFKHCWDTFVDHQGCPFPQPDGLDEHSQDLSGRLRAILQ

hAPOBEC-4 (*Homo sapiens*):

5 MEPIYEEYLANHGTVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTF
 YELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLD SAIYNND SIRHIILYSNNSPCNEANHCCIS
 KMYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVLSPISSGGIWHSVLHSFISG
 VSGSHVFQPILTGRALADRHNAEINAITGVKPYFTDVLQTKRNPNTKAQEALESYPLNNAFPQGQFF
 QMPSGQLQPNLPPDLRAPVVFVLPLRDLPPMHMGQNPKNPRNIVRHLNMPQMSFQETKDLGRLPTGR
 10 SVEIVEITEQFASSKEADEKKKKKGKK

mAPOBEC-4 (*Mus musculus*):

MDSLLMQKKFLYHFKNVRWAKGRHETYLCEVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDW
 DLDPGRCYRVTFWTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQI
 GIMTFKDYFYCWNTFVENRERTFKAWEGHENSVRRLTRQLRRILLPLYEVDDLDAFRMLGF

15 **rAPOBEC-4 (*Rattus norvegicus*):**

MEPLYEEYLTSGTIVKPYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFY
 ELRSSGNLIQKGLASNCTGSHTHPESMLFERDGYLD SLIFHDSNIRHIILYSNNSPCDEANHCCISK
 MYNFLMNYPEVTLVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLAISGGIWQSILETFVSGI
 SEGLTAVRPFTAGRTLTDRYNAYEINCITEVKPYFTDALHSWQKENQDQKVAASENQPLHNTPAQW
 20 QPDMSQDCRTPAVFMLVPYRDLPIHVNPSQKPRTVVRHLNLTQLSASKVKALRKSPSGRPVKKEEA
 RKGSTRSQEANETNKSQKWKQTLFIKSNICHLLEREQKKIGILSSWSV

mfAPOBEC-4 (*Macaca fascicularis*):

MEPTYEEYLANHGTVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTPQTKHLTF
 YELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLD SAIYNND SIRHIILYCNNSPCNEANHCCIS
 25 KVYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVLSPISSGGIWHSVLHSFVSG
 VSGSHVFQPILTGRALTDRYNAYEINAITGVKPFPTDVLHTRKRNPNNTKAQMALESYPLNNAFPQGSF
 QMTSGIPPDLRAPVVFVLLPLRDLPPMHMGQDPNKPNNIIRHLNMPQMSFQETKDLERLPTRRSVETV
 EITERFASSKQAEETKKKKGKK

pmCDA-1 (*Petromyzon marinus*):

30 MAGYECVRVSEKLDFTFEFQFENLHYATERHRTYVIFDVKPSAGGRSRRLWGYIINNPNVCHAEI
 LMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPLWKNLLEEQGHITLTMHFSRIYDRDREGDHRGL
 RGLKHVSNSFRMGVVGRAEVKECLA EYVEASRRTLWLDTTESMAAKMRKLCILVRCAGMRESGIP
 LHLFTLQTPLLSGRVWWRV

pmCDA-2 (*Petromyzon marinus*):

35 MELREVVDALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIH
 AEVLLLSAVRAALLRRRRRCEDGEEATRGTLCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYEL

DVNRRRSEAEGLRSLRSLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDM
 TGFGEEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDPKFPFLAEFLAQTSVEPSGTPRETRGRP
 RGASSRGPEIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP

pmCDA-5 (*Petromyzon marinus*):

5 MAGDENVRVSEKLDFTFEFQFENLHYATERHRTYVIFDVKPKSAGGRSRRLWGYIINNPNVCHAEI
 LMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPLWKNLLEEQGHITLMMHFSRIYDRDREGDHRGL
 RGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLWLDTTESMAAKMRRKLFILVRCAGMRESGMP
 LHFT

yCD (*Saccharomyces cerevisiae*):

10 MVTGGMASKWDQKGMEDIAEYEEAALGYKEGGVPIGGCLINNKGSVLGRGHNMRFQKGSATLHGEISTL
 ENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVGENVNFKSKGEKYLQTRGHEVVVDDER
 CKKIMKQFIDERPDQDFEDIGE

rAPOBEC-1 (delta 177-186):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
 15 KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFYIYIARLYHHADPNRQGLRDLIS
 SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRGLPPCLNILRRKQPQLTFFTIALQSCHY
 QRLPPHILWATGLK

rAPOBEC-1 (delta 202-213):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
 20 KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFYIYIARLYHHADPNRQGLRDLIS
 SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYVLELYCIILGLPPCLNILRRKQPQHY
 QRLPPHILWATGLK

Mouse APOBEC-3:

MGPFCCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHG
 25 VFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSL
 DIFSSRLYNVQDPETQQNLCLRVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLTNF
 RYQDSKLQEILRPCYIPVPSSSSSTLSNICLTGKLPETRFCVEGRMDPLSEEEFYFYSQFYNQ
 RVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAELFLDKIRSMELSQVTITCY
 LTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQF
 30 TDCWTNFVNPKRPFWPWKGLEIISRRTQRRRLRIKESWGLQDLVNDFGNLQLGPPMS

(italic: nucleic acid editing domain)

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
 35 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.

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.

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.

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.

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.

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.

Details of C to T nucleobase editing proteins are described in International PCT Application No. PCT/US2016/058344 (WO2017/070632) 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.

25 **Cytidine deaminases**

The fusion proteins provided herein comprise 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 cytidine 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).

5 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
10 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,
15 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,
20 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.

A fusion protein of the invention comprises two or more nucleic acid editing domains. 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
25 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 APOBEC1 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
30 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
 5 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
 10 embodiments, the deaminase is a rat deaminase, *e.g.*, rAPOBEC1. In some embodiments, the deaminase is a *Petromyzon marinus* cytidine deaminase 1 (pmCDA1). In some embodiments, the deaminase is a human APOBEC3G. In some embodiments, the deaminase is a fragment
 15 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 comprises 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
 20 least 99%), or at least 99.5% identical to the deaminase domain of any deaminase described herein.

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
 20 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).

Cas9 complexes with guide RNAs

Some aspects of this disclosure provide complexes comprising any of the fusion
 25 proteins provided herein, and a guide RNA bound to a Cas9 domain (*e.g.*, a dCas9, a nuclease active Cas9, or a Cas9 nickase) of fusion protein. 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,
 30 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
 35 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 a gene of interest (*e.g.*, a gene associated with a disease or disorder).

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, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence.

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.

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
 5 guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

Additional Domains

A base editor described herein can include any domain which helps to facilitate the
 10 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
 15 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.

In some embodiments, a base editor can comprise a uracil glycosylase inhibitor (UGI)
 20 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
 25 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.

30 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:eaa04774 (2017), the entire content of which is hereby incorporated by reference.

Additionally, in some embodiments, a Gam protein can be fused to an N terminus of a base editor. In some embodiments, 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:eaa04774 (2017). In some embodiments, 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.

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

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 an adenosine

deaminase domain or a cytidine deaminase domain, wherein the aforementioned domains are fused to a polynucleotide binding domain, thereby forming a nucleobase editor capable of inducing changes at one or more bases within a nucleic acid molecule as described herein and at least one guide polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence

5 comprises a targeted nucleobase pair; (b) inducing strand separation of the target region; (c) converting a first nucleobase of the target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of the 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

10 embodiments, step (b) is omitted. In some embodiments, the 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

15 genes, wherein at least one gene is located in a different locus.

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

20 some embodiments, the second base is inosine.

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,

25 without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system. In some embodiments, the base editor system comprises a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a

30 nucleobase editing domain (*e.g.*, deaminase domain) for editing the nucleobase; and 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 base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, deaminase domain) for editing the nucleobase, and a guide

polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. 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

5 polynucleotide programmable RNA binding domain. In some cases, a deaminase domain can be a cytosine deaminase or a cytidine deaminase, an adenine deaminase or an adenosine deaminase. In some embodiments, the terms “cytosine deaminase” and “cytidine deaminase” can be used interchangeably. In some embodiments, the terms “adenine deaminase” and “adenosine deaminase” can be used interchangeably. In some cases, a deaminase domain can

10 be a cytosine deaminase or a cytidine deaminase. In some cases, a deaminase domain can be an adenine deaminase or an adenosine deaminase. Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), 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

15 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:eaa04774 (2017), the entire

20 contents of which are hereby incorporated by reference.

In some embodiments, a single guide polynucleotide may be utilized to target a deaminase 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.

25 The nucleobase components 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, the 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

30 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 sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or a RNA recognition motif.

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.

5 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
10 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
15 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
20 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
25 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
30 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
 5 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.

In some embodiments, the base editor inhibits base excision repair of the edited
 10 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,
 15 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.

In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site.
 20 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.

In some embodiments, the target region comprises a target window, wherein the target
 25 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
 30 performed using any of the base editors provided herein. In some embodiments, a target window is a deamination window.

In some embodiments, non-limiting exemplary cytidine base editors (CBE) include 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.

In some embodiments, the adenosine base editor (ABE) 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 an 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.

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.

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).

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).

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 9. 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 9. 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 Table 9 below.

Table 9. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
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
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

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
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	F	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	T	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	L	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
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	F	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	L	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

In some embodiments, the base editor is an eighth generation ABE (ABE8). In some embodiments, the ABE8 contains a Tada*8 variant. In some embodiments, the ABE8 has a monomeric construct containing a Tada*8 variant (“ABE8.x-m”). In some

5 embodiments, the ABE8 is ABE8.1-m, which has a monomeric construct containing

TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-m, which has a monomeric construct containing TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-m, which has a monomeric construct containing TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-m, which has a monomeric construct containing TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-m, which has a monomeric construct containing TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-m, which has a monomeric construct containing TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-m, which has a monomeric construct containing TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-m, which has a monomeric construct containing TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-m, which has a monomeric construct containing TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-m, which has a monomeric construct containing TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-m, which has a monomeric construct containing

TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-m, which has a monomeric construct containing TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-m, which has a monomeric construct

5 containing TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-m, which has a monomeric construct containing TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T

10 mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing wild-type E. coli TadA fused to a TadA*8 variant ("ABE8.x-d"). In some embodiments, the ABE8 is ABE8.1-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is

15 ABE8.2-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to

20 TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-d, which

25 has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10

30 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some

embodiments, the ABE8 is ABE8.12-d, which has heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y123H (Y123H reverted from H123Y),

5 Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the

10 ABE8 is ABE8.16-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-d, which

15 has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-d,

20 which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-d, which has a

25 heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-d, which has a heterodimeric construct containing wild-

30 type *E. coli* TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing TadA*7.10 fused to a TadA*8 variant ("ABE8.x-7"). In some embodiments, the ABE8 is ABE8.1-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10

with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H

- (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-7, which has a
- 5 heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some
- 10 embodiments, the ABE8 is ABE8.22-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-7, which has a
- 15 heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24

- In some embodiments, the ABE is ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d,
- 20 ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d, or ABE8.24-d as shown in **Table 10** below.

25 **Table 10: Adenosine Deaminase Base Editor 8 Variants**

ABE8	Adenosine Deaminase	Adenosine Deaminase Description
ABE8.1-m	TadA*8.1	Monomer_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	Monomer_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	Monomer_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	Monomer_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	Monomer_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	Monomer_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	Monomer_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	Monomer_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	Monomer_TadA*7.10 + Y147R_Q154R_I76Y

ABE8.10-m	TadA*8.10	Monomer_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	Monomer_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	Monomer_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	Monomer_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	Monomer_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	Monomer_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	Monomer_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	Monomer_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	Monomer_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	Monomer_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	Monomer_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R
ABE8.21-m	TadA*8.21	Monomer_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	Monomer_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	Monomer_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	Monomer_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	Heterodimer_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	Heterodimer_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	Heterodimer_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	Heterodimer_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	Heterodimer_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	Heterodimer_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	Heterodimer_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154R)
ABE8.12-d	TadA*8.12	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	Heterodimer_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S)
ABE8.15-d	TadA*8.15	Heterodimer_(WT) + (TadA*7.10 + V82S_Y147R)
ABE8.16-d	TadA*8.16	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R)
ABE8.21-d	TadA*8.21	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H)
ABE8.24-d	TadA*8.24	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147T)
ABE8.25	TadA*8.25	
ABE8.26	TadA*8.26	

In some embodiments, the ABE8 is ABE8a-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-m, which has a monomeric construct containing TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-m, which has a monomeric construct containing TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some
 10 embodiments, the ABE8 is ABE8e-m, which has a monomeric construct containing TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments,
 15 the ABE8 is ABE8b-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some
 20 embodiments, the ABE8 is ABE8d-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).
 25

In some embodiments, the ABE8 is ABE8a-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10
 30 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10

with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

- 5 In some embodiments, the ABE is ABE8a-m, ABE8b-m, ABE8c-m, ABE8d-m, ABE8e-m, ABE8a-d, ABE8b-d, ABE8c-d, ABE8d-d, or ABE8e-d, as shown in Table 11 below. In some embodiments, the ABE is ABE8e-m or ABE8e-d. ABE8e shows efficient adenine base editing activity and low indel formation when used with Cas homologues other than SpCas9, for example, SaCas9, SaCas9-KKH, Cas12a homologues, *e.g.*, LbCas12a, enAs-Cas12a, SpCas9-NG and circularly permuted CP1028-SpCas9 and CP1041-SpCas9. In addition to the mutations shown for ABE8e in Table 11, off-target RNA and DNA editing were reduced by introducing a V106W substitution into the TadA domain (as described in M. Richter et al., 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein).

15

Table 11: Additional Adenosine Deaminase Base Editor 8 Variants

ABE8 Base Editor	Adenosine Deaminase	Adenosine Deaminase Description
ABE8a-m	TadA*8a	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	Monomer_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	Monomer_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	Heterodimer_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	Heterodimer_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	Heterodimer_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

- In some embodiments, base editors (*e.g.*, ABE8) are generated by cloning an adenosine deaminase variant (*e.g.*, TadA*8) into a scaffold that includes a circular permutant Cas9 (*e.g.*, CP5 or CP6) and a bipartite nuclear localization sequence. In some embodiments, the base

- editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP5 variant (*S. pyrogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP5 variant (*S. pyrogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP6 variant (*S. pyrogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP6 variant (*S. pyrogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in **Table 12** below.

Table 12. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
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

10

As shown in Table 13 below, genotypes of 40 ABE8s are described. Residue positions in the evolved *E. coli* TadA portion of ABE are indicated. Mutational changes in ABE8 are shown when distinct from ABE7.10 mutations. In some embodiments, the ABE has a genotype of one of the ABEs as shown in Table 13 below.

15

Table 13. Residue Identity in Evolved TadA

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				

ABE8.19-m					S				H		R		R				
ABE8.20-m				Y	S				H		R		R				
ABE8.21-m											R		S				
ABE8.22-m					S								S				
ABE8.23-m					S				H								
ABE8.24-m					S				H		T						
ABE8.1-d											T						
ABE8.2-d											R						
ABE8.3-d													S				
ABE8.4-d									H								
ABE8.5-d					S												
ABE8.6-d																	R
ABE8.7-d													R				
ABE8.8-d									H		R		R				
ABE8.9-d				Y							R		R				
ABE8.10-d											R		R				R
ABE8.11-d											T		R				
ABE8.12-d											T		S				
ABE8.13-d				Y					H		R		R				
ABE8.14-d				Y	S												
ABE8.15-d					S						R						
ABE8.16-d					S				H		R						
ABE8.17-d					S								R				
ABE8.18-d					S				H				R				
ABE8.19-d					S				H		R		R				
ABE8.20-d				Y	S				H		R		R				
ABE8.21-d											R		S				
ABE8.22-d					S								S				
ABE8.23-d					S				H								
ABE8.24-d					S				H		T						

In some embodiments, the base editor is ABE8.1, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

5 ABE8.1_Y147T_CP5_NGC PAM_monomer
MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD**SGGSSGGSSGSETPGTSES**
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
10 GRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKGGMQPT

VAYSVLVVAKEGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPK
 YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQH
 KHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPRAF
 KYFDTTIARKEYRSTKEVLDATLIHQSI TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGG
 5 **GGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE**
 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDLNPDNSDVKL
 FIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSL
 GLTPNFKS NFDAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
 10 NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
 EFYKF I KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
 LKDNREKIEKILTFRIPIYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIER
 MTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILED
 15 IVLTTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
 DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
 VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
 QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGK
 SDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
 20 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

25 In some embodiments, the base editor is ABE8.1, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_monomer:

30 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
 GRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
 35 VAYSVLVVAKEGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPK

YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFYLYLASHYEKLKGPS PEDNEQKQLFVEQH
 KHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENI IHLFTLTNLGAPRAF
 KYFDTTIARKEYRSTKEVL DATLIHQ SITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGG
 GGMDDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
 5 ATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDD SFFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVKL
 FIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSL
 GLTPNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
 NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
 10 EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYF
 LKDNREKIEKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASAQSFIER
 MTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
 RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
 IVLTLTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
 15 DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVK
 VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
 QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
 SDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLN
 20 AVVG TALIKKYPKLESEFVYG DYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italics sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

In some embodiments, the base editor is ABE8.14, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

pNMG-357_ABE8.14 with NGC PAM CP5

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 30 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDGGSSGGSSGSETPGTSESAT
 TPESGGSSGGSMSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
 LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSSGSSGG
 SSGSETPGTSESATPESGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

TNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
 DPKKYGGFMQPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYK
 EVKCDLIIKLPKYSLELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPE
 DNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHL
 5 FTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQ SITGLYETRIDLSQLGGD *GGSGGS*
GGSGGS *GGSGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI*
 GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEED
 KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEG
 DLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 10 NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNL
 SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
 YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELH
 AILRRQEDFY PFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVV
 DKGASAQSFIERMTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
 15 KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
 LDNEENEDI EDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
 GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
 PAIKKGILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EG IKELG
 SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
 20 KVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
 IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREI
 NNYHHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMIAKSE QEGADKRTADGSEF
ESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 25 sequence indicates sequence derived from Cas9, the italics sequence denotes a linker
 sequence, and the underlined sequence denotes a bipartite nuclear localization sequence.

In some embodiments, the base editor is ABE8.8-m, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

30 ABE8.8-m
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGG LVMQNYRLIDATLYVT FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLC RFFRMPRR VFNAQKKAQSSTD *SGGSSGGSSGSETPGTSES*
*ATP*ESSGGSSGSDKKYSIGL AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDL
 NPDNSDVKLFQVLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNL
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD
 5 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 10 NEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 15 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQIEGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
 TIMERSSEFEKNPIDFLEAKGYKEVKDLIIKLPKYSLELENGRKRMLASAGELQKGNELAL
 20 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLNDKV
 LSAYNKHDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 25 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.8-d, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

30 ABE8.8-d
 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG

LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFFRMPRRVFNAQKKAQSSSTDSSGSSGG
 SSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 5 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
 KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLN
 LIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDAEDAKLQLSKDITYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 10 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEET
 ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM
 RKPFLSGEQKKAIVDLLFKTNRKVTQKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 LLKIIKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSL
 15 HEHIANLAGSPAIKKILQTVKVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERG
 GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
 20 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAVKVEKGKS
 KKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSK
 RVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 25 EVLDATLIHQSI TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations.

30 In some embodiments, the base editor is ABE8.13-m, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLYDATLYVTFEPVCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 **LLFDSGETAEATRLKRTARRRYTRRKNRICYLQE**IFSNEMAKVDD**SSFFHRLEESFLVEEDKK**
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSNFDLAEDAKQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 10 GYIDGGASQEEFYKFIFKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 15 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPA
 IKKGILQTVKVVDDELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEIGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIINN
 20 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEI GKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKEISILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
 TIMERSSEFEKNPIDFLEAKGYKEVKDLIIKLPKYSLFELNGRKRMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEDEQQLFVEQHKKHYLDEIIEQISEFSKRVLADANLDKV
 25 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 30 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.13-d, which comprises or consists
 essentially of the following sequence or a fragment thereof having adenosine deaminase
 activity:

ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATP**ESSGGSSGSS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIG
 5 LHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFFRMPRVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSES**ATP**ESSGGSSGSS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
 10 KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLN
 LIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQY
 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEET
 15 ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
 RKP AFLSGEQKKAIVDLLFKTNRKVTQVQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 LLKIIKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDLSL
 HEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 20 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERG
 GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 25 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS
 KKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIIEQISEFSK
 RVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

30 In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.17-m, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.17-m

5 MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
ATPESSSGSSGSSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
10 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSDTKADLRILIYLAHAMIKFRGHFLIEGDL
NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
LFGNLIALSLGLTPNFKSDFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAI
15 LRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLD
NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWRLSRKLINGI
RDQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPA
20 IKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNPSEEVVKMKMYWRQLLNALITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAYFFYSNI
25 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
TIMERSSEKPNIDFLEAKGYKEVKDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
LSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI
30 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations.

In some embodiments, the base editor is ABE8.17-d, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.17-d

5 MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSTDSSGSSGGSSGSETPGTSES
ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
10 AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSSGSSGG
SSGSETPGTSESATPESSGGSSGSSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLN
15 LIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEET
ITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM
20 RKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDQSGKTILDFLKS DGFANRNFQMLIHDDSLTFKEDIQKAQVSGQGD SL
HEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMK
RIE EG IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
25 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI G
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKDWDPKKYGGFDSPTVAYSVLVVAKEVGKS
30 KKLKSVKELLGITIMERS SFEKNPIDFLEAGYKEVKKDLIIKLPKYSLFELENGKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEII EQISEFSK
RVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTK
EVLDATLIHQ SITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations.

- 5 In some embodiments, the base editor is ABE8.20-m, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMA
 10 LRQGGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
 ATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 15 NPDNSDVKLFILQVLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNG
 LFGNLIALLSLGLTPNFKSDFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 20 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 25 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVKKMKNYWRQLLNALITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDKYDVVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 30 FSKEISILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKDLIIKLPKYSLELENKRKMLASAGELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEQNEQQLFVEQHKHYLDEIIIEQISEFSKRVLADANLNDKV
 LSAYNKHARDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations.

- 5 In some embodiments, the base editor is ABE8.20-d, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
 10 LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLSDFFRMRQEIKAKKQSSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRRAIG
 LHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
 AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRVFNAQKKAQSSSTDSSGSSGG
 15 SSGSETPGTSESATPESSGGSSGSSDKKYSIGLLAIGTNSVGWAVITDEYKVPSKKFKVLGNT
 DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
 KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLN
 LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQY
 20 ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
 EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
 PHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEET
 ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGM
 RKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
 25 LLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSL
 HEHIANLAGSPAIKKILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
 RIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
 SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNALITQRKFDNLTKAERG
 30 GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRK
 DFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG
 KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV
 NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKS
 KKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLAS

AGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSK
 RVILADANLDKVL SAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
 EVLDATLIHQSI TGLYETRIDLSQLGGDE GADKRTADGSEFESPKKKRKV*

- 5 In the above sequence, the plain text denotes an adenosine deaminase sequence, bold
 sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker
 sequence, underlined sequence denotes a bipartite nuclear localization sequence, and double
 underlined sequence indicates mutations.

In some embodiments, an ABE8 of the invention is selected from the following
 sequences:

- 10 01. monoABE8.1_bpNLS + Y147T
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 15 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALS LGLTPNFKS NFDLAEDAKLQLSKDTYDDDL DNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 20 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 25 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVRINN
 30 YHHAHDAYLNAVVG TALIKKYPKLESEFVYG DYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV

LSAYNKHHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

02. monoABE8.1_bpNLS + Y147R

5 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
ATPESSSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
10 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKN
LFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
15 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
20 IKKGILQTVKVVDDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKMKMYWRQLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
25 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGS PEDNEQQLFVEQHKKHYLDEIEQISEFSKRVILADANLDKV
LSAYNKHHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVL DATLIHQSI
30 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

03. monoABE8.1_bpNLS + Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

GMNHRVEITEGILADECAALLCYFFRMPSVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 5 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKN
 LFGNLIALLSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 10 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLS GEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMI EERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 15 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVINN
 YHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGG
 20 FSKESILPKRNSDKLIARKKDWDPPKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRKMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKKHYLDEII EQISEFSKRVLADANLDKV
 LSAYNKH RDKPIREQAENI IHLFTLTNLGAPAAF KYFDTTIDRKQYRSTKEVL DATLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

25

04. monoABE8.1_bpNLS + Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCYFFRMPSVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 30 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSEFFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKN
 LFGNLIALLSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD

AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 5 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 10 LTRSDKNRGKSDNPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPPKKGFFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 15 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

 20 05. monoABE8.1_bpNLS + V82S
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYSTFEPVCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGA
 25 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDKAILSARLSKSRLENLIAQLPGEKKNL
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYA
 30 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI

RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDDELVKVMGRHKPENIVIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 5 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARELQKGNELAL
 10 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVI
 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

06. monoABE8.1_bpNLS + T166R

15 MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHRLEESFLVEEDKK
 20 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 25 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 30 IKKGILQTVKVVDDELVKVMGRHKPENIVIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI

MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLNDKV
 5 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

07. monoABE8.1_bpNLS + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 10 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 15 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNK
 LFGNLIALSLGLTPNFKSDFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 20 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 25 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKTEVQTGG
 30 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLEFELNGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLNDKV
 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

08. monoABE8.1_bpNLS + Y147R_Q154R_Y123H
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
5 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEGDL
NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKN
10 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
15 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
20 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNALITQRKFDNLTKAERGGLSELDAKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
25 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKKHYLDEIEQISEFSKRVLADANLDKV
LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFYFDTTIDRKQYRSTKEVLDTLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV
- 30 09. monoABE8.1_bpNLS + Y147R_Q154R_I76Y
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 5 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 10 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 15 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLASARELQKGNELAL
 20 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKKHYLDEIIIEQISEFSKRVLADANLNDKV
 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

10. monoABE8.1_bpNLS + Y147R_Q154R_T166R

25 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSRDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFFHRLEESFLVEEDKK
 30 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI

LRRQEDFYFPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEDEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 5 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDDELVKVMGRHKPENIVIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINN
 10 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLNDKV
 15 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

11. monoABE8.1_bpNLS + Y147T_Q154R
 MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 20 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCTFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 25 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALSLGLTPNFKSDFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEDEVVDK
 30 GASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDDELVKVMGRHKPENIVIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ

ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 5 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKV
 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 10 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

12. monoABE8.1_bpNLS + Y147T_Q154S
 MSEVEFSHEYWMRHALLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 15 GMNHRVEITEGILADECAALLCTFFRMPRSVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHRLSEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 20 LFGNLIALLSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 25 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKS DGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDLVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEELGELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 30 LTRSDKNRGKSDNPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINN
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI

TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRKMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKV
 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

5

13. monoABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y
 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
 GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSES

10 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNL
 LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD

15 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD

20 NEENEDILEDIVLTTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKMYWRQLNAKLITQRKFDNLTKAERGGLSELDKAGFIK

25 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNI
 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDKYVDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGI
 TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGKRKMLASARELQKGNELAL

30 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKV
 LSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

14. monoABE8.1_bpNLS + V82S + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
 LRQGGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
 GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSSGGSSGGSSGSETPGTSES
 ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 5 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEGDL
 NPDNSDVKDLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKN
 LFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSD
 AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 10 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITPWNFEVVDK
 GASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
 AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
 15 RDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
 IKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKNYWRQLNAKLITQRKFDNLTKAERGGLSELDAKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINN
 20 YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI
 MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGG
 FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKEKGKSKKLKSVKELLGI
 TIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARELQKGNELAL
 PSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKV
 25 LSAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDTLIHQSI
 TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

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
 30 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.

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

5 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

10 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.

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

15 XTEN linker domain). In some embodiments, 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.*, an adenosine deaminase domain or a cytidine deaminase domain). In some embodiments, a linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond,

20 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

25 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

30 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.*, UGI, cytidine deaminase, etc.).

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*, (GGGG)*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.

Linkers

In certain embodiments, linkers may be used to link any of the peptides or peptide domains of the invention. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length. In certain embodiments, the linker is a polypeptide or based on amino acids. In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may 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, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is a bond (*e.g.*, a covalent bond), an organic molecule, group, polymer, or chemical moiety. 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.

In some embodiments, the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via a linker that is 4, 16, 32, or 104 amino acids in length. In some embodiments, the linker is about 3 to about 104 amino acids in length. In some

embodiments, any of the fusion proteins provided herein, comprise a cytidine deaminase, adenosine deaminase and a Cas9 domain that are fused to each other via a linker. Various linker lengths and flexibilities between the deaminase domain (*e.g.*, an engineered ecTadA) and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers of the form (GGGS)_n, (GGGGS)_n, and (G)_n to more rigid linkers of the form (EAAAK)_n, (SGGS)_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) and (XP)_n 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 cytidine deaminase and adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker (*e.g.*, an XTEN linker) comprising the amino acid sequence SGSETPGTSESATPES.

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.

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:eao4774 (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.

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.

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.

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.

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.

Non-limiting examples of protein domains which can be included in the fusion protein include deaminase domains (*e.g.*, cytidine deaminase, adenosine deaminase), a uracil

5 glycosylase inhibitor (UGI) domain, epitope tags, and reporter gene sequences.

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-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase,

10 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)
15 fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

Base Editor Efficiency

CRISPR-Cas9 nucleases have been widely used to mediate targeted genome editing.

20 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
25 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
30 that can more efficiently and cleanly make precise point mutations are needed. Base editing systems as provided herein provide a new way to provide genome editing without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

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
5 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.

10 In some embodiments, any of base editor systems provided herein result 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
15 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.

20 In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result 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
25 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.

30 In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in at most 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.3% indel formation in the target polynucleotide sequence. In some embodiments, any of base

editor systems comprising one of the ABE8 base editor variants described results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising one of ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising an ABE7.10.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has at least 0.01%, 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%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, a base editor system comprising one of the ABE8 base editor variants described herein has at least 0.01%, 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%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising an ABE7.10.

The disclosure provides adenosine deaminase variants (*e.g.*, ABE8 variants) that have increased efficiency and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (*e.g.*, “bystanders”).

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations. In some embodiments, an unintended editing or mutation is a bystander mutation or bystander editing, for example, base editing of a target base (*e.g.*, A or C) in an unintended or non-target position in a target window of a target nucleotide sequence. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander

editing or mutations 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor

5 system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold,

10 at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing. In some embodiments, an unintended editing or mutation is a spurious mutation or spurious editing, for example, non-

15 specific editing or guide independent editing of a target base (*e.g.*, A or C) in an unintended or non-target region of the genome. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor

20 variants described herein has reduced spurious editing 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some

25 embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9

30 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

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 (*i.e.*, mutation of bystanders). 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

5 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.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, 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

10 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%, or at least 99% base editing efficiency. In some embodiments, the base editing efficiency may be measured by calculating the percentage of edited nucleobases in a population of cells. In some embodiments, any of the ABE8 base editor variants described herein have base editing efficiency of at least 0.01%, at least 1%, at

15 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has

20 higher base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least

25 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least

30 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6

fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% on-target base editing efficiency. In some embodiments, any of the ABE8 base editor variants described herein have on-target base editing efficiency of at least 0.01%, 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited target nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has higher on-target base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

The ABE8 base editor variants described herein may be delivered to a host cell via a plasmid, a vector, a LNP complex, or an mRNA. In some embodiments, any of the ABE8 base editor variants described herein is delivered to a host cell as an mRNA. In some embodiments, an ABE8 base editor delivered via a nucleic acid based delivery system, *e.g.*, an mRNA, has on-target editing efficiency of at least 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases. In some embodiments, an ABE8 base editor delivered by an mRNA system has higher base editing efficiency compared to an ABE8 base editor delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein

has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

10 In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result 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% off-target editing in the target polynucleotide sequence.

20 In some embodiments, any of the ABE8 base editor variants described herein has lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least about 2.2 fold decrease in guided off-target editing

efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the ABE8 base editor variants described herein has lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has 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%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 5.0 fold, at least 10.0 fold, at least 20.0 fold, at least 50.0 fold, at least 70.0 fold, at least 100.0 fold, at least 120.0 fold, at least 130.0 fold, or at least 150.0 fold lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein has 134.0 fold decrease in guide-independent off-target editing efficiency (*e.g.*, spurious RNA deamination) when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein does not increase guide-independent mutation rates across the genome.

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. In some embodiments, any of the base editors provided herein are capable of generating at least 0.01% of intended mutations (*e.g.*, spurious off-target editing or bystander editing). In some embodiments, an intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to alter or correct a mutation in a target gene. 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 in a nucleic acid

(*e.g.* a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to alter or correct an intended mutation. In some embodiments, the intended mutation is a mutation that

5 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. In some embodiments, the intended mutation is a mutation that alters the splicing of a gene. In some embodiments, the intended mutation is a mutation that alters the regulatory sequence of a gene (*e.g.*, a gene promotor or gene repressor).

10 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

15 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.

The number of intended mutations and indels can be determined using any suitable

20 method, for example, as described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632); 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

25 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:eaa04774 (2017); the entire contents of which are hereby incorporated by reference.

In some embodiments, to calculate indel frequencies, sequencing reads are scanned

30 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.

5 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
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.

 In some embodiments, the base editors provided herein are capable of limiting
15 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. In some embodiments, any of the base editors provided herein are capable of limiting the formation of indels at a region of a nucleic acid to less than 1%, less than 1.5%, less than 2%, less than 2.5%, less than 3%, less than 3.5%, less
20 than 4%, less than 4.5%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, less than 10%, less than 12%, less than 15%, or less than 20%. The number of indels formed at a nucleic acid region may 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, any number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least
25 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 a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) to a base editor.

Multiplex Editing

30 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

5 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 requires a PAM sequence to target

10 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

15 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.

In some embodiments, the plurality of nucleobase pairs are in one more genes. In

20 some embodiments, the plurality of nucleobase pairs is in the same gene. In some embodiments, at least one gene in the one more genes is located in a different locus.

In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein coding region. In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein non-coding region. In some embodiments, the

25 editing is editing of the plurality of nucleobase pairs in at least one protein coding region and at least one protein non-coding region.

In some embodiments, the editing is in conjunction with one or more guide polynucleotides. In some embodiments, the base editor system can comprise one or more base editor system. In some embodiments, the base editor system can comprise one or more

30 base editor systems in conjunction with a single guide polynucleotide. In some embodiments, the base editor system can comprise one or more base editor system in conjunction with a plurality of guide polynucleotides. In some embodiments, the editing is in conjunction with one or more guide polynucleotide with a single base editor system. In some embodiments, the 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, the 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, the editing is in conjunction with a mix of at least one guide polynucleotide
5 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
10 be appreciated that the editing can comprise a sequential editing of a plurality of nucleobase pairs.

METHODS OF USING BASE EDITORS

Base editing of genes and alleles opens up new strategies for therapeutics and basic
15 research.

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
20 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.

It will be understood that the numbering of the specific positions or residues in the respective sequences, *e.g.*, polynucleotide or amino acid sequences of a disease-related gene
25 or its encoded protein, respectively, 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
30 sequence alignment and determination of homologous residues.

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.

In some embodiments, the target DNA sequence encodes a protein (*e.g.*, LKKR2), 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.

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 Cpf1 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.

Generating an Intended Mutation

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.

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.

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

Details of base editor efficiency are described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), 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:eaa04774 (2017), the entire contents of which are hereby incorporated by reference.

In some embodiments, the editing of the plurality of nucleobase pairs in one or more genes result in formation of at least one intended mutation. In some embodiments, the formation of the 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.

Expression of Fusion Proteins in a Host Cell

Fusion proteins of the disclosure comprising an adenosine deaminase variant may be expressed in virtually any host cell of interest using routine methods known to the skilled artisan. For example, a DNA encoding an adenosine deaminase of the disclosure can be cloned by designing suitable primers for the upstream and downstream of CDS based on the cDNA sequence. The cloned DNA may be directly, or after digestion with a restriction enzyme when desired, or after addition of a suitable linker and/or a nuclear localization signal ligated with a DNA encoding one or more additional components of a base editing system. The base editing system is translated in a host cell to form a complex.

A DNA encoding a protein domain described herein can be obtained by chemically synthesizing the DNA, or by connecting synthesized partly overlapping oligoDNA short chains by utilizing the PCR method and the Gibson Assembly method to construct a DNA encoding the full length thereof. The advantage of constructing a full-length DNA by chemical synthesis or a combination of PCR method or Gibson Assembly method is that the

codon to be used can be designed in CDS full-length according to the host into which the DNA is introduced. In the expression of a heterologous DNA, the protein expression level is expected to increase by converting the DNA sequence thereof to a codon highly frequently used in the host organism. As the data of codon use frequency in host to be used, for
5 example, the genetic code use frequency database (kazusa.or.jp/codon/index.html) disclosed in the home page of Kazusa DNA Research Institute can be used, or documents showing the codon use frequency in each host may be referred to. By reference to the obtained data and the DNA sequence to be introduced, codons showing low use frequency in the host from
10 amino acid and showing high use frequency.

An expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme can be produced, for example, by linking the DNA to the downstream of a promoter in a suitable expression vector. In some embodiments, animal cell expression plasmids (*e.g.*, pA1-11, pXT1,
15 pRc/CMV, pRc/RSV, pcDNA1/Neo), and animal virus vectors such as retrovirus, vaccinia virus, adenovirus, and the like are used.

In some embodiments, any promoter appropriate for a host to be used for gene expression can be used. In a conventional method using DSB, since the survival rate of the host cell sometimes decreases markedly due to the toxicity, it is desirable to increase the
20 number of cells by the start of the induction by using an inductive promoter. However, since sufficient cell proliferation can also be afforded by expressing the nucleic acid-modifying enzyme complex of the present disclosure, a constitution promoter can also be used without limitation.

For example, when the host is an animal cell, SR α promoter, SV40 promoter, LTR
25 promoter, CMV (cytomegalovirus) promoter, RSV (*Rous sarcoma virus*) promoter, MoMuLV (*Moloney mouse leukemia virus*) LTR, HSV-TK (simple herpes virus thymidine kinase) promoter and the like are used. Of these, CMV promoter, SR α promoter and the like are preferable.

As the expression vector, besides those mentioned above, one containing enhancer,
30 splicing signal, terminator, polyA addition signal, a selection marker such as drug resistance gene, auxotrophic complementary gene and the like, replication origin and the like on demand can be used.

An RNA encoding a protein domain described herein can be prepared by, for example, transcription to mRNA in a vitro transcription system known per se by using a

vector encoding DNA encoding the above-mentioned nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme as a template.

5 A fusion protein of the disclosure can be intracellularly expressed by introducing an expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme into a host cell, and culturing the host cell.

As the animal cell, cell lines such as monkey COS-7 cell, monkey Vero cell, Chinese hamster ovary (CHO) cell, dhfr gene-deficient CHO cell, mouse L cell, mouse AtT-20 cell, mouse myeloma cell, rat GH3 cell, human FL cell and the like, pluripotent stem cells such as iPS cell, ES cell and the like of human and other mammals, and primary cultured cells
10 prepared from various tissues are used. Furthermore, zebrafish embryo, *Xenopus* oocyte and the like can also be used.

All the above-mentioned host cells may be haploid (monoploid), or polyploid (*e.g.*, diploid, triploid, tetraploid and the like). In the conventional mutation introduction methods, mutation is, in principle, introduced into only one homologous chromosome to produce a
15 hetero gene type. Therefore, desired phenotype is not expressed unless dominant mutation occurs, and homozygousness inconveniently requires labor and time. In contrast, according to the present disclosure, since mutation can be introduced into any allele on the homologous chromosome in the genome, desired phenotype can be expressed in a single generation even in the case of recessive mutation, which is extremely useful since the problem of the
20 conventional method can be solved.

An expression vector can be introduced by a known method (*e.g.*, lysozyme method, competent method, PEG method, CaCl₂ coprecipitation method, electroporation method, the microinjection method, the particle gun method, lipofection method, Agrobacterium method and the like) according to the kind of the host.

25 An animal cell can be introduced into a vector according to the methods described in, for example, Cell Engineering additional volume 8, New Cell Engineering Experiment Protocol, 263-267 (1995) (published by Shujunsha), and Virology, 52, 456 (1973). A cell introduced with a vector can be cultured according to a known method according to the kind of the host.

30 As a medium for culturing an animal cell, for example, minimum essential medium (MEM) containing about 5 to about 20% of fetal bovine serum (Science, 122, 501 (1952)), Dulbecco's modified Eagle medium (DMEM) (Virology, 8, 396 (1959)), RPMI 1640 medium (The Journal of the American Medical Association, 199, 519 (1967)), 199 medium (Proceeding of the Society for the Biological Medicine, 73, 1 (1950)) and the like are used.

The pH of the medium is preferably about 6 to about 8. The culture is performed at generally about 30°C to about 40°C. Where necessary, aeration and stirring may be performed.

When a higher eukaryotic cell, such as animal cell, is used as a host cell, a DNA encoding a base editing system of the present disclosure (*e.g.*, comprising an adenosine deaminase variant) is introduced into a host cell under the regulation of an inducible promoter (*e.g.*, metallothionein promoter (induced by heavy metal ion), heat shock protein promoter (induced by heat shock), Tet-ON/Tet-OFF system promoter (induced by addition or removal of tetracycline or a derivative thereof), steroid-responsive promoter (induced by steroid hormone or a derivative thereof) etc.), the induction substance is added to the medium (or removed from the medium) at an appropriate stage to induce expression of the nucleic acid-modifying enzyme complex, culture is performed for a given period to carry out a base editing and, introduction of a mutation into a target gene, transient expression of the base editing system can be realized.

Alternatively, the above-mentioned inductive promoter can also be utilized as a vector removal mechanism when higher eukaryotic cells, such as animal cell and the like are used as a host cell. That is, a vector is mounted with a replication origin that functions in a host cell, and a nucleic acid encoding a protein necessary for replication (*e.g.*, SV40 on and large T antigen, oriP and EBNA-1 etc. for animal cells), of the expression of the nucleic acid encoding the protein is regulated by the above-mentioned inducible promoter. As a result, while the vector is autonomously replicatable in the presence of an induction substance, when the induction substance is removed, autonomous replication is not available, and the vector naturally falls off along with cell division (autonomous replication is not possible by the addition of tetracycline and doxycycline in Tet-OFF system vector).

25 DELIVERY SYSTEM

A base editor disclosed herein can be encoded on a nucleic acid that is contained in a viral vector. Viral vectors can include lentivirus, Adenovirus, Retrovirus, and Adeno-associated viruses (AAVs). Viral vectors can be selected based on the application. For example, AAVs are commonly used for gene delivery *in vivo* due to their mild immunogenicity. Adenoviruses are commonly used as vaccines because of the strong immunogenic response they induce. Packaging capacity of the viral vectors can limit the size of the base editor that can be packaged into the vector. For example, the packaging capacity of the AAVs is ~4.5 kb including two 145 base inverted terminal repeats (ITRs).

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.

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.

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.

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.

5 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,
10 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.

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-
15 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
20 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.

The disclosed strategies for designing base editors can be useful for generating base editors capable of being packaged into a viral vector. The use of RNA or DNA viral based systems for the delivery of a base editor takes advantage of highly evolved processes for
25 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
30 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.

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
5 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
10 (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); Sommerfelt *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).

15 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
20 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.

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
25 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. 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
30 (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). The 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).

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.

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

Non-Viral Delivery of Base Editors

Non-viral delivery approaches for 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 14** (below).

Table 14

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-Dioleoyloxy)propyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocetadecylamidoglycylspermine	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-Dioleoyloxy-N-[2(sperminecarboxamido-ethyl)-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristoxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3 β -[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diideoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocetadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl)-oxymethyloxy]ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyl dimyristoylphosphatidylcholine	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-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolinium chloride	DOTIM	Cationic
N1 -Cholesterylloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoley-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoley-1-methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 15 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 15

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

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

Table 16 summarizes delivery methods for a polynucleotide encoding a fusion protein described herein.

5 **Table 16**

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

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Non-Viral	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
	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 liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

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).

5 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
10 toxicity due to over expression of the chosen nuclease.

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
15 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.

20 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.

25 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).

A base editor described herein with or without one or more guide nucleic can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral
30 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 example, 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.

For *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.

AAV has a packaging limit of 4.5 or 4.75 Kb. This means disclosed base editor as well as a promoter and transcription terminator can fit into 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.

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)).

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.

5 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
10 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 µl 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
15 preferred.

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
20 immediately frozen at -80°C.

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

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
30 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.

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.

The disclosure in some embodiments comprehends a method of modifying a cell or
5 organism. The cell can be a prokaryotic cell or a eukaryotic cell. The cell can be a mammalian cell. The mammalian cell may 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,
10 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.

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,
15 preferably a mammalian cell or a human cell.

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
20 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.or.jp/codon/ (visited Jul. 9, 2002), and these tables can be adapted in a number
25 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
30

codons) in a sequence encoding an engineered nuclease correspond to the most frequently used codon for a particular amino acid.

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.

Inteins

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.

Inteins (intervening protein) are auto-processing domains found in a variety of diverse organisms, which carry out a process known as protein splicing. Protein splicing is a multi-step biochemical reaction comprised of both the cleavage and formation of peptide bonds. While the endogenous substrates of protein splicing are proteins found in intein-containing organisms, inteins can also be used to chemically manipulate virtually any polypeptide backbone.

In protein splicing, the intein excises itself out of a precursor polypeptide by cleaving two peptide bonds, thereby ligating the flanking extein (external protein) sequences via the formation of a new peptide bond. This rearrangement occurs post-translationally (or possibly co-translationally). Intein-mediated protein splicing occurs spontaneously, requiring only the
5 folding of the intein domain.

About 5% of inteins are split inteins, which are transcribed and translated as two separate polypeptides, the N-intein and C-intein, each fused to one extein. Upon translation, the intein fragments spontaneously and non-covalently assemble into the canonical intein structure to carry out protein splicing in trans. The mechanism of protein splicing entails a
10 series of acyl-transfer reactions that result in the cleavage of two peptide bonds at the intein-extein junctions and the formation of a new peptide bond between the N- and C-exteins. This process is initiated by activation of the peptide bond joining the N-extein and the N-terminus of the intein. Virtually all inteins have a cysteine or serine at their N-terminus that attacks the carbonyl carbon of the C-terminal N-extein residue. This N to O/S acyl-shift is facilitated by a
15 conserved threonine and histidine (referred to as the TXXH motif), along with a commonly found aspartate, which results in the formation of a linear (thio)ester intermediate. Next, this intermediate is subject to trans-(thio)esterification by nucleophilic attack of the first C-extein residue (+1), which is a cysteine, serine, or threonine. The resulting branched (thio)ester intermediate is resolved through a unique transformation: cyclization of the highly conserved
20 C-terminal asparagine of the intein. This process is facilitated by the histidine (found in a highly conserved HNF motif) and the penultimate histidine and may also involve the aspartate. This succinimide formation reaction excises the intein from the reactive complex and leaves behind the exteins attached through a non-peptidic linkage. This structure rapidly rearranges into a stable peptide bond in an intein-independent fashion.

In some embodiments, an N-terminal fragment of a base editor (*e.g.*, ABE, CBE) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. 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
30 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.

In some embodiments, an ABE was split into N- and C- terminal fragments at Ala, Ser, Thr, or Cys residues within selected regions of SpCas9. These regions correspond to loop regions identified by Cas9 crystal structure analysis. The N-terminus of each fragment is fused to an intein-N and the C- terminus of each fragment is fused to an intein C at amino acid positions S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, and S590, which are indicated in bold capital letters in the sequence below.

1 mdkkysigld igtntsvwav itdeykvpsk kfkvlgntr hsikknliga llfdsgetae
 61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
 121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlndnsd
 181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknlgfng
 241 lialslgltp nfksnfdlae daklqlskdt yddldnlla qigdqyadlf laaknlsdai
 301 ll**S**dilrvn**T** ei**T**kaplsas mikrydehhq dltllkalvr qqlpekykei ffdq**S**kngya
 361 gyidggasqe efykfikpil ekmdgteell vknlnredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**A**rgn**S** rf**A**wm**T**rk**S**e e**T**i**T**pwnfee
 481 vvdkgasaqs fiermtndfk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**C**fd**S**vei sgvedrfrn**A**S lgtyhdllki
 601 ikdkdflrne enedilediv ltltlfedre mieerlktya hlfdckvmkq lkrrrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediql aqvsgggsd
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdyvdh
 841 ivpqsfldkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqlnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrnm tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
 1021 miaksegeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
 1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
 1141 ysvlvvakve kgksklkksv kellgitime rssfeknpid fleakgykev kkdliiklpk
 1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
 1261 qhkhyldei eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltlnlga
 1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd

PHARMACEUTICAL COMPOSITIONS

Other aspects of the present disclosure relate to pharmaceutical compositions comprising any of the base editors, fusion proteins, the fusion protein-guide polynucleotide complexes, or the edited cells 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).

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.).

Some nonlimiting 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 alcohols, 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,” “vehicle,” or the like are used interchangeably herein.

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.

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.

In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, *e.g.*, for gene editing. Suitable routes of administering 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, intraosseous, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

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.

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 al.*, 1989, J. Neurosurg. 71: 105.) Other controlled release systems are discussed, for example, in Langer, *supra*.

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.

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 al.*, Gene Ther. 1999, 6: 1438-47). Positively charged lipids such as N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammoniummethylsulfate, 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.

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
5 calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

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
10 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.

15 In another aspect, an article of manufacture containing materials useful for the treatment of the diseases described above is included. In some embodiments, the article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers can be formed from a variety of materials such as glass or plastic. In some embodiments, the container holds a composition
20 that is effective for treating a disease described herein and can have a sterile access port. For example, the container can be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle. The active agent in the composition is a compound of the invention. In some embodiments, the label on or associated with the container indicates that the composition is used for treating the disease of choice. The article
25 of manufacture can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as 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.

30 In some embodiments, any of the fusion proteins, gRNAs, complexes and/or cells 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, the pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. In some embodiments, the pharmaceutical composition comprises any of the edited cells as provided herein. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

Kits

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. In some embodiments, the kit comprises a nucleic acid construct comprising a nucleotide sequence encoding at least one guide RNA. In some embodiments, the kit comprises any of the edited cells as provided herein. In some embodiments, the kit comprises any of the pharmaceutical compositions as provided herein.

The kit provides, in some embodiments, instructions for using the kit to edit one or more 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.

In certain embodiments, the kit is useful for conditioning a subject for transplantation or engraftment.

EXAMPLES

5 The following examples are provided for illustrative purposes only and are not intended to limit the scope of the claims provided herein.

EXAMPLE 1. PAM Variant Validation in Base Editors

To evaluate a novel method for conditioning, a base editing strategy was developed
10 that targets the cell surface protein CD117/cKIT to create at least one amino acid substitution that alters an epitope binding domain, thereby preventing an anti-CD117 antibody from binding the edited cells without altering SCF binding or activity. Typical antibodies used for conditioning include AMG 191 and SR1, which may also be used in methods of the invention (FIG. 1). AMG 191 and SR1 are also known to deplete human hematopoietic stem and
15 progenitor cells and myelodysplastic syndromes cells *in vivo* (FIG. 2).

The conditioning methods described here provides for the post-HSCT use of anti-CD117 therapies, which cannot bind and deplete the edited cells, and thus could be used to expand gene edited cells *in vivo*, or allow for cancer treatment without hematopoietic toxicity. CD117 is altered (*e.g.*, using base editing) in a cell for transplantation to prevent
20 binding of anti-CD117 antibody (AMG 191; SR1). This alteration does not interfere with normal SCF signaling (FIG. 3). Using base editing, a nucleobase change is generated to create an amino acid substitution in CD117. Stem cell factor (SCF) drives HSC self-renewal and differentiation into progenitor cells (FIG. 3, at 1). Administering anti-CD117 antibody blocks SCF binding to CD117, thereby depleting HSCs and progenitor cells in the patient
25 (conditioning) (FIG. 3, at 2). Autologous gene edited HSCs are transplanted into the patient (FIG. 3, at 3). Gene-edited cells compete with residual host HSCs to repopulate bone marrow (BM) (FIG. 3, at 4). Anti-CD117 antibody blocks SCF binding to wildtype (WT) CD117, but cannot bind to HSCs with an edited CD117. Native, wild-type HSCs are targeted by anti-CD117 antibody, but gene edited HSCs are not (FIG. 3, at 5). Thus, both wild-type and
30 edited HSC cells express CD117 polypeptides that are activated by SCF binding (FIG. 4). However, binding of anti-CD117 antibody to wild-type CD117 disrupts SCF binding and results in inhibition of SCF signaling in wild-type cells (FIG. 5). In contrast, gene-edited HSCs are refractory to anti-CD117 antibody because the amino acid substitution introduced

in CD117 prevents the binding of anti-CD117 antibody, but does not interfere with normal SCF binding and signaling.

EXAMPLE 2. CD117/C-Kit Guide Screen

5 A screen was conducted to identify guide RNAs that introduce alterations in a CD117 polypeptide without altering SCF binding to CD117 or CD117 biological activity.

First, 189 guides capable of introducing a mutation in CD117 were screened in HEK293T cells. Guide numbers were assigned to each guide (*see* Table 23). Guide numbers 1-132 were used with an ABE8.8 base editor and guide numbers 133-189 were used with an IBE-NGC base editor (internal Cas9 deaminase fusion with NGC PAM). A guide for Hereditary Persistence of Fetal Hemoglobin (HPFH) was used as a positive control for the ABE8.8 editor, and a guide for Makassar-WT was used as a positive control for the IBE-NGC editor (Maxcyte oc-25, 0.3 M, 50 nM editor and 10 μ M guide).

The Makassar-WT guide nucleic acid sequence is as follows:

15 ACUUCUCCUCAGGAGUCAGGUGC. The Makassar-WT guide hybridizes to the complement of the following target sequence: ACTTCTCCTCAGGAGTCAGGTGC.

The HPFH guide nucleic acid sequence is as follows: CUUGACCAAUAGCCUUGACA.

The HPFH guide hybridizes to the complement of the following target sequence: CTTGACCAATAGCCTTGACA.

20 FIG. 6 shows the base editing efficiencies at different nucleotide positions at the target site for either the Makassar-WT guide (top) or the HPFH guide (bottom).

From the 189 guides, 31 guides were selected for ABE8.8 (HPFH) and 14 guides were selected for IBE-NGC (Makassar) based on-target editing efficiency. As shown in Table 17, five (5) guides, cc119, cc103, cc128, cc131, and cc80, used with ABE8.8 showed greater than 80% base editing efficiency (Maxcyte oc-25, 0.3 M, 50 nM editor and 10 μ M guide) (FIGS. 7A and 7B). Use of Guide cc103 resulted in a Y221C mutation, which is a naturally occurring mutation in CD117.

Table 17. Five (5) guides (ABE8.8) showed >80% editing efficiency

Guide Number	Epitope Group	Mutation	Efficiency (%)	Mutation	Efficiency (%)
Guide-CC119	3-Other epitope	D45G	87.19	N/A	N/A
Guide-CC103	3-Other epitope	Y221C	84.13	S220G	0.28
Guide-CC128	4-Functionally important residues	Y259C	83.54	N260V	78.76

Guide-CC131	4-Functionally important residues	K383G	16.31	N/A	N/A
Guide-CC80	2-Dimer interfering epitope	K342G/R	5.4	N/A	N/A

As shown in Table 18, eleven (11) guides, cc100, cc91, cc132, cc101, cc102, cc79, cc84, cc126, cc90, cc114 and cc93, used with ABE8.8 showed about 60% to 80% base editing efficiency (Maxcyte oc-25, 0.3 M, 50 nM editor and 10 uM guide) (FIGS. 8A and 8B). Guide cc102 resulted in a Y221C mutation, which is a naturally occurring mutation in CD117.

Table 18. Eleven (11) guides (ABE8.8) showed 60% to 80% editing efficiency

Guide Number	Epitope Group	Mutation	Efficiency (%)	Mutation	Efficiency (%)
Guide-CC100	3-Other epitope	I163T	19.84	F162P/L	3.98
Guide-CC91	3-Other epitope	T461A	75.79	Q460R	33.12
Guide-CC132	4-Functionally important residues	R381G	22.52	N/A	N/A
Guide-CC101	3-Other epitope	L472P	76.68	V473A	39.93
Guide-CC102	3-Other epitope	Y221C	70.14	S220G	7.21
Guide-CC79	2-Dimer interfering epitope	M351T	68.67	Y350H	7.48
Guide-CC84	3-Other epitope	Q79R	68.46	N80G	5.87
Guide-CC126	4-Functionally important residues	L124P	67.84	Y125H	60.51
Guide-CC90	3-Other epitope	E81G/D	66.2	N80S	1.01
Guide-CC114	3-Other epitope	D45G	65.51	N/A	N/A
Guide-CC93	3-Other epitope	V195A	64.67	L196P	0.4

As shown in Table 19, nine (9) guides, cc98, cc76, cc125, cc78, cc89, cc73, and cc110, cc72, and cc115, used with ABE8.8 showed about 40% to 60% base editing efficiency (Maxcyte oc-25, 0.3 M, 50 nM editor and 10 uM guide) (FIGS. 9A and 9B). Guide cc78 resulted in a D419G mutation, guide cc89 resulted in a T144A mutation, and guide cc110 resulted in a T322A, which are all naturally occurring mutations in CD117.

Table 19. Nine (9) guides (ABE8.8) showed 40% to 60% editing efficiency

Guide Number	Epitope Group	Mutation	Efficiency (%)	Mutation	Efficiency (%)
Guide-CC98	3-Other epitope	F324L	59.4	V323A	6.11
Guide-CC76	2-Dimer interfering epitope	T354A	39.04	R353G	1.37
Guide-CC125	3-Other epitope	I163V	56.44	N/A	N/A

Guide-CC78	2-Dimer interfering epitope	Y418C	52.25	D419G	10.86
Guide-CC89	3-Other epitope	T144A	50.26	N145G/Y	3.56
Guide-CC73	1-SCF interfering epitope	N77G/Y	50.23	E76G/K78E	20.25/3.25
Guide-CC110	3-Other epitope	T322A	47.48	N/A	N/A
Guide-CC72	1-SCF interfering epitope	E257G	43.28	Q256R/K258E	18.03/1.09
Guide-CC115	3-Other epitope	K484G	25.17	N/A	N/A

As shown in FIG. 10, four (4) guides, cc71, cc87, cc112, and cc124, used with ABE8.8 and all fourteen (14) guides, cc146, cc148, cc156, cc158, cc161, cc162, cc166, cc167, cc169, cc180, cc182, and cc184 used with IBE-NGC showed less than 40% base editing efficiency (Maxcyte oc-25, 0.3 M, 50 nM editor and 10 uM guide).

A summary of accessible mutations with a greater than 60% conversion rate is shown in Table 20.

Table 20. Summary of accessible mutations (>60% conversion)

Mutation	Domain	Guide number	Efficiency
D45G	3-Other epitope	Guide-CC119/114	87.19/65.51
Y221C	3-Other epitope	Guide-CC103/102	84.13/70.14
Y259C	4-Functionally important residues	Guide-CC128	83.54
N260V	4-Functionally important residues	Guide-CC128	78.76
T461A	3-Other epitope	Guide-CC91	75.79
L472P	3-Other epitope	Guide-CC101 (?)	71.99
M351T	2-Dimer interfering epitope	Guide-CC79	68.67
Q79R	3-Other epitope	Guide-CC84	68.46
L124P	4-Functionally important residues	Guide-CC126	67.84
E81G/D	3-Other epitope	Guide-CC90	66.2
V195A	3-Other epitope	Guide-CC93	64.67
Y125H	4-Functionally important residues	Guide-CC126	60.51

A summary of accessible mutations with about 25% to 60% conversion is shown in Table 21.

Table 21. Summary of accessible mutations (25% to 60% conversion)

Mutation	Domain	Guide number	Efficiency
-----------------	---------------	---------------------	-------------------

F324L	3-Other epitope	Guide-CC98	59.4
I163V	3-Other epitope	Guide-CC125/124	56.44/30/31
Y418C	2-Dimer interfering epitope	Guide-CC78	52.25
T144A	3-Other epitope	Guide-CC89/87	50.26/35.61
N77G/Y	1-SCF interfering epitope	Guide-CC73	50.23
T322A	3-Other epitope	Guide-CC110	47.48
E257G	1-SCF interfering epitope	Guide-CC72	43.28
T354A	2-Dimer interfering epitope	Guide-CC76	39.04
V473A	3-Other epitope	Guide-CC101	36.04
Q460R	3-Other epitope	Guide-CC91	33.12
S44P	3-Other epitope	Guide-CC112	29.4
K484G	3-Other epitope	Guide-CC115	25.17

The structural location of the mutations on the CD117 protein targeted by guides cc119, cc128, cc103, cc91, cc101, cc84, cc79, cc126, cc90, cc126, cc93, cc126, cc98, cc78, cc125, cc89, cc73, cc72, cc110, and cc76 are shown in FIGS. 11A-11T, respectively. Table 22 shows the guides with natural occurring mutations. In particular, guides cc20, cc102 and 103 resulted in the Y221C naturally occurring mutation; guides cc27, cc87, cc88, and cc89 resulted in the T144A naturally occurring mutation; and guides cc78 and cc137 resulted in the D419G naturally occurring mutation.

Table 22. Guides with natural occurring mutations

Guide Number	Group	Possible Mutations	Editing Efficiency
Guide-CC20	3-Other epitope	S220G Y221C	not selected
Guide-CC27	3-Other epitope	T144A N145D	not selected
Guide-CC29	3-Other epitope	I438M D439G	not selected
Guide-CC30	3-Other epitope	N486G T488A	not selected
Guide-CC35	3-Other epitope	I39V H40R	not selected
Guide-CC38	3-Other epitope	T303A T304A	not selected
Guide-CC41	3-Other epitope	K65E	not selected
Guide-CC42	3-Other epitope	M318V	not selected
Guide-CC70	4-Functionally important residues	I201V	not selected
Guide-CC78	2-Dimer interfering epitope	Y418C D419G	52.253
Guide-CC87	3-Other epitope	T144A N145G	35.613
Guide-CC88	3-Other epitope	T144A N145G	not selected
Guide-CC89	3-Other epitope	T144A N145G	50.255
Guide-CC96	3-Other epitope	K65R T67A	not selected
Guide-CC102	3-Other epitope	S220G Y221C	70.142

Guide-CC103	3-Other epitope	S220G Y221C	84.135
Guide-CC104	3-Other epitope	M171T I172T	not selected
Guide-CC105	3-Other epitope	M171T I172T	not selected
Guide-CC106	3-Other epitope	I438M D439G	not selected
Guide-CC110	3-Other epitope	T322A	47.484
Guide-CC127	4-Functionally important residues	K199G I201V	not selected
Guide-CC137	2-Dimer interfering epitope	Y418C D419G R420G	not selected
Guide-CC145	3-Other epitope	D52G E53G I54V R55G	not selected
Guide-CC146	3-Other epitope	D129G N130G D131G T132A	4.471
Guide-CC150	3-Other epitope	N99G K100G H101R	not selected
Guide-CC155	3-Other epitope	N145S Y146C	not selected
Guide-CC164	3-Other epitope	M171T I172T	not selected
Guide-CC179	3-Other epitope	D52G	not selected
Guide-CC182	3-Other epitope	T59A	18.415
Guide-CC188	4-Functionally important residues	K383G T385A	not selected
Guide-CC189	4-Functionally important residues	I201V	not selected

A CD117 target nucleotide sequence and corresponding amino acid sequence for the S220G and/or Y221C mutations is shown in FIG. 12A. Guide cc102 hybridizes to the complement of the CD117 target sequence shown below:

5 GCAAGCTATCTTCTTAGGGA

The editing efficiency at different base pair positions along the CD117 target site for guide cc102 is shown in FIG. 12B.

Guide cc103 hybridizes to the complement of the CD117 target sequence shown below:

10 CAAGCTATCTTCTTAGGGAA

The editing efficiency at different base pair positions along the CD117 target site for guide cc103 is shown in FIG. 12C.

A CD117 target nucleotide sequence and corresponding amino acid sequence for the Y418C and/or D419G mutations is shown in FIG. 12D. Guide cc78 hybridizes to the complement of the CD117 target sequence shown below:

15

ACTTACGACAGGCTCGTGAA

The editing efficiency at different base pair positions along the CD117 target site for guide cc78 is shown in FIG. 12E.

5 A CD117 target nucleotide sequence and corresponding amino acid sequence for the T144A and/or N145G mutations is shown in FIG. 12F. Guide cc87 hybridizes to the complement of the CD117 target sequence shown below:

TGACCAATTATTCCTCAAG

The editing efficiency at different base pair positions along the CD117 target site for guide cc87 is shown in FIG. 12G.

10 Guide cc89 hybridizes to the complement of the CD117 target sequence shown below:

AGTGACCAATTATTCCTCA

The editing efficiency at different base pair positions along the CD117 target site for guide cc89 is shown in FIG. 12H.

15 A CD117 target nucleotide sequence and corresponding amino acid sequence for the T322A mutation is shown in FIG. 12I. Guide cc110 hybridizes to the complement of the CD117 target sequence shown below:

ACTACAGTATTTGTAAACGA

20 The editing efficiency at different base pair positions along the CD117 target site for guide cc110 is shown in FIG. 12J.

A CD117 target nucleotide sequence and corresponding amino acid sequence for the D129G, N130G, D131G, and/or T132A mutations is shown in FIG. 12K. Guide cc146 hybridizes to the complement of the CD117 target sequence shown below:

AAGACAACGACACGCTGGTC

25 The editing efficiency at different base pair positions along the CD117 target site for guide cc146 is shown in FIG. 12L.

A CD117 target nucleotide sequence and corresponding amino acid sequence for the T59A mutation is shown in FIG. 12M. Guide cc182 hybridizes to the complement of the CD117 target sequence shown below:

30 GGCTGTTATGCACTGATCCG

The editing efficiency at different base pair positions along the CD117 target site for guide cc182 is shown in FIG. 12N.

The guide screen identified several potential target amino acid substitutions that can be introduced into the CD117 cell surface protein through the use of guide RNAs in

combination with base editors. The amino acid substitutions include one or more of the following: T13A, S35P, I39V, H40R, K43G, K43R, S44P, D45G, I47V, D52G, E53G, I54V, R55G, L56P, L57P, T59A, F63P, V64A, K65E, K65R, W66R, T67A, D72G, E73G, T74A, N75D, N75G, E76G, N77G, N77S, N77Y, K78E, K78R, Q79R, N80G, N80S, E81G, E81D, W82R, I83T, I83V, T84A, E85G, K86E, E88G, T90A, N99G, K100G, H101R, K116R, V120A, S123P, L124P, Y125H, K127G, K127R, E128G, D129G, D129E, N130G, N130D, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, K149E, K149G, Q152R, K154E, K154G, R161G, F162P, F162L, I163T, I163V, D165G, M171T, I172T, I172V, K173G, S174G, K176G, Q190R, E191G, K193G, V195A, L196P, S197P, E198G, K199G, F200P, I201T, I201V, L202P, V213A, V214A, S215P, V216A, K218G, K218R, S220G, Y221C, E225G, E227G, E228G, T230A, S240G, Y243C, K247G, R248G, Q256R, E257G, E257D, K258E, K258G, Y259C, Y259H, N260G, D266G, N268D, N268G, Y269C, T276A, I277T, I277V, S279P, R281G, V282A, S285P, N293G, N294G, T295A, F296P, S298P, N300G, N300S, T302A, T303A, T304A, M318V, T322A, V323A, F324L, N326G, D327G, D332G, I334V, K342E, K342G, K342R, Q347R, Y350H, M351T, R353G, T354A, T354I, K358G, E360G, D361G, K364G, E366G, N367G, H378R, T380A, R381G, K383G, T385A, T389A, D398G, V399A, N400G, V407A, Y408H, E414G, I415V, T417A, Y418C, Y418H, D419G, R420G, L421P, V422A, N423S, M425V, E435G, I438M, D439G, V454A, L455P, V457A, V459A, Q460R, T461A, N463G, S464P, S465P, F469P, K471E, K471G, L472P, V473A, Q475R, S476G, I478M, I478V, D479G, S481G, F483P, K484G, N486G, N486S, T488A, Y494C, N495D, D496G, K499E, Y503C, Y503H, F504P, and N505G.

Exemplary amino acid substitutions include one or more of the following: T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, and T488A. These potential target CD117 amino acid substitutions include naturally occurring mutations. Nonlimiting examples of CD117 naturally occurring mutations include: T13A, N77S, D129E, N130D, D131N, T144A, Y221C, E257D, T322A, T354I, D419G. Overall,

one or more of these mutations can be used to identify anti-CD117 antibodies that bind to wild-type CD117 cells and do not bind CD117 in gene edited cells.

EXAMPLE 3: Multiplex Editing

5 The guides identified in Example 2 may be used in methods for multiplex editing. Multiplex editing may be used, for example, for the treatment of hemoglobinopathy, such as sickle cell disease (SCD) or hereditary persistence of fetal hemoglobin (HPFH), hematologic cancer, or myeloproliferative disease.

10 In one experiment, guides 1-132 are used in combination with the HPFH guide along with the ABE8.8 editor to introduce an amino acid substitution in the CD117 cell surface protein for conditioning and in an HPFH polynucleotide for correcting a mutation associated with HPFH. In particular, 4μg of a guide selected from cc72, cc73, cc78, cc79, cc84, cc89, cc90, cc91, c93, cc102, cc103, cc110, cc114, cc119, or cc125 was used in combination with 4μg of the HPFH guide along with the ABE8.8 editor to induce A to G edits (FIG. 13A).
15 Single editing with no HPFH guide was used as a control. Similarly, HPFH editing was also tested using the dual guides at either 72 or 120 hours (FIG. 13B). Single editing with only an HPFH guide was used as a control.

In another experiment, guides 133-189 are used in combination with the Makassar guide along with the IBE-NGC editor to introduce an amino acid substitution in the CD117
20 cell surface protein for conditioning and in a beta globin (*HBB*) polynucleotide (Makassar) for correcting a mutation associated with sickle cell. The multiplex editing methods described herein can be further optimized.

EXAMPLE 4: Materials and Methods

25 General Methods:

Cloning can be conducted via USER enzyme (New England Biolabs) cloning methods (*see* Geu-Flores *et al.*, USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res* 35, e55, doi:10.1093/nar/gkm106 (2007)) and templates for PCR amplification were purchased as
30 bacterial or mammalian codon optimized gene fragments (GeneArt). Vectors created were transformed into Mach T1^R Competent Cells (ThermoFisher Scientific) and maintained at -80 C for long-term storage. Primers used in this work can be purchased from Integrated DNA Technologies and PCRS were carried out using either Phusion U DNA Polymerase Green MultiPlex PCR Master Mix (ThermoFisher) or Q5 Hot Start High-Fidelity 2x Master Mix

(New England Biolabs). Plasmids used in this work can be freshly prepared from 50 mL of Mach1 culture using ZymoPURE Plasmid Midiprep (Zymo Research Corporation) which involves an endotoxin removal procedure. Molecular biology grade, Hyclone water (GE Healthcare Life Sciences) was used in all assays, transfections, and PCR reactions to ensure exclusion of DNase activity.

Amino acid sequences of sgRNAs used for Hek293T mammalian cell transfection are provided in Table 23 below. When a target DNA sequence did not start with a 'G,' a 'G' was added to the 5' end of the primer since it has been established that the human U6 promoter prefers a 'G' at the transcription start site (*see* Cong, L. *et al.*, Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823, doi:10.1126/science.1231143 (2013)). The pFYF sgRNA plasmid described previously was used as a template for PCR amplification. Table 23 also identifies the nucleic acid target sequence, PAM sequence, and potential amino acid substitutions.

The spCas9 guide scaffold polynucleotide sequence used for the guide RNAs in Table 23 is as follows:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
CACCGAGTCGGTGC

Table 23: Sequences of sgRNAs used for Hek293T mammalian cell transfection.

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC1	AUGAAACGAAUGAGAAUAAG	ATGAAACGAATGAGAATAAG	CAG	E73G T74A N75D
Guide-CC2	UGGAUGAAACGAAUGAGAAU	TGGATGAAACGAATGAGAAT	AAG	D72G E73G T74A
Guide-CC3	GCCAGGGGAAGCCUCUCCCC	GCCAGGGGAAGCCTCTTCCC	AAG	Q152R K154E
Guide-CC4	UCUGUGUACUCAACGUGGAA	TCTGTGTACTCAACGTGGAA	AAG	Y243C
Guide-CC5	UGUGUACUCAACGUGGAAAA	TGTGTACTCAACGTGAAAA	GAG	Y243C
Guide-CC6	GGGAAGAUUAUCCCAAGUCU	GGGAAGATTATCCCAAGTCT	GAG	E360G D361G
Guide-CC7	CCAGAAAUCCUGACUUACGA	CCAGAAATCCTGACTTACGA	CAG	E414G I415V
Guide-CC8	UCUGAGAAUGAAAGUAAUUAU	TCTGAGAATGAAAGTAATAT	CAG	E366G N367G

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC9	CUAUUUUAACUUUGCAUUUA	CTATTTTAACTTTGCATTTA	AAG	Y503C N505G
Guide-CC10	UUCACGAGCCUGUCGUAAGU	TTCACGAGCCTGTCGTAAGT	CAG	L421P V422A
Guide-CC11	CAGCAGUGGAUCUAUAUGAA	CAGCAGTGGATCTATATGAA	CAG	Q347R
Guide-CC12	CAUCCCCCAAACCGAACAC	CATTCCCCAAACCTGAACAC	CAG	K342E
Guide-CC13	UAUCCCAAGUCUGAGAAUGA	TATCCCAAGTCTGAGAAATGA	AAG	K364G
Guide-CC14	AUAAAUGGGAAGAUUAUCCC	ATAAATGGGAAGATTATCCC	AAG	K358G
Guide-CC15	UCCCCAAACCGAACACCAG	TCCCCAAACCTGAACACCAG	CAG	K342G
Guide-CC16	GCCAUUCACGAGCCUGUCGU	GCCATTACGAGCCTGTCTGT	AAG	V422A
Guide-CC17	AAAUGUCACAACAACCUUGG	AAATGTCACAACAACCTTGG	AAG	N300S T302A
Guide-CC18	CGUGGAAAAGAGAAAACAGU	CGTGGAAAAGAGAAAACAGT	CAG	K247G R248G
Guide-CC19	CCAUUUGACAAAGCCCGGAU	CCATTTGACAAAGCCCGGAT	CAG	V64A W66R
Guide-CC20	AGCAAGCUAUCUUCUAGGG	AGCAAGCTATCTTCTTAGGG	AAG	S220G Y221C
Guide-CC21	UGAGGUUUAUUCUGACCCC	TGAGGTTTATTCCTGACCCC	AAG	R161G I163V
Guide-CC22	UGAAUGGAUCACGGAAAAGG	TGAATGGATCACGGAAAAGG	CAG	E81G I83V
Guide-CC23	CCACUGGCAGUACAGAAGCA	CCACTGGCAGTACAGAAGCA	GAG	L455P V457A
Guide-CC24	AGAAUGAAUGGAUCACGGAA	AGAATGAATGGATCACGGAA	AAG	N80G E81G
Guide-CC25	UGUAAACGAUGGAGAAAUG	TGTAAACGATGGAGAAAATG	TAG	N326G D327G
Guide-CC26	AGGCUUACAACGAUGUGGGC	AGGCTTACAACGATGTGGGC	AAG	Y494C N495D
Guide-CC27	AAGUGACCAAUUAUCCUC	AAGTGACCAATTATTCCTC	AAG	T144A N145D
Guide-CC28	UGUCACAACAACCUUGGAAG	TGTCACAACAACCTTGAAG	TAG	T302A T303A
Guide-CC29	AAUAGAUUGGUUUUUUGUC	AATAGATTGGTATTTTGTGTC	CAG	I438M D439G
Guide-CC30	ACAAUGGCACGGUUGAAUGU	ACAATGGCACGGTTGAATGT	AAG	N486G T488A
Guide-CC31	ACACAACAGGCACAGCUUUG	ACACAACAGGCACAGCTTTG	AAG	V213A V214A
Guide-CC32	GUUCAGAGUUCUAUAGAUGC	GTTTCAGAGTTCTATAGATTC	TAG	Q475R S476G
Guide-CC33	AGAUUUGAUUGUUGAAUAUG	AGATTTGATTGTTGAATATG	AAG	D332G I334V
Guide-CC34	AUAAAACAUUAAUUGCUAUGG	ATAAACATTAAATGCTATGG	CAG	V407A Y408H
Guide-CC35	AUCCAUCCAUCAGGAAAAU	ATCCATCCATCCAGGAAAAAT	CAG	I39V H40R
Guide-CC36	AUGGAUCACGGAAAAGGCAG	ATGGATCACGGAAAAGGCAG	AAG	I83V T84A

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC37	GUUGACUAUCAGUUCAGCGA	GTTGACTATCAGTTCAGCGA	GAG	T276A I277V
Guide-CC38	CACAACAACCUUGGAAGUAG	CACAACAACCTTGGAAAGTAG	TAG	T303A T304A
Guide-CC39	CUAUAGAUUCUAGUGCAUUC	CTATAGATTCTAGTGCATTTC	AAG	I478V D479G
Guide-CC40	UGCCAAUAAUACUUUUGGAU	TGCCAATAATACTTTTGGAT	CAG	N293G N294G
Guide-CC41	GCUUUGUCAAAUGGACUUUU	GCTTTGTCAAATGGACTTTT	GAG	K65E
Guide-CC42	CUUCCCCAUGAUAAACACUA	CTTCCCCATGATAAACACTA	CAG	M318V
Guide-CC43	CGUUUGGAAAGCUAGUGGUU	CGTTTGGAAAGCTAGTGGTT	CAG	K471E
Guide-CC44	AAACUACAGGAGAAAAUUA	AAACTACAGGAGAAATATAA	TAG	Q256R
Guide-CC45	AGGAGGCACUACACAUUCC	AGGAGGCACCTACACATTCC	TAG	T389A
Guide-CC46	CUUGAAUGCACUAGAAUCUA	CTTGAATGCACTAGAATCTA	TAG	F483P
Guide-CC47	UCCAAACGGUGGCCCAGAU	TTCCAAACGGTGGCCCAGAT	GAG	F469P
Guide-CC48	UCCAGGAAAAUCAGACUUA	TCCAGGAAAATCAGACTTAA	TAG	K43G
Guide-CC49	CCCUCAAGGGGUGCCAGGGG	CCCTCAAGGGGTGCCAGGGG	AAG	K149G
Guide-CC50	UCCAAAGCAAGCUAUCUUCU	TCCAAAGCAAGCTATCTTCT	TAG	K218G
Guide-CC51	UAGGGAAGGGGAAGAAUUC	TAGGGAAGGGGAAGAATTCA	CAG	E225G
Guide-CC52	AGCGAGAGUUAUGAUUCUG	AGCGAGAGTTAATGATTCTG	GAG	R281G
Guide-CC53	UUUGGAAAGCUAGUGGUUCA	TTTGGAAAGCTAGTGGTTCA	GAG	K471G
Guide-CC54	AGACGGUCCCCUGGACUCA	AGACGGTTCCCCCTGGACTCA	CAG	S35P
Guide-CC55	ACGUUGACUAUCAGUUCAGC	ACGTTGACTATCAGTTCAGC	GAG	T276A
Guide-CC56	CUGCACAUCCACUGGCAGUA	CTGCACATCCACTGGCAGTA	CAG	V459A
Guide-CC57	CACAUCCACUGGCAGUACAG	CACATCCACTGGCAGTACAG	AAG	V457A
Guide-CC58	AUCCACUGGCAGUACAGAAG	ATCCACTGGCAGTACAGAAG	CAG	V457A
Guide-CC59	UUGACGUCAGAAUUGGACAC	TTGACGTCAGAATTGGACAC	TAG	V399A
Guide-CC60	AUUAACUCUCGUGAACUGA	ATTAACCTCTCGCTGAAGTGA	TAG	V282A
Guide-CC61	CCAUACAAGGAGCGGUCAAC	CCATACAAGGAGCGGTCAAC	AAG	L124P Y125H
Guide-CC62	UUCAGGAUGAAUUUUUCCGA	TTCAGGATGAATTTTCCGA	CAG	I201T L202P
Guide-CC63	GUGACUUCAAUUAUGAACGU	GTGACTTCAATTATGAACGT	CAG	D266G N268D
Guide-CC64	GUCGGAAAAAUUCAUCCUGA	GTCGGAAAAATTCATCCTGA	AAG	E198G K199G
Guide-CC65	AUUAUAUUUCUCCUGUAGUU	ATTATATTTCTCCTGTAGTT	TAG	Y259H
Guide-CC66	AAGGAGCGGUCAACAAGGAA	AAGGAGCGGTCAACAAGGAA	AAG	S123P
Guide-CC67	GAGAUUAAAAGGCACCGAAG	GAGATTAAAAGGCACCGAAG	GAG	K383G
Guide-CC68	AACGAGAUUAAAAGGCACCG	AACGAGATTAAAAGGCACCG	AAG	R381G
Guide-CC69	GAAAAAUUCAUCCUGAAAGU	GAAAAATTCATCCTGAAAGT	GAG	K199G
Guide-CC70	AUUCAUCCUGAAAGUGAGGC	ATTATCCTGAAAGTGAGGC	CAG	I201V
Guide-CC71	GAAAGAAGACAACGACACGC	GAAAGAAGACAACGACACGC	TGG	K127R E128G D129G

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC72	UACAGGAGAAAUAUAAUAGC	TACAGGAGAAATATAATAGC	TGG	Q256R E257G K258E
Guide-CC73	AUGAGAAUAAGCAGAAUGAA	ATGAGAATAAGCAGAATGAA	TGG	E76G N77G K78E
Guide-CC74	CCAGGGGAAGCCUCUCCCCA	CCAGGGGAAGCCTCTTCCCA	AGG	Q152R K154G
Guide-CC75	CAGAAAUCCUGACUUACGAC	CAGAAATCCTGACTTACGAC	AGG	E414G I415V
Guide-CC76	ACAGAACCUUCACUGAUAAA	ACAGAACCTTCACTGATAAA	TGG	R353G T354A
Guide-CC77	UCACGAGCCUGUCGUAAGUC	TCACGAGCCTGTCGTAAGTC	AGG	L421P V422A
Guide-CC78	ACUUACGACAGGCUCUGGAA	ACTTACGACAGGCTCGTGAA	TGG	Y418C D419G
Guide-CC79	UGUUCAUUAUAGAUCACUGC	TGTTCATATAGATCCACTGC	TGG	M351T
Guide-CC80	CCAAACCUGAACACCAGCAG	CCAAACCTGAACACCAGCAG	TGG	K342G
Guide-CC81	CUGUCGUAAGUCAGGAUUUC	CTGTCTGTAAGTCAGGATTTC	TGG	Y418H
Guide-CC82	CAGAACCUUCACUGAUAAA	CAGAACCTTCACTGATAAAT	GGG	T354A
Guide-CC83	UAUUUUAACUUUGCAUUUAA	TATTTTAACTTTGCATTTAA	AGG	N505G
Guide-CC84	UAAGCAGAAUGAAUGGAUCA	TAAGCAGAATGAATGGATCA	CGG	K78R Q79R N80G
Guide-CC85	GAAUGGCAUGCUCCAAUGUG	GAATGGCATGCTCCAATGTG	TGG	N423S M425V
Guide-CC86	GUGCAGACACUAAACUCAUC	GTGCAGACACTAAACTCATC	TGG	Q460R T461A
Guide-CC87	UGACCAAUUAUUCCCUCAAG	TGACCAATTATTCCCTCAAG	GGG	T144A N145G
Guide-CC88	GUGACCAAUUAUUCCCUCAA	GTGACCAATTATTCCCTCAA	GGG	T144A N145G
Guide-CC89	AGUGACCAAUUAUUCCCUCA	AGTGACCAATTATTCCCTCA	AGG	T144A N145G
Guide-CC90	GAAUGAAUGGAUCACGGAAA	GAATGAATGGATCACGGAAA	AGG	N80S E81G
Guide-CC91	UGCAGACACUAAACUCAUCU	TGCAGACACTAAACTCATCT	GGG	Q460R T461A
Guide-CC92	CCAAAGCAAGCUAUCUUCUU	CCAAAGCAAGCTATCTTCTT	AGG	K218G S220G
Guide-CC93	ACAGCACUGACUUGCCCUCC	ACAGCACTGACTTGCCCTCC	TGG	V195A L196P
Guide-CC94	GGAGGGCAAGUCAGUGCUGU	GGAGGGCAAGTCAGTGCTGT	CGG	E191G K193G
Guide-CC95	CACAUAAACAUAUAAUUGCUA	CACATAAACATTAAATGCTA	TGG	V407A Y408H
Guide-CC96	CAAAUGGACUUUUGAGAUCC	CAAATGGACTTTTGAGATCC	TGG	K65R T67A
Guide-CC97	CAAUGGCACGGUUGAAUGUA	CAATGGCACGGTTGAATGTA	AGG	N486S T488A

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC98	AAAUACUGUAGUGUUUAUCA	AAATACTGTAGTGTATTATCA	TGG	V323A F324L
Guide-CC99	CAAAGCAAGCUAUCUUCUUA	CAAAGCAAGCTATCTTCTTA	GGG	K218R S220G
Guide-CC100	GGAAUAAAACCUCAAGUCCUU	GGAATAAACCTCAAGTCCTT	GGG	F162P I163T
Guide-CC101	ACCACUAGCUUUCCAAACGG	ACCACTAGCTTTCCAAACGG	TGG	L472P V473A
Guide-CC102	GCAAGCUAUCUUCUAGGGA	GCAAGCTATCTTCTTAGGGA	AGG	S220G Y221C
Guide-CC103	CAAGCUAUCUUCUAGGGAA	CAAGCTATCTTCTTAGGGAA	GGG	S220G Y221C
Guide-CC104	UUUGAUC AUGAUGCCCGCCU	TTTGATCATGATGCCCCGCTT	TGG	M171T I172T
Guide-CC105	UUGAUC AUGAUGCCCGCCUU	TTGATCATGATGCCCCGCTT	GGG	M171T I172T
Guide-CC106	AUAGAUUGGUAUUUUUGUCC	ATAGATTGGTATTTTGTGCC	AGG	I438M D439G
Guide-CC107	AGGAAUAAAACCUCAAGUCCU	AGGAATAAACCTCAAGTCCT	TGG	F162P I163T
Guide-CC108	CACAACAGGCACAGCUUUGA	CACAACAGGCACAGCTTTGA	AGG	V213A
Guide-CC109	UGACGUCAGAAUUGGACACU	TGACGTCAGAATTGGACACT	AGG	V399A
Guide-CC110	ACUACAGUAUUUGUAAACGA	ACTACAGTATTTGTAAACGA	TGG	T322A
Guide-CC111	AGCAAAUGUCACAACAACCU	AGCAAATGTCACAACAACCT	TGG	N300G
Guide-CC112	GUCUGAUUUUCCUGGAUGGA	GTCTGATTTTCCTGGATGGA	TGG	S44P
Guide-CC113	UGCUGAUCCAAAAGUAUUUAU	TGCTGATCCAAAAGTATTAT	TGG	S298P
Guide-CC114	AUCAGACUAAUAGUCCGCG	ATCAGACTTAATAGTCCGCG	TGG	D45G
Guide-CC115	UGCAUUC AAGCACAAUGGCA	TGCATTCAAGCACAAATGGCA	CGG	K484G
Guide-CC116	UCCCAGAGCCCACAAUAGAU	TCCCAGAGCCCACAATAGAT	TGG	E435G
Guide-CC117	UCAGCGAGAGUUAUUGAUUC	TCAGCGAGAGTTAATGATTC	TGG	R281G
Guide-CC118	GUGUGAAACGCGCCUACCAU	GTGTGAAACGCGCCTACCAT	CGG	K176G
Guide-CC119	UCAGACUAAUAGUCCGCGU	TCAGACTTAATAGTCCGCGT	GGG	D45G
Guide-CC120	UCUAGUGCAUUC AAGCACAA	TCTAGTGCATTCAAGCACAA	TGG	S481G
Guide-CC121	AUUCCCUCAAGGGGUGCCAG	ATTCCCTCAAGGGGTGCCAG	GGG	K149E
Guide-CC122	AAGGCAGAAGCCACCAACAC	AAGGCAGAAGCCACCAACAC	CGG	E88G
Guide-CC123	GAGGUUU AUUCCUGACCCCA	GAGGTTTATTCCTGACCCCA	AGG	I163V
Guide-CC124	UUUAUUCUGACCCCAAGGC	TTTATTCCTGACCCCAAGGC	GGG	I163V
Guide-CC125	GUUU AUUCCUGACCCCAAGG	GTTTATTCCTGACCCCAAGG	CGG	I163V
Guide-CC126	CAUACAAGGAGCGGUCAACA	CATACAAGGAGCGGTCAACA	AGG	L124P Y125H
Guide-CC127	AAAAAUUCAUCCUGAAAGUG	AAAAATTCATCTGAAAGTG	AGG	K199G I201V
Guide-CC128	AAAUUAUAGCUGGCAUCA	AAATATAATAGCTGGCATCA	CGG	Y259C N260G
Guide-CC129	UGACUUC AAUUAUGAACGUC	TGACTTCAATTATGAACGTC	AGG	D266G N268G
Guide-CC130	CUAGUUCUGUGUACUCAACG	CTAGTTCTGTGTACTCAACG	TGG	S240G
Guide-CC131	AGAUUAAAAGGCACCGAAGG	AGATTA AAAAGGCACCGAAGG	AGG	K383G

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC132	ACGAGAUUAAAAGGCACCGA	ACGAGATTAAAAGGCACCGA	AGG	R381G
Guide-CC133	GGAUGAAACGAAUGAGAAUA	GGATGAAACGAATGAGAATA	AGC	E73G T74A N75G
Guide-CC134	AACUACAGGAGAAAUAUAAU	AACTACAGGAGAAATATAAT	AGC	Q256R E257G K258G
Guide-CC135	UGGGAAAGAAGACAACGACA	TGGGAAAGAAGACAACGACA	CGC	K127G E128G D129G
Guide-CC136	GCCAAUAAUACUUUUGGAUC	GCCAATAATACTTTTGGATC	AGC	N293G N294G T295A
Guide-CC137	CUUACGACAGGCUCGUGAAU	CTTACGACAGGCTCGTGAAT	GGC	Y418C D419G R420G
Guide-CC138	GUUAAAAUAGGCAGAAGUCU	GTTAAAATAGGCAGAAGTCT	TGC	Y503H F504P
Guide-CC139	UUCUGUUCAUAUAGAUCAC	TTCTGTTCATATAGATCCAC	TGC	Y350H M351T
Guide-CC140	AGAAAUCCUGACUUACGACA	AGAAATCCTGACTTACGACA	GGC	I415V T417A
Guide-CC141	AACGAUGUGGGCAAGACUUC	AACGATGTGGGCAAGACTTC	TGC	D496G K499E
Guide-CC142	AUUCGCCAAACCGAAGACAC	ATTCCCCAAACCTGAACACC	AGC	K342G
Guide-CC143	CGACAGGCUCGUGAAUGGCA	CGACAGGCTCGTGAATGGCA	TGC	R420G
Guide-CC144	ACUUCUGCCUAUUUUAACUU	ACTTCTGCCTATTTTAACTT	TGC	Y503C
Guide-CC145	GCGACGAGAUUAGGCUGUUA	GCGACGAGATTAGGCTGTTA	TGC	D52G E53G I54V R55G
Guide-CC146	AAGACAACGACACGCUGGUC	AAGACAACGACACGCTGGTC	CGC	D129G N130G D131G T132A
Guide-CC147	UGGAUCACGGAAAAGGCAGA	TGGATCACGGAAAAGGCAGA	AGC	I83V T84A E85G K86E
Guide-CC148	GACACAGACACAACAGGCAC	GACACAGACACAACAGGCAC	AGC	V213A V214A S215P V216A
Guide-CC149	GGGAAGAAUUCACAGUGACG	GGGAAGAATTACAGTGACG	TGC	E227G E228G T230A
Guide-CC150	GCACCAACAAACACGGCUUA	GCACCAACAAACACGGCTTA	AGC	N99G K100G H101R
Guide-CC151	UAUAGAUUCUAGUGCAUUA	TATAGATTCTAGTGCATTCA	AGC	I478M D479G S481G

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC152	UCAUGAUCAAAAGUGUGAAA	TCATGATCAAAAGTGTGAAA	CGC	I172V K173G S174G
Guide-CC153	AUGAUCAAAAGUGUGAAACG	ATGATCAAAAGTGTGAAACG	CGC	I172V K173G S174G
Guide-CC154	AAUGAAUGGAUCACGGAAAA	AATGAATGGATCACGGAAAA	GGC	E81G I83V T84A
Guide-CC155	CCAAUUAUCCCUCAAGGGG	CCAATTATTCCTCAAGGGG	TGC	N145S Y146C
Guide-CC156	CAGACUAAUAGUCCGCGUG	CAGACTTAATAGTCCGCGTG	GGC	D45G I47V
Guide-CC157	UCUGACGUCAAUGCUGCCAU	TCTGACGTCAATGCTGCCAT	AGC	D398G N400G
Guide-CC158	GCAUAAACAGCCUAAUCUCGU	GCATAACAGCCTAATCTCGT	CGC	L56P L57P
Guide-CC159	UUUUUCCGACAGCACUGACU	TTTTTCCGACAGCACTGACT	TGC	L196P S197P
Guide-CC160	UUUCCGUGAUCCAUAUCAUUC	TTTCCGTGATCCATTCATTC	TGC	W82R I83T
Guide-CC161	UUUGACAAAGCCCGGAUCAG	TTTGACAAAGCCCGGATCAG	TGC	F63P V64A
Guide-CC162	GCCCAGAUGAGUUUAGUGUC	GCCCAGATGAGTTTAGTGTC	TGC	S464P S465P
Guide-CC163	GCUGAACUGAUAGUCAACGU	GCTGAACTGATAGTCAACGT	TGC	I277T S279P
Guide-CC164	ACACUUUUGAUCAUGAUGCC	ACACTTTTGATCATGATGCC	CGC	M171T I172T
Guide-CC165	GAAAAUCAGACUAAUAGUC	GAAATCAGACTTAATAGTC	CGC	K43R D45G
Guide-CC166	GCAGACACUAAACUCAUCUG	GCAGACACTAACTCATCTG	GGC	T461A N463G
Guide-CC167	GGACCAGGAGGGCAAGUCAG	GGACCAGGAGGGCAAGTCAG	TGC	Q190R E191G
Guide-CC168	GCUGAUCCAAAAGUAUUAUU	GCTGATCCAAAAGTATTATT	GGC	F296P S298P
Guide-CC169	ACAUCCACUGGCAGUACAGA	ACATCCACTGGCAGTACAGA	AGC	L455P V457A
Guide-CC170	CACUGGCAGUACAGAAGCAG	CACTGGCAGTACAGAAGCAG	AGC	V454A L455P
Guide-CC171	AGGCAGAAGCCACCAACACC	AGGCAGAAGCCACCAACACC	GGC	E88G T90A
Guide-CC172	AGGUUUAUCCUGACCCCAA	AGGTTTATTCCTGACCCCAA	GGC	I163V
Guide-CC173	CCAAGCUUUUCCUUGUUGAC	CCAAGCTTTTCCTTGTTGAC	CGC	K116R
Guide-CC174	UUAGUGUCUGCACAUCCACU	TTAGTGTCTGCACATCCACT	GGC	V459A
Guide-CC175	CCACUAGCUUCCAAACGGU	CCACTAGCTTTCCAAACGGT	GGC	L472P
Guide-CC176	CCUCAAGGGGUGCCAGGGGA	CCTCAAGGGGTGCCAGGGGA	AGC	K149G
Guide-CC177	UGUGAAACGCGCCUACCAUC	TGTGAAACGCGCCTACCATC	GGC	K176G
Guide-CC178	CUAGUGCAUUAAGCACAAU	CTAGTGCATTCAAGCACAAT	GGC	K484G
Guide-CC179	CCGCGUGGGCGACGAGAUUA	CCGCGTGGGCGACGAGATTA	GGC	D52G

Guide No.	Guide Sequence	Target Sequence	PAM	Mutations
Guide-CC180	UUAUUCUGACCCCAAGGCG	TTATTCCTGACCCCAAGGCG	GGC	D165G
Guide-CC181	ACUCCAGAAUCAUUAACUCU	ACTCCAGAATCATTAACCTCT	CGC	S285P
Guide-CC182	GGCUGUUAUGCACUGAUCCG	GGCTGTTATGCACTGATCCG	GGC	T59A
Guide-CC183	UUCAUCUAAACGAGAUUAAAA	TTCATCTAACGAGATTAAAA	GGC	H378R T380A R381G
Guide-CC184	ACAGGAGAAAUAUAAUAGCU	ACAGGAGAAATATAATAGCT	GGC	E257G K258G Y259C
Guide-CC185	UCAGGAUGAAUUUUUCCGAC	TCAGGATGAATTTTTCGAC	AGC	F200P I201T
Guide-CC186	GACUUCAAUUAUGAACGUCA	GACTTCAATTATGAACGTCA	GGC	N268G Y269C
Guide-CC187	AGGAGCGGUCAACAAGGAAA	AGGAGCGGTCAACAAGGAAA	AGC	V120A S123P
Guide-CC188	GAUUAAAAGGCACCGAAGGA	GATTAAAAGGCACCGAAGGA	GGC	K383G T385A
Guide-CC189	AAAAUUCAUCCUGAAAGUGA	AAAATTCATCCTGAAAGTGA	GGC	I201V

General HEK293T mammalian culture conditions

Cells were cultured at 37 °C with 5% CO₂. HEK293T cells [CLBTx013, American Type Cell Culture Collection (ATCC)] were cultured in Dulbecco's modified Eagles medium plus Glutamax (10566-016, Thermo Fisher Scientific) with 10% (v/v) fetal bovine serum (A31606-02, Thermo Fisher Scientific). Cells were tested negative for mycoplasma after receipt from supplier.

Hek293T plasmid transfection and gDNA extraction

HEK293T cells were seeded onto 48-well well Poly-D-Lysine treated BioCoat plates (Corning) at a density of 35,000 cells/well and transfected 18-24 hours after plating. Cells were counted using a NucleoCounter NC-200 (Chemometec). To these cells were added 750 ng of base editor or nuclease control, 250 ng of sgRNA, and 10 ng of GFP-max plasmid (Lonza) diluted to 12.5 μ L total volume in Opti-MEM reduced serum media (ThermoFisher Scientific). The solution was combined with 1.5 μ L of Lipofectamine 2000 (ThermoFisher) in 11 μ L of Opti-MEM reduced serum media and left to rest at room temperature for 15 min. The entire 25 μ L mixture was then transferred to the pre-seeded Hek293T cells and left to incubate for about 120 h. Following incubation, media was aspirated and cells were washed two times with 250 μ L of 1x PBS solution (ThermoFisher Scientific) and 100 μ L of freshly prepared lysis buffer was added (100 mM Tris-HCl, pH 7.0, 0.05% SDS, 25 μ g/mL Proteinase K (Thermo Fisher Scientific). Transfection plates containing lysis buffer were

incubated at 37 °C for 1 hour and the mixture was transferred to a 96-well PCR plate and heated at 80 °C for 30 min.

Analysis of DNA and RNA off-target editing

5 HEK293T cells were plated on 48-well poly-D-lysine coated plates (Corning) 16 to 20 hours before lipofection at a density of 30,000 cells per well in DMEM + Glutamax medium (Thermo Fisher Scientific) without antibiotics. 750 ng nickase or base editor expression plasmid DNA was combined with 250ng of sgRNA expression plasmid DNA in 15 µl OPTIMEM + Glutamax. This was combined with 10 µl of lipid mixture, comprising 1.5
10 µl Lipofectamine 2000 and 8.5 µl OPTIMEM + Glutamax per well. Cells were harvested 3 days after transfection and either DNA or RNA was harvested. For DNA analysis, cells were washed once in 1X PBS, and then lysed in 100 µl QuickExtract™ Buffer (Lucigen) according to the manufacturer's instructions. For RNA harvest, the MagMAX™ mirVana™ Total RNA Isolation Kit (Thermo Fisher Scientific) was used with the KingFisher™ Flex Purification
15 System according to the manufacturer's instructions.

Targeted RNA sequencing was performed largely as previously described (*see* Rees, H. A. *et al.*, Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv* **5**, eaax5717, doi:10.1126/sciadv.aax5717 (2019)). cDNA was prepared from the isolated RNA using the SuperScript IV One-Step RT-PCR System with EZDnase
20 (Thermo Fisher Scientific) according to the manufacturer's instructions. The following program was used: 58 °C for 12 min; 98°C for 2 min; followed by PCR cycles which varied by amplicon: for CTNNB1 and IP90: 32 cycles of [98°C for 10 sec; 60°C for 10 sec; 72°C for 30 sec] and for RSL1D1 35 cycles of [98°C for 10 sec; 58°C for 10 sec; 72°C for 30 sec]. No RT controls were run concurrently with the samples. Following the combined RT-PCR,
25 amplicons were barcoded and sequenced using an Illumina Miseq as described above.

Off-target DNA sequencing was performed using previously published primers (*see* Komor, A. C. *et al.*, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016); Rees, H. A. *et al.*, Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci*

Adv 5, eaax5717, doi:10.1126/sciadv.aax5717 (2019)) using a two-step PCR and barcoding method to prepare samples for sequencing using Illumina Miseq sequencers as above.

Table 24: HTS Primers used to amplify genomic sites

5

Primer Name	Sequence
fwd_site_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAGCCCCATCTGTCAAACCT
rev_site_1	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGAATGGATTCCCTGGAAACAATGA
fwd_site_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGAGGGAGAGCCGTGTAGTT
rev_site_2	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCTCTCAAAGTGCTGGGAT
fwd_site_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCATCAGGCTCTCAGCTCAG
rev_site_3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGGGTTTGTGGTTGC
fwd_site_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCATTCCTCTTTAGCCA
rev_site_4	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGCCGTTCCCTCTTTGCTA
fwd_site_5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAACCTGTGTGACACTTGGCA
rev_site_5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGGCCCAAGATCACACA
fwd_site_6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCACGGATAAAGACGCTGGGA
rev_site_6	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTCCCAGGTGCTGAC
fwd_site_7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNTTGATTGTCTCCTTTGCCGC
rev_site_7	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGACCCAGTGTGATAGATCAGT
fwd_site_8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCACCCCTTCAGTCCATGCTT
rev_site_8	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGATGGGGAGGAACGAGT
fwd_site_9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCTCAGCCTGAGTGTTGA
rev_site_9	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCCACCCTAGTCATTGGAG
fwd_site_10	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCAGAGGGACACACTGTGG
rev_site_10	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACACTCACTCACCCACACA
fwd_site_11	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTGTGTGGGTGAGTGAGTGTG
rev_site_11	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACCAAGGTTACAGCCTGA

fwd_site_12	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGTCTCTGCCTGTAGCTGC
rev_site_12	TGGAGTTCAGACGTGTGCTCTTCCGATCTCGCTCTGGGCTTCATCTTCA
fwd_site_13	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGGGATTATGGGTGTGAGCC
rev_site_13	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGCTTCCTCCTCTCTCTCC
fwd_site_14	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGCAGACCAGATTCGGAGAA
rev_site_14	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTCAGTTTCCAGGGGGTCC
fwd_site_15	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTCCGCACAGCCTTAGTTCAA
rev_site_15	TGGAGTTCAGACGTGTGCTCTTCCGATCTAACTTGAAGAGACGGCAGCA
fwd_site_16	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCCCAGCTACAGAAAGGTC
rev_site_16	TGGAGTTCAGACGTGTGCTCTTCCGATCTATTTCCACCGCAAAATGGCC
fwd_site_17	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCACTTCAGCCCAGGAGTAT
rev_site_17	TGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGTATGGTGAGAGGTAGGGA
fwd_site_18	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTCTGAGGTCACACAGTGGG
rev_site_18	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGAGCAGGGACCACATC
fwd_site_19	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAGGTGGAGAGAGGATGT
rev_site_19	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTCTTCCTGAGGTCTAGGAACCCG
fwd_site_20	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCTGTTCTAAAGCCCACC
rev_site_20	TGGAGTTCAGACGTGTGCTCTTCCGATCTACTCTCTGGTTCTGTTTGTGGCCA
fwd_CTNNB1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNATTTGATGGAGTTGGACATG GCC
rev_CTNNB1	TGGAGTTCAGACGTGTGCTCTCCAGCTACTTGTCTTGAGTGAAGG
fwd_RSL1D1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGGCTTTCCAAATCAGTGGG TC
rev_RSL1D1	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTCATAAGCTTAGACCAACAAGC
fwd_IP90	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTGGTTGACCAATCTGTGGT G
rev_IP90	TGGAGTTCAGACGTGTGCTCTCTGCGTCTGGATCAGGTACG
fwd_HEK293_site2_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGTGGAGAGTGAGTAAGCC A
rev_HEK293_site2_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTACGGTAGGATGATTTTCAGGCA

fwd_HEK293_site2_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCACAAAGCAGTGTAGCTCAGG
rev_HEK293_site2_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTTTGGTACTCGAGTGTTATTCAG
fwd_HEK293_site3_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTCCCCTGTTGACCTGGAGAA
rev_HEK293_site3_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTCACTGTACTTGCCCTGACCA
fwd_HEK293_site3_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTTGGTGTGACAGGGAGCAA
rev_HEK293_site3_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGATGTGGGCAGAAGGG
fwd_HEK293_site3_off3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGAGAGGGAACAGAAGGGCT
rev_HEK293_site3_off3	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCAAAGGCCAAGAACCT
fwd_HEK293_site3_off4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTAGCACTTTGGAAGGTCG
rev_HEK293_site3_off4	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCATCTTAATCTGCTCAGCC
fwd_HEK293_site3_off5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNAAAGGAGCAGCTCTTCCTGG
rev_HEK293_site3_off5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTGCACCATCTCCCACAA
fwd_HEK293_site4_off1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGCATGGCTTCTGAGACTCA
rev_HEK293_site4_off1	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCCTTGCACTCCCTGTCTTT
fwd_HEK293_site4_off2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTTGGCAATGGAGGCATTGG
rev_HEK293_site4_off2	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGAGGCTGCCCATGAGAG
fwd_HEK293_site4_off3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGTCTGAGGCTCGAATCCTG
rev_HEK293_site4_off3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTGTGGCCTCCATATCCCTG
fwd_HEK293_site4_off4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTTCCACCAGAACTCAGCCC
rev_HEK293_site4_off4	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCGGTTCTCCACAACAC
fwd_HEK293_site4_off5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCACGGAAGGACAGGAGAAG
rev_HEK293_site4_off5	TGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGGGGAGGGATAAAGCAG
fwd_HEK_site_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGAAACGCCCATGCAATTAGTC
rev_HEK_site_3	TGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGTCAACCAGTATCCCGGTG
fwd_HEK_site_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGAATGGATTCTTTGGAAACAATG

rev_HEK_site_2	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGCCCCATCTGTCAAACCT
fwd_HEK_site_4	TGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTTCAACCCGAACGGAG
rev_HEK_site_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTGGTCTTCTTTCCCTCC
fwd_LDLR	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCCCTGCTTCTTTTTCTCTGT GT
rev_LDLR	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCATTAAACGCAGCCAACTTCA
fwd_TRAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGAGGTCTATGGACTTCAAGAG CAA
Rev_TRAC	TGGAGTTCAGACGTGTGCTCTTCCGATCTCATCATTGACCAGAGCTCTGGGCAGAA
fwd_CBLB	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCACTTACCAGCATTACTTCTTAA ACC
Rev_CBLB	TGGAGTTCAGACGTGTGCTCTTCCGATCTATGGGCTCCACTTTTCAGCTCTGTAA
fwd_CD7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTTCAGGCACATGTAGGAGGGA
Rev_CD7	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCGCCTGCAGCTGTCTGGACACTGGCA
fwd_B2M	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAAGATGAGTATGCCTGCCGTG
Rev_B2M	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGATTGTTTATATCAGATGGGATGGG
fwd_CIITA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCAAGTTTGGTCTGAGCCCTC CC
Rev_CIITA	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTGGGTCCCTGCGCTCTGCA
fwd_PDCD1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAGGGACTGAGGGTGGAAGGTCC
Rev_PDCD1	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCTCCGCCTGAGCAGTGGAGAA

mRNA production for ABE editors used in CD34+ cells

- Editors were cloned into a plasmid encoding a dT7 promoter followed by a 5'UTR,
- 5 Kozak sequence, ORF, and 3'UTR. The dT7 promoter carries an inactivating point mutation within the T7 promoter that prevents transcription from circular plasmid. This plasmid templated a PCR reaction (Q5 Hot Start 2X Master Mix), in which the forward primer corrected the SNP within the T7 promoter and the reverse primer appended a 120A tail to the 3' UTR. The resulting PCR product was purified on a Zymo Research 25µg DCC column
- 10 and used as mRNA template in the subsequent in vitro transcription. The NEB HiScribe High-Yield Kit was used as per the instruction manual but with full substitution of N1-methyl-pseudouridine for uridine and co-transcriptional capping with CleanCap AG

(Trilink). Reaction cleanup was performed by lithium chloride precipitation. Primers used for amplification can be found in **Table 25**.

Table 25: Primers used for ABE8 T7 in vitro transcription reactions

Name	Sequence
fwd_IVT	TCGAGCTCGGTACCTAATACGACTCAC
rev_IVT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTCTCCTACTCAGGCTTTATTCAAAGACCA

CD34+ cell preparation

Mobilized peripheral blood was obtained and enriched for Human CD34⁺ HSPCs (HemaCare, M001F-GCSF/MOZ-2). The CD34⁺ cells were thawed and put into X-VIVO 10 (Lonza) containing 1% Glutamax (Gibco), 100ng/mL of TPO (Peprotech), SCF (Peprotech) and Flt-3 (Peprotech) at 48 hours prior to electroporation

Electroporation of CD34⁺ cells

48 hours post thaw, the cells were spun down to remove the X-VIVO 10 media and washed in MaxCyte buffer (HyClone) with 0.1% HSA (Akron Biotechnologies). The cells were then resuspended in cold MaxCyte buffer at 1,250,000 cells per mL and split into multiple 20 μ L aliquots. The ABE mRNA (0.15 μ M) and -198 HBG1/2 sgRNA (4.05 μ M) were then aliquoted as per the experimental conditions and raised to a total of 5 μ L in MaxCyte buffer. The 20 μ L of cells was added into the 5 μ L RNA mixture in groups of 3 and loaded into each chamber of an OC25x3 MaxCyte cuvette for electroporation. After receiving the charge, 25 μ L was collected from the chambers and placed in the center of the wells in a 24-well untreated culture plate. The cells recovered for 20 minutes in an incubator (37°C, 5% CO₂). After the 20 minutes recovery, X-VIVO 10 containing 1% Glutamax, 100ng/mL of TPO, SCF and Flt-3 was added to the cells for a concentration of 1,000,000 cells per mL. The cells were then left to further recover in an incubator (37°C, 5% CO₂) for 48hrs.

Erythrocyte differentiation post ABE electroporation

Following 48 h post electroporation rest (day 0 of culture), the cells were spun down and moved to “Phase 1” IMDM media (ATCC) containing 5% human serum, 330µg/mL transferrin (Sigma), 10µg/mL human insulin (Sigma), 2U/mL heparin sodium (Sigma), 3U/mL EPO (Peprotech), 100ng/mL SCF (Peprotech), 5µg/mL IL3 and 50µM hydrocortisone (Sigma) at 20,000 cells per mL. On day 4 of culture, the cells were fed 4x volume of the same media. On day 7, the cells were spun down and moved to “Phase 2” IMDM media containing 5% human serum (Sigma), 330µg/mL transferrin, 10µg/mL human insulin, 2U/mL heparin sodium, 3U/mL EPO and 100ng/mL SCF at 200,000 cells per mL. On day 11, cells were spun down and moved to “Phase 3” IMDM media containing 5% human serum, 330µg/mL of transferrin, 10µg/mL human insulin, 2U/mL of heparin sodium and 3U/mL of EPO at 1,000,000 cells per mL. On day 14, the cells were spun down and resuspended in the same media as day 11 but at 5,000,000 cells per mL. On day 18, the differentiated red blood cells were collected in 500,000 cell aliquots, washed once in 500µL DPBS (Gibco) and frozen at -80°C for 24 hours before UHPLC processing.

15

Genomic DNA extraction for CD34+ cells

Following ABE electroporation (48h later), an aliquot of cells was cultured in X-VIVO 10 media (Lonza) containing 1% Glutamax (Gibco), 100ng/mL of TPO (Peprotech), SCF (Peprotech) and Flt-3 (Peprotech). Following 48 h and 144 h post culturing, 100,000 cells were collected and spun down. 50 µL of Quick Extract (Lucigen) was added to the cell pellet and the cell mixture was transferred to a 96-well PCR plate (Bio-Rad). The lysate was heated for 15 minutes at 65°C followed by 10 minutes at 98°C. The cell lysates were stored at -20°C.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference.

Claims:

1. An *in vitro* or *ex vivo* method of producing a hematopoietic stem cell or progenitor thereof for the treatment of a hemoglobinopathy, hematologic cancer, or myeloproliferative disease, the method comprising:

(a) expressing in the hematopoietic stem cell or progenitor thereof a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; and

(b) contacting the hematopoietic stem cell or progenitor thereof with a guide RNA that targets a nucleic acid molecule encoding a CD117 protein, and introducing a T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, or T488A mutation in the CD117 protein.

2. Use of a hematopoietic stem cell or progenitor thereof in the preparation of a medicament for the treatment of a hemoglobinopathy, hematologic cancer, or myeloproliferative disease, wherein the hematopoietic stem cell or progenitor thereof is produced according to a method comprising:

(a) expressing in the hematopoietic stem cell or progenitor thereof a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; and

(b) contacting the hematopoietic stem cell or progenitor thereof with a guide RNA that targets a nucleic acid molecule encoding a CD117 protein, and introducing a T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G,

K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, or T488A mutation in the CD117 protein.

3. The method of claim 1 or the use of claim 2, wherein the hemoglobinopathy is selected from the group consisting of sickle cell anemia, sickle cell disease, thalassemia, Fanconi anemia, aplastic anemia, and Wiskott-Aldrich syndrome;

wherein the hematologic cancer is selected from the group consisting of acute myeloid leukemia, acute lymphoid leukemia, chronic myeloid leukemia, chronic lymphoid leukemia, multiple myeloma, diffuse large B-cell lymphoma, and non-Hodgkin's lymphoma; and

wherein the myeloproliferative disease is a myelodysplastic syndrome.

4. The method of claim 1, or the use of claim 2, wherein the cell is selected from a hematopoietic stem cell or a progenitor thereof, or a common myeloid progenitor, proerythroblast, or erythroblast; optionally, wherein the cell is a mammalian cell.

5. The method or use of any one of claims 1-4, wherein the method further comprises contacting the cell with one or more additional guide RNAs that target a protein selected from the group consisting of CXCR4, CD135, CD90, CD45, and CD34.

6. An *in vitro* or *ex vivo* method of base editing a gene encoding a CD117 protein expressed by a hematopoietic stem cell or a progenitor thereof, the method comprising:

(a) expressing in a hematopoietic stem cell or a progenitor thereof comprising a CD117 protein a nucleobase editor polypeptide, wherein the nucleobase editor polypeptide comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; and

(b) contacting the cell with a guide RNA capable of targeting a nucleic acid molecule encoding the CD117 protein, thereby base editing the gene encoding the CD117 protein and introducing a T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C,

T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, or T488A mutation in the CD117 protein.

7. The method of any one of claims 1-6, wherein the deaminase is an adenosine deaminase or a cytidine deaminase.

8. The method of claim 7, wherein the adenosine deaminase is a TadA*8.

9. The method of claim 8, wherein the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

10. The method or use of any one of claims 1-9, wherein the deaminase is a monomer or heterodimer;

wherein the nucleobase editor polypeptide is an internal base editor comprising the deaminase inserted at an internal location of the napDNAbp;

wherein the nucleobase editor polypeptide further comprises one or more uracil glycosylase inhibitors; and/or

wherein the nucleobase editor polypeptide further comprises one or more nuclear localization sequences.

11. Use of a hematopoietic stem cell in the preparation of a medicament for hematopoietic stem cell transplantation, wherein the medicament comprises i) an edited hematopoietic stem cell and ii) an antibody, antibody drug conjugate, or chimeric antigen receptor expressing T cell (CAR-T) that selectively binds a wild-type version of a CD117 protein, wherein the edited hematopoietic stem cell is prepared according to a method comprising:

(a) expressing in a hematopoietic stem cell of the subject a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase; and

(b) contacting the hematopoietic stem cell with a guide RNA capable of targeting a nucleic acid molecule encoding a CD117 protein, thereby introducing a T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G, N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, or T488A mutation in the CD117 protein and generating an edited hematopoietic stem cell.

12. The use of claim 11, wherein the deaminase is an adenosine deaminase or a cytidine deaminase.

13. The use of claim 12, wherein the adenosine deaminase is a TadA*8.

14. The use of claim 13, wherein the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

15. The use of any one of claims 11-14, wherein the deaminase is a monomer or heterodimer;

wherein the nucleobase editor polypeptide is an internal base editor comprising the deaminase inserted at an internal location of the napDNAbp;

wherein the nucleobase editor polypeptide further comprises one or more uracil glycosylase inhibitors;

wherein the nucleobase editor polypeptide further comprises one or more nuclear localization sequences; and/or

wherein the mutation fails to alter the biological activity of the CD117 protein.

16. The method or use of any one of claims 1-15, wherein the CD117 protein comprising the mutation is capable of binding Stem Cell Factor (SCF); or is capable of SCF signaling.

17. The method or use of any one of claims 1-16, wherein the mutation results from the modification of one or more single target nucleobases.
18. The method or use of claim 17, wherein the single target nucleobase is a cytosine (C) and wherein the modification comprises conversion of the C to a thymine (T); or
 wherein the single target nucleobase is an adenosine (A) and wherein the modification comprises conversion of the A to a guanine (G).
19. The method or use of any one of claims 1-18, wherein the cell is a CD34+ cell; and/or
 wherein the cell is from a subject having hemoglobinopathy, hematologic cancer, or a myeloproliferative disease.
20. The method or use of claim 19, wherein the cell is from a subject having sickle cell disease.
21. The method or use of any one of claims 1-20, wherein the deaminase is fused to the napDNAbp;
 wherein the deaminase is inserted at an internal location of the napDNAbp;
 wherein the napDNAbp is nuclease inactive or a nickase;
 wherein the napDNAbp comprises a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polypeptide or a portion thereof;
 wherein the napDNAbp comprises a dead Cas9 or a Cas9 nickase;
 wherein the napDNAbp is a modified *Staphylococcus aureus* Cas9, *Streptococcus thermophilus* 1 Cas9 (St1Cas9), or *Streptococcus pyogenes* Cas9 (SpCas9), or
 wherein the napDNAbp comprises an SpCas9 having an altered protospacer-adjacent motif (PAM) specificity; optionally,
 wherein the altered PAM has specificity for the nucleic acid sequence 5'-NGC-3';
22. The use of any one of claims 11-21, wherein the deaminase is a cytidine deaminase or an adenosine deaminase.
23. The use of claim 22, wherein the cytidine deaminase is an APOBEC deaminase; or

wherein the adenosine deaminase is a TadA deaminase,
a TadA*8, or

TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8,
TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15,
TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22,
TadA*8.23, or TadA*8.24.

24. The method or use of any one of claims 1-23, wherein the guide RNA comprises the nucleic acid sequence selected from Table 23.

25. The method or use of any one of claims 1-24, wherein the guide RNA comprises a nucleic acid sequence that hybridizes to the complement of a CD117 target sequence selected from the group consisting of:

CAAGCTATCTTCTTAGGGA;
GCAAGCTATCTTCTTAGGGAA;
ACTTACGACAGGCTCGTGAA;
TGACCAATTATTCCCTCAAG;
AGTGACCAATTATTCCCTCA;
ACTACAGTATTTGTAAACGA;
GTGACCAATTATTCCCTCAA;
TGTTTCATATAGATCCACTGC;
AAGTGACCAATTATTCCCTC;
TTCTGTTTCATATAGATCCAC;
ACAGGAGAAATATAATAGCT;
AAATATAATAGCTGGCATCA;
AAGACAACGACACGCTGGTC; and
GGCTGTTATGCACTGATCCG.

26. A base editor system comprising a fusion protein or a polynucleotide encoding the fusion protein, wherein the fusion protein comprises a nucleic acid programmable DNA binding protein (napDNAbp) and a deaminase, and a guide polynucleotide comprising a nucleic acid sequence selected from Table 23; or

a guide polynucleotide comprising a nucleic acid sequence that hybridizes to the

complement of a target nucleic acid sequence selected from the group consisting of:

CAAGCTATCTTCTTAGGGA;
 GCAAGCTATCTTCTTAGGGAA;
 ACTTACGACAGGCTCGTGAA;
 TGACCAATTATTCCCTCAAG;
 AGTGACCAATTATTCCCTCA;
 GTGACCAATTATTCCCTCAA;
 TGTTTCATATAGATCCACTGC;
 AAGTGACCAATTATTCCCTC;
 TTCTGTTTCATATAGATCCAC;
 ACAGGAGAAATATAATAGCT;
 AAATATAATAGCTGGCATCA;
 ACTACAGTATTTGTAAACGA;
 ACTACAGTATTTGTAAACGA;
 AAGACAACGACACGCTGGTC; and
 GGCTGTTATGCACTGATCCG.

27. The base editor system of claim 26, wherein the fusion protein further comprises one or more uracil glycosylase inhibitors; and/or one or more nuclear localization sequences.

28. The base editor system of any one of claims 26 or 27, wherein the base editor system is capable of modifying one or more single target nucleobases to effect at least one amino acid substitution in a CD117 polypeptide.

29. The base editor system of claim 28, wherein the single target nucleobase is a cytosine (C) and wherein the modification comprises conversion of the C to a thymine (T); or wherein the single target nucleobase is an adenosine (A) and wherein the modification comprises conversion of the A to a guanine (G).

30. The base editor system of claim 29, wherein the at least one amino acid substitution is in domain 1, 2, 3, 4, or 5 of CD117; or wherein the at least one amino acid substitution is selected from the group consisting of T13A, I39V, H40R, D45G, D52G, E53G, I54V, R55G, T59A, K65E, K65R, T67A, E76G,

N77G, N77S, N77Y, K78E, Q79R, N80G, E81G, E81D, N99G, K100G, H101R, L124P, Y125H, D129E, D129G, N130D, N130G, D131G, D131N, T132A, T144A, N145D, N145G, N145Y, N145S, Y146C, F162P, F162L, I163T, I163V, M171T, I172T, V195A, L196P, K199G, I201V, S220G, Y221C, Q256R, E257D, E257G, K258E, Y259C, T303A, T304A, M318V, T322A, V323A, F324L, K342G, K342R, Y350H, M351T, R353G, T354A, T354I, R381G, K383G, T385A, Y418C, D419G, R420G, I438M, D439G, Q460R, T461A, K484G, N486G, and T488A.

31. The base editor system of any one of claims 26-30, wherein the napDNAbp is nuclease inactive or a nickase;

wherein the napDNAbp comprises a Cas9, Cas12a/Cpf1, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polypeptide or a portion thereof;

wherein the napDNAbp comprises a Cas9 polypeptide or a portion thereof;

wherein the napDNAbp comprises a dead Cas9 or a Cas9 nickase;

wherein the napDNAbp is a modified *Staphylococcus aureus* Cas9, *Streptococcus thermophilus* 1 Cas9, or *Streptococcus pyogenes* Cas9 (SpCas9); or

wherein the napDNAbp comprises an SpCas9 having an altered protospacer-adjacent motif (PAM) specificity; optionally,

wherein the altered PAM has specificity for the nucleic acid sequence 5'-NGC-3'.

32. The base editor system of any one of claims 26-31, wherein the deaminase is capable of deaminating cytidine or adenine in DNA;

wherein the deaminase is a cytidine deaminase;

wherein the cytidine deaminase is an APOBEC deaminase;

wherein the deaminase is an adenosine deaminase;

wherein the adenosine deaminase is a TadA*8;

wherein the adenosine deaminase is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24; and/or

wherein the deaminase is a monomer or heterodimer.

33. A polynucleotide encoding the base editor system of any one of claims 26-32.

34. A cell produced by the method of claim 1 or 3.
35. A cell produced by introducing into a cell, or a progenitor thereof, the base editor system of any one of claims 26-32 or the polynucleotide of claim 33.
36. The cell of claim 34 or 35, wherein the cell is produced *ex vivo* or *in vitro*;
wherein the cell is a hematopoietic stem cell, a CD34+ cell, or a common myeloid progenitor, proerythroblast, or erythroblast; and/or
wherein the cell is a mammalian cell.
37. A pharmaceutical composition comprising an effective amount of the cell of any one of claims 34-36.
38. The pharmaceutical composition of claim 37 for use in treating a subject with a hemoglobinopathy, hematologic cancer, or myeloproliferative disease.
39. Use of the cell of any one of claims 34-36 in the preparation of a pharmaceutical composition for treating a subject with a hemoglobinopathy, hematologic cancer, or myeloproliferative disease.
40. The cell of any one of claims 34-36 for use in treating a subject with a hemoglobinopathy, hematologic cancer, or myeloproliferative disease, or the use of claim 39, wherein a conditioning regimen comprising an antibody, antibody drug conjugate or chimeric antigen receptor expressing T-cell that selectively binds a wild-type CD117 protein is to be administered to the subject.
41. The cell for use or the use of claim 40, wherein the antibody, antibody drug conjugate or chimeric antigen receptor expressing T-cell selectively binds domain 1, 2, 3, 4, or 5 of CD117; optionally,
wherein the antibody is a monoclonal antibody, which is to be administered sequentially or concurrently; and/or
wherein the hemoglobinopathy is sickle cell disease; and/or wherein the subject is a mammal.

42. A kit comprising the cell of any one of claims 34-36, the base editor system of any one of claims 26-32, the polynucleotide of claim 33, or the pharmaceutical composition of claim 37 or 38.