

COMPOSITIONS AND METHODS FOR GENOME EDITING THE NEONATAL FC RECEPTOR

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims priority to and the benefit of U.S. Provisional Application No. 63/255,290, filed October 13, 2021, the entire contents of which are incorporated herein by reference.

SEQUENCE LISTING

10 This application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. The Sequence Listing XML file, created on October 9, 2022, is named 180802_049101_PCT_SL.xml and is 970,484 bytes in size.

TECHNICAL FIELD

15 The present disclosure relates to the field of genome editing. Specifically, the disclosure relates to compositions and methods for editing, modifying expression, and/or silencing the neonatal Fc receptor (FcRn) gene, *FCGRT*.

BACKGROUND OF THE INVENTION

Immunoglobulin G (IgG) is the most common type of antibody found in blood circulation and extracellular fluids, where it controls infection of body tissues. While IgG
25 can directly bind antigen, the fragment crystallizable (Fc) region of IgG also binds receptors on cells to effect an immune response. The family of Fc gamma receptors (FcγR) includes the atypical neonatal Fc receptor (FcRn), encoded by the *FCGRT* gene. FcRn functions to recirculate and maintain IgG and albumin, as well as transport IgG and albumin across polarized cellular barriers, thereby increasing the half-life of IgG and albumin in circulation.
30 FcRn also interacts with and facilitates antigen presentation of peptides derived from IgG immune complexes (IC).

FcRn was first identified as the receptor that transports maternal IgG antibodies from mother to child. Initially, it was believed that FcRn was only present in placental and intestinal tissues during the fetal and newborn stages. However, FcRn is now known to be
35 expressed in many tissues throughout the body, including epithelia, endothelia, and cells of

hematopoietic origin. Specifically, FcRn expression in the epithelia has been detected in the intestines, placenta, kidney, and liver.

Several autoimmune disorders are caused by the reaction of IgG to autoantigens, including myasthenia gravis, warm autoimmune hemolytic anemia (wAIHA), idiopathic thrombocytopenia purpura (ITP), Grave's disease, chronic inflammatory demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and hemolytic diseases of fetus and newborn (HDFN). As FcRn functions to maintain IgG levels in circulation, FcRn also extends the half-life of antibodies that give rise to such autoimmune disorders. Intravenous immunoglobulin (IVIg) is a recently developed therapy that saturates FcRn's IgG recycling capacity and reduces the levels of pathogenic IgG binding to FcRn, thereby facilitating the reduction in levels of IgG autoantibodies. Other strategies for treating autoimmune disorders include injection of higher affinity antibodies to reduce the inflammatory response to autoantigen.

A need remains for improved compositions and methods for targeted treatment of FcRn-mediated autoimmune disorders.

SUMMARY OF THE INVENTION

Provided herein are compositions and methods for modifying the neonatal Fc receptor for IgG (FcRn) protein and/or expression or activity thereof in a mammalian cell. The compositions and methods disclosed herein yield production of modified, variant FcRn proteins having a reduced ability to bind to an Fc region of an IgG antibody. Such compositions and methods are useful in ameliorating IgG-mediated autoimmune disorders. Advantageously, the compositions and methods disclosed herein specifically target FcRn binding to IgG without interfering with albumin half-life in a subject.

Accordingly, in one embodiment, a method of modifying FcRn protein in a mammalian cell is provided, the method comprising contacting the cell with a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of an *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene.

In another embodiment, a method of treating an IgG-mediated autoimmune disorder in a subject in need thereof is provided, the method comprising modifying FcRn protein in a mammalian cell of the subject.

In another embodiment, a composition is provided, comprising a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of the *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene.

In another embodiment, lipid nanoparticles (LNP) that are surface-functionalized to incorporate an Fc fragment of an IgG antibody or other targeting moiety are provided. The disclosed LNPs can target the neonatal Fc receptor (FcRn) on epithelial surfaces, fuse or become internalized, and deliver their payload to the targeted cells. The LNPs disclosed herein may comprise siRNA for silencing FcRn, thereby limiting the half-life of IgG in circulation and treating an IgG-mediated autoimmune disorder in a subject in need thereof.

In another embodiment, a LNP is provided, comprising: a lipid monolayer membrane comprising at least one fragment crystallizable (Fc) region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane.

In another embodiment, a LNP is provided, comprising: a lipid monolayer membrane comprising at least one fragment Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one nucleic acid.

In another embodiment, a LNP is provided, comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one siRNA or guide RNA that modulates expression of or silences an *FCGRT* gene.

In another embodiment, a pharmaceutical composition is provided, comprising: at least one LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one nucleic acid; and at least one pharmaceutically-acceptable excipient.

In another embodiment, a method of treating an IgG-mediated autoimmune disorder in a subject in need thereof is provided, the method comprising administering to the subject a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or other targeting moiety as disclosed herein, or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein

the lipid core matrix comprises at least one siRNA or guide RNA that that modulates expression of or silences an *FCGRT* gene.

In another embodiment, a method of silencing FcRn expression in a cell is provided, the method comprising contacting the cell with a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one siRNA that silences an *FCGRT* gene. In one aspect, the disclosure features a method of altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide. The method involves contacting the FcRn polynucleotide with a base editor system containing one or more guide polynucleotides and a base editor, or one or more polynucleotides encoding the base editor system, thereby altering the nucleobase of the FcRn polynucleotide. The base editor contains a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain. In the base editor system, (a) the one or more guide polynucleotides contain a nucleic acid sequence containing at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in Table 2B; or (b) the one or more guide polynucleotides targets the base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from one or more of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130, W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:

FcRn amino acid sequence

AESHL SLLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEP CGAWVWENQVSWYWE
KETTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQ
GTWGGDWPEALAI SQRWQQQDKAANKELTFLLFSCPHRLREHLERGRGNLEWKEPPSMRLKA
25 RPSSPGFSVLTCSAFSFYPPQLRFLRNGLAAGTGQDGFGPSDGSFHASSSLTVKSGDEH
HYCCIVQHAGLAQPLRVELESPAKSSVLVVGIVIGVLLLLTAAVGGALLWRRMRSGLPAPWI
SLRGDDTGVL LPTPGEAQDADLKDVNVIPATA (SEQ ID NO: 530), or a corresponding
position in another FcRn polypeptide sequence.

In another aspect, the disclosure features a cell produced by the method of any of the aspects of the disclosure, or embodiments thereof.

In another aspect, the disclosure features a base editor system for altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide. The base editor system contains: (i) one or more guide polynucleotides, or one or more polynucleotides encoding the one or more guide polynucleotides, and (ii) a base editor containing a nucleic

acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain, or one or more polynucleotides encoding the base editor. In the base editor system, (a) the one or more guide polynucleotides contain a nucleic acid sequence containing at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in Table 2B; or (b) the one or more guide polynucleotides targets the base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from one or more of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130, W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:

FcRn amino acid sequence

10 AESHLSLLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWE
KETTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQ
GTWGGDWPEALAISQRWQQQDKAANKELTFLLFSCPHRLREHLERGRGNLEWKEPPSMRLKA
RPSSPGFSVLTCSAFSFYPPELQLRFLRNGLAAGTGQGDGFGPNSDGSFHASSSLTVKSGDEH
HYCCIVQHAGLAQPLRVELESPAKSSVLVVGIVIGVLLLLTAAAVGGALLWRRMRSGLPAPWI
15 SLRGDDTGVLPTPGEAQDADLKDVNVI PATA (SEQ ID NO: 530), or a corresponding
position in another FcRn polypeptide sequence.

In another aspect, the disclosure features a polynucleotide encoding the base editor system of any of the aspects of the disclosure, or embodiments thereof.

In another aspect, the disclosure features a vector containing the polynucleotide of any of the aspects of the disclosure, or embodiments thereof.

In another aspect, the disclosure features a cell containing the polynucleotide or vector of any of the aspects of the disclosure, or embodiments thereof.

In another aspect, the disclosure features a composition containing the base editor system, polynucleotide, vector, or cell of any of the aspects of the disclosure, or embodiments thereof.

In another aspect, the disclosure features a pharmaceutical composition containing the composition of any of the aspects of the disclosure, or embodiments thereof, and a pharmaceutically acceptable excipient.

In another aspect, the disclosure features a method of treating an autoimmune disorder mediated by immunoglobulin G in a subject in need thereof. The method involves altering a nucleobase of an FcRn polynucleotide in the subject by administering to the subject a base editor system, or one or more polynucleotides encoding the base editor system, thereby treating the autoimmune disorder. The base editor system contains one or more guide polynucleotides and a base editor. The base editor contains a nucleic acid programmable

DNA binding protein (napDNAbp) domain and a deaminase domain. In the base editor system, (a) the one or more guide polynucleotides contains a nucleic acid sequence containing at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in Table 2B; or (b) the one or more guide polynucleotides targets the base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from one or more of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130, W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:

FcRn amino acid sequence

10 AESHLSELLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWE
KETTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQ
GTWGGDWPEALAIISQRWQQQDKAANKELTFLLFSCPHRLREHLERGRGNLEWKEPPSMRLKA
RPSSPGFSVLTCFAFSFYPPQLRFLRNGLAAGTGQGDGFGPNSDGSFHASSSLTVKSGDEH
HYCCIVQHAGLAQPLRVELESPAKSSVLVVGIVIGVLLLLTAAAVGGALLWRRMRSGLPAPWI
15 SLRGDDTGVLPTPGEAQDADLKDVNVI PATA (SEQ ID NO: 436), or a corresponding
position in another FcRn polypeptide sequence.

In another aspect, the disclosure features a kit suitable for use in the method of any of the aspects of the disclosure, or embodiments thereof, and containing a guide polynucleotide containing a sequence listed in Table 2A or Table 2B.

20 In another aspect, the disclosure features a method of altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide. The method involves contacting the FcRn polynucleotide with a base editor system, thereby altering the nucleobase of the FcRn polynucleotide. The base editor system contains one or more guide polynucleotides selected from one or more of gRNA1583, gRNA1578, gRNA3265, or one or more polynucleotides encoding the same, and a base editor containing a nucleic acid
25 programmable DNA binding protein (napDNAbp) domain and an adenosine deaminase domain, or one or more polynucleotides encoding the base editor.

In another aspect, the disclosure features a base editor system containing one or more guide polynucleotides selected from one or more of gRNA1583, gRNA1578, gRNA3265, or one or more polynucleotides encoding the same, and a base editor containing a nucleic acid
30 programmable DNA binding protein (napDNAbp) domain and an adenosine deaminase domain, or one or more polynucleotides encoding the base editor.

In another aspect, the disclosure features a guide polynucleotide containing a sequence listed in Table 2A or Table 2B.

In any of the aspects of the disclosure, or embodiments thereof, the alteration of the nucleobase results in one or more of the following amino acid alterations in the FcRn polypeptide encoded by the FcRn polynucleotide relative to the reference sequence: F110L, F110S, F110P, L112P, N113S, N113D, E115G, E115K, E116G, E116K, E116Q, F117P, M118N, M118V, M118I, M118T, N119G, N119D, N119S, N119C, D121G, L122F, L122A, L122P, T126I, T126S, T126N, T126A, W127R, G128S, D130G, D130N, D130H, W131R, W131Q, P132L, P132S, P132P, E133G, A134V, L135P, I137V, I137T. In any of the aspects of the disclosure, or embodiments thereof, the one or more guide polynucleotides target the base editor to effect an alteration of a nucleobase in a codon encoding the amino acid M118 or W131 in the reference sequence. In any of the aspects of the disclosure, or embodiments thereof, the alteration of the nucleobase results in an amino acid alteration in the FcRn polypeptide encoded by the FcRn polynucleotide selected from one or more of M118V, M118I, M118T, W131R, and W131Q.

In any of the aspects of the disclosure, or embodiments thereof, the one or more amino acid alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to IgG1, IgG2, IgG3, and/or IgG4. In any of the aspects of the disclosure, or embodiments thereof, the one or more amino acid alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to an Fc region of IgG1, IgG2, IgG3, and/or IgG4. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations has a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is greater than 3000 nM.

In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide encoded by the FcRn polynucleotide containing an altered nucleobase is capable of binding albumin. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 2000 nM. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 1000 nM. In any of the aspects of the disclosure, or embodiments thereof, binding of the FcRn polypeptide containing the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 500 nM.

In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations may have a K_D in solution for binding with albumin that is no more than 1.5 times that of a reference FcRn polypeptide that has the same

amino acid sequence except that it does not contain the one or more amino acid alterations. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations may have a K_D in solution for binding with albumin that is between 0.5 and 1.5 times that of a reference FcRn polypeptide that has the same amino acid sequence except that it does not contain the one or more amino acid alterations. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations may have a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is at least 5 times that of a reference FcRn polypeptide that has the same amino acid sequence except that it does not contain the one or more amino acid alterations. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations may have a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is at least 10 times that of a reference FcRn polypeptide that has the same amino acid sequence except that it does not contain the one or more amino acid alterations. In any of the aspects of the disclosure, or embodiments thereof, the FcRn polypeptide containing the one or more amino acid alterations does not bind to IgG1, IgG2, IgG3, and/or IgG4 at detectable levels, e.g., as measured in a suitable assay such as an SPR assay described herein.

In any of the aspects of the disclosure, or embodiments thereof, the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 20%. In any of the aspects of the disclosure, or embodiments thereof, the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 40%. In any of the aspects of the disclosure, or embodiments thereof, the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 50%.

In any of the aspects of the disclosure, or embodiments thereof, the deaminase domain is capable of deaminating cytidine or adenine in DNA. In any of the aspects of the disclosure, or embodiments thereof, the deaminase domain is an adenosine deaminase domain or a cytidine deaminase domain. In any of the aspects of the disclosure, or embodiments thereof, the adenosine deaminase converts a target A•T to G•C in the FcRn polynucleotide. In any of the aspects of the disclosure, or embodiments thereof, the cytidine deaminase converts a target C•G to T•A in the FcRn polynucleotide. In any of the aspects of the disclosure, or embodiments thereof, the cytidine deaminase domain is an APOBEC deaminase domain or a derivative thereof.

In any of the aspects of the disclosure, or embodiments thereof, the base editor is a BE4 base editor.

In any of the aspects of the disclosure, or embodiments thereof, the adenosine deaminase domain is a TadA deaminase domain. In any of the aspects of the disclosure, or embodiments thereof, the deaminase domain is an adenosine deaminase domain. In any of the aspects of the disclosure, or embodiments thereof, the adenosine deaminase is a TadA*8 or Tad*9 variant. In any of the aspects of the disclosure, or embodiments thereof, the adenosine deaminase is a 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 In any of the aspects of the disclosure, or embodiments thereof, the deaminase domain is a monomer or heterodimer.

In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain is Cas9 or Cas12. In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain is a nuclease inactive or nickase variant. In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain contains a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polynucleotide or a functional portion thereof. In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain contains a dead Cas9 (dCas9) or a Cas9 nickase (nCas9). In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain contains a variant of SpCas9 or SaCas9 having an altered protospacer-adjacent motif (PAM) specificity. In any of the aspects of the disclosure, or embodiments thereof, the SpCas9 or SaCas9 has specificity for a PAM sequence selected from one or more of NGG, NGA, NGC, NNGRRT, and NNNRRT, where N is any nucleotide and R is A or G.

In any of the aspects of the disclosure, or embodiments thereof, the napDNAbp domain contains a nuclease active Cas9.

30 In any of the aspects of the disclosure, or embodiments thereof, the base editor further contains one or more uracil glycosylase inhibitors (UGIs), or the method further involves expressing a UGI in a cell in trans with the base editor.

In any of the aspects of the disclosure, or embodiments thereof, the base editor further contains one or more nuclear localization signals (NLS). In any of the aspects of the disclosure, or embodiments thereof, the NLS is a bipartite NLS.

In any of the aspects of the disclosure, or embodiments thereof, the one or more guide polynucleotides contain a scaffold containing one of the following nucleotide sequences:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU (SpCas9 scaffold; SEQ ID NO: 317) or

- 5 GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAUGCCGUGUUUA
UCUCGUCAACUUGUUGGCGAGAUUUU (SaCas9 scaffold; SEQ ID NO: 436). In any of the
aspects of the disclosure, or embodiments thereof, the one or more guide polynucleotides
contain one or more modified nucleotides. In any of the aspects of the disclosure, or
embodiments thereof, the one or more modified polynucleotides are at the 5' terminus and/or
10 the 3' terminus of the one or more guide polynucleotides. In any of the aspects of the
disclosure, or embodiments thereof, the one or more modified nucleotides are 2'-O-methyl-3'-
phosphorothioate nucleotides. In any of the aspects of the disclosure, or embodiments
thereof, the one or more guide polynucleotides contain a spacer containing only 19 to 23
nucleotides. In any of the aspects of the disclosure, or embodiments thereof, the one or more
15 guide polynucleotides contain a spacer containing only 19 or 20 nucleotides.

In any of the aspects of the disclosure, or embodiments thereof, the base editor
contains a complex containing the deaminase domain, the napDNAbp domain, and the guide
polynucleotide, or the base editor is a fusion protein containing the napDNAbp domain fused
to the deaminase domain.

- 20 In any of the aspects of the disclosure, or embodiments thereof, the FcRn
polynucleotide is in a cell. In any of the aspects of the disclosure, or embodiments thereof,
the cell is a hepatocyte, an endothelial cell, a myeloid cell, or an epithelial cell. In any of the
aspects of the disclosure, or embodiments thereof, the cell is *in vivo* or *ex vivo*. In any of the
aspects of the disclosure, or embodiments thereof, the cell is in a subject. In any of the
25 aspects of the disclosure, or embodiments thereof, the cell is a mammalian cell. In any of the
aspects of the disclosure, or embodiments thereof, the cell is a human cell.

In any of the aspects of the disclosure, or embodiments thereof, the subject is a
mammal. In any of the aspects of the disclosure, or embodiments thereof, the mammal is a
human.

- 30 In any of the aspects of the disclosure, or embodiments thereof, the base editor further
contains one or more uracil glycosylase inhibitors (UGIs), or the base editor system further
contains a UGI in trans with the base editor.

In any of the aspects of the disclosure, or embodiments thereof, the vector contains a lipid nanoparticle. In any of the aspects of the disclosure, or embodiments thereof, the lipid nanoparticle contains a lipid monolayer containing a lipid selected from one or more of lecithin, phosphatidylcholines, phosphatidic acid, phosphatidylethanolamines, phosphatidylglycerols, phosphatidylserines, phosphatidylinositols, cardiolipins, lipid-polyethyleneglycol conjugates, and combinations thereof. In any of the aspects of the disclosure, or embodiments thereof, the lipid monolayer contains a PEGylated lipid. In any of the aspects of the disclosure, or embodiments thereof, the lipid monolayer further contains a cholesterol. In any of the aspects of the disclosure, or embodiments thereof, the lipid nanoparticle contains an ionizable cationic lipid selected from one or more of: N-methyl-N-(2-(arginoylamino) ethyl)- N, N- Di octadecyl aminium chloride or di stearyl arginyl ammonium chloride] (DSAA); N,N-di-myristoyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride (DMGLA); N,N-dimyristoyl-N-methyl-N-2[N2-guanidino-L- lysiny] aminoethyl ammonium chloride; N,N-dimyristoyl-N-methyl-N-2[N'-(N2, N6- di-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-di-stearyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-dioleoyl-N,N-dimethylammonium chloride (DODAC); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTAP); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTMA); N,N-distearyl- N,N-dimethylammonium bromide (DDAB); 3-(N-(N',N'-dimethylaminoethane)- carbamoyl) cholesterol (DC-Choi); N-(1,2-dimyristyloxyprop-3-yl)-N,N- dimethyl-N-hydroxyethyl ammonium bromide (DMRIE); 1,3-dioleoyl-3- trimethylammonium-propane, N-(1-(2,3-dioleoyloxy)propyl)-N-(2-(sperminecarboxamido)ethyl)-N,N-dimethy- 1 ammonium trifluoro-acetate (DOSPA); GAP-DLRIE; DMDHP; 3-p[4N-(H8N-diguanidino spermidine)-carbamoyl] cholesterol (BGSC); 3-P[N,N-diguanidinoethyl-aminoethane)-carbamoyl] cholesterol (BGTC); N,N,N2,N3 Tetramethyltetrapalmitylspermine (cellfectin); N-t-butyl-N'- tetradecyl-3-tetradecyl-aminopropion-amidine (CLONfectin); dimethyldioctadecyl ammonium bromide (DDAB); 1,3-dioleoyloxy-2-(6-carboxyspermyl)-propyl amide (DOSPER); 4-(2,3-bis-palmitoyloxy-propyl)- 1-methyl-1H-imidazole (DPIM) N,N,N',N'-tetramethyl-N,N'-bis(2-hydroxyethyl)-2,3 dioleoyloxy- 1 ,4-butanediammonium iodide (Tfx-50); 1,2 dioleoyl-3-(4'-trimethylammonio) butanol-sn-glycerol (DOBT); cholesteryl (4'-trimethylammonia) butanoate (ChOTB) where the trimethylammonium group is connected via a butanol spacer arm to either the double chain (for DOTB) or cholesteryl group (for ChOTB); DL-1,2-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORI); DL-1,2-0-dioleoyl-3- dimethylaminopropyl-P-

hydroxyethylammonium (DORIE); 1,2-dioleoyl-3-succinyl-sn-glycerol choline ester (DOSC);
 cholesteryl hemisuccinate ester (ChOSC); dioctadecylamidoglycylspermine (DOGS);
 dipalmitoyl phosphatidylethanolamylspermine (DPPES); cholesteryl-3P- carboxyl-amido-
 ethylenetrimethylammonium iodide; 1-dimethylamino-3- trimethylammonio-DL-2-propyl-
 5 cholesteryl carboxylate iodide; cholesteryl-3-β- carboxyamidoethyleneamine; cholesteryl-3-
 P-oxy succinamido- ethylenetrimethylammonium iodide; 1-dimethylamino-3-
 trimethylammonio-DL-2- propyl-cholesteryl-3-P-oxy succinate iodide; 2-(2-
 trimethylammonio)- ethylmethylamino ethyl-cholesteryl-3-P-oxy succinate iodide; 3-β-N-
 (polyethyleneimine)-carbamoylcholesterol, DC-cholesterol; N4-cholesteryl-spermine HCl
 10 salt (GL67); N1-[2-((1 S)-1-[(3- aminopropyl)amino]-4-[di(3-amino-
 propyl)amino]butylcarboxamido)ethyl]-3,4-di[oleoyloxy]- benzamide (MVL5); and
 combinations thereof.

In any of the aspects of the disclosure, or embodiments thereof, the vector contains a
 polymer nanoparticle. In any of the aspects of the disclosure, or embodiments thereof, the
 15 vector is a viral vector. In any of the aspects of the disclosure, or embodiments thereof, the
 viral vector is a retroviral vector or an adeno-associated virus vector.

In any of the aspects of the disclosure, or embodiments thereof, the disorder is
 selected from one or more of myasthenia gravis (gMG), warm autoimmune hemolytic anemia
 (wAIHA), idiopathic thrombocytopenia purpura (ITP), Grave's disease, chronic
 20 inflammatory demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and hemolytic
 diseases of fetus and newborn (HDFN).

In any of the aspects of the disclosure, or embodiments thereof, the base editor further
 contains one or more uracil glycosylase inhibitors (UGIs), or the method further involves
 expressing a UGI in a cell in trans with the base editor.

25 In any of the aspects of the disclosure, or embodiments thereof, the administration is
 local administration. In any of the aspects of the disclosure, or embodiments thereof, the
 administration is systemic administration.

In any of the aspects of the disclosure, or embodiments thereof, the base editor system
 is administered to the subject using a vector.

30 In any of the aspects of the disclosure, or embodiments thereof, the vector is a lipid
 nanoparticle. In any of the aspects of the disclosure, or embodiments thereof, the vector
 targets the liver.

In any of the aspects of the disclosure, or embodiments thereof, the subject is a mammal. In any of the aspects of the disclosure, or embodiments thereof, the mammal is a human.

5 These and other objects, features, embodiments, and advantages will become apparent to those of ordinary skill in the art from a reading of the following detailed description and the appended claims.

Definitions

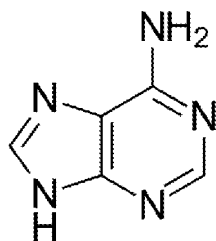
10 While the following terms are believed to be well understood in the art, definitions are set forth to facilitate explanation of the presently-disclosed subject matter. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the presently-disclosed subject matter belongs.

15 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.

20 Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about.” Accordingly, unless indicated to the contrary, the numerical parameters set forth in this specification and claims are approximations that can vary depending upon the desired properties sought to be obtained by the presently-disclosed subject matter.

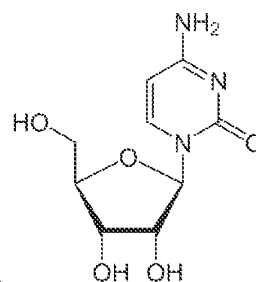
25 It should be understood that every maximum numerical limitation given throughout this specification includes every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation, as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. By

“adenine” or “9*H*-Purin-6-amine” is meant a purine nucleobase with the molecular formula



$C_5H_5N_5$, having the structure , and corresponding to CAS No. 73-24-5.

By “adenosine” or “4-Amino-1-[(2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2(1*H*)-one” is meant an adenine molecule attached to



5 a ribose sugar via a glycosidic bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_{10}H_{13}N_5O_4$.

By “adenosine deaminase” or “adenine deaminase” is meant a polypeptide or fragment thereof capable of catalyzing the hydrolytic deamination of adenine or adenosine. In some embodiments, the deaminase or deaminase domain is an adenosine deaminase catalyzing the hydrolytic deamination of adenosine to inosine or deoxy adenosine to deoxyinosine. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in deoxyribonucleic acid (DNA). The adenosine deaminases (*e.g.* engineered adenosine deaminases, evolved adenosine deaminases) provided herein may be from any organism (*e.g.*, eukaryotic, prokaryotic), including but not limited to 15 algae, bacteria, fungi, plants, invertebrates (*e.g.*, insects), and vertebrates (*e.g.*, amphibians, mammals). In some embodiments, the adenosine deaminase is an adenosine deaminase variant with one or more alterations and is capable of deaminating both adenine and cytosine in a target polynucleotide (*e.g.*, DNA, RNA) and may be referred to as a “dual deaminase”. Non-limiting examples of dual deaminases include those described in 20 PCT/US22/22050. In some embodiments, the target polynucleotide is single or double stranded. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in single-stranded DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and

cytosine in RNA. In embodiments, the adenosine deaminase variant is selected from those described in PCT/US2020/018192, PCT/US2020/049975, and PCT/US2017/045381.

By “adenosine deaminase activity” is meant catalyzing the deamination of adenine or adenosine to guanine in a polynucleotide. In some embodiments, an adenosine deaminase variant as provided herein maintains adenosine deaminase activity (*e.g.*, at least about 30%, 40%, 50%, 60%, 70%, 80%, 90% or more of the activity of a reference adenosine deaminase (*e.g.*, TadA*8.20 or TadA*8.19)).

By “Adenosine Base Editor (ABE)” is meant a base editor comprising an adenosine deaminase.

By “Adenosine Base Editor (ABE) polynucleotide” is meant a polynucleotide encoding an ABE.

By “Adenosine Base Editor 8 (ABE8) polypeptide” or “ABE8” is meant a base editor as defined herein comprising an adenosine deaminase or adenosine deaminase variant comprising one or more of the alterations listed in **Table 15**, one of the combinations of alterations listed in **Table 15**, or an alteration at one or more of the amino acid positions listed in **Table 15**, such alterations are relative to the following reference sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or a

corresponding position in another adenosine deaminase. In embodiments, ABE8 comprises alterations at amino acids 82 and/or 166 of SEQ ID NO: 1. In some embodiments, ABE8 comprises further alterations, as described herein, relative to the reference sequence.

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

“Administering” is referred to herein as providing one or more compositions described herein to a patient or a subject. By way of example and without limitation, composition administration (*e.g.*, injection) can be performed by intravenous (*i.v.*) injection, sub-cutaneous (*s.c.*) injection, intradermal (*i.d.*) injection, intraperitoneal (*i.p.*) injection, or intramuscular (*i.m.*) injection. One or more such routes can be employed. Parenteral administration can be, for example, by bolus injection or by gradual perfusion over time. In some embodiments, parenteral administration includes infusing or injecting intravascularly, intravenously, intramuscularly, intraarterially, intrathecally, intratumorally, intradermally, intraperitoneally, transtracheally, subcutaneously, subcuticularly, intraarticularly, subcapsularly, subarachnoidly and intrasternally. Alternatively, or concurrently,

administration can be by the oral route. In embodiments, one or more compositions described herein are administered by subretinal or subfoveal injection. In some instances, subretinal injection creates a bleb in the fovea.

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

By “albumin polypeptide” is meant a protein with at least about 85% amino acid sequence identity to GenBank Accession No. CAA23754.1, provided below, or a fragment thereof capable of binding to an FcRn polypeptide.

>CAA23754.1 serum albumin [Homo sapiens]

10 MKWVTFISLLFLFSSAYSRGVFRDDAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFED
HVKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNE
CFLQHKDDNPNLPRLVREVDVMCTAFHDNEETFLKKYLYEIAARRHPYFYAPELLFFAKRYK
AAFTECCQAADKAACLLPKLDEL RDEGKASSAKQRLKCSLQKFGERAFKAWAVARLSQRFP
KAEFAEVSKLVTDLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLEK
15 SHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLRLR
LAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCLEFKQLGEYKFQNALLV
YTKKVPQVSTPTLVEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVVLNQLCVLHEKTPVSDR
VTKCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKQTALVELVK
HKPKATKEQLKAVMDDFAAFVEKCKADDKETCFAEEGKKLVAASQAALGL (SEQ ID NO:
20 425).

By “albumin polynucleotide” is meant a nucleic acid molecule encoding an albumin polypeptide, as well as the introns, exons, 3′ untranslated regions, 5′ untranslated regions, and regulatory sequences associated with its expression, or fragments thereof. In embodiments, an albumin polynucleotide is the genomic sequence, cDNA, mRNA, or gene associated with
25 and/or required for albumin expression. An exemplary albumin nucleotide sequence from *Homo sapiens* is provided below (GenBank: V00495.1:76-1905):

ATGAAGTGGGTAACCTTTATTTCCCTTCTTTTTCTCTTTAGCTCGGCTTATTCCAGGGGTGT
GTTTCGTCGAGATGCACACAAGAGTGAGGTTGCTCATCGGTTTAAAGATTTGGGAGAAGAAA
ATTTCAAAGCCTTGGTGTTGATTGCCTTTGCTCAGTATCTTCAGCAGTGTCCATTTGAAGAT
30 CATGTAAAATTAGTGAATGAAGTAACTGAATTTGCAAAAACATGTGTAGCTGATGAGTCAGC
TGAAAATTGTGACAAATCACTTCATACCCTTTTTGGAGACAAATTATGCACAGTTGCAACTC
TTCGTGAAACCTATGGTGAAATGGCTGACTGCTGTGCAAAACAAGAACCTGAGAGAAATGAA
TGCTTCTTGCAACACAAAGATGACAACCCAAACCTCCCCGATTGGTGAGACCAGAGGTTGA
TGTGATGTGCACTGCTTTTCATGACAATGAAGAGACATTTTTGAAAAAATACTTATATGAAA

TTGCCAGAAGACATCCTTACTTTTATGCCCCGGAACCTCTTTTCTTTGCTAAAAGGTATAAA
GCTGCTTTTACAGAATGTTGCCAAGCTGCTGATAAAGCTGCCTGCCTGTTGCCAAAGCTCGA
TGAACCTTCGGGATGAAGGGAAGGCTTCGTCTGCCAAACAGAGACTCAAATGTGCCAGTCTCC
AAAAATTTGGAGAAAGAGCTTTCAAAGCATGGGCAGTGGCTCGCCTGAGCCAGAGATTTCCC
5 AAAGCTGAGTTTGCAGAAGTTTCCAAGTTAGTGACAGATCTTACCAAAGTCCACACGGAATG
CTGCCATGGAGATCTGCTTGAATGTGCTGATGACAGGGCGGACCTTGCCAAGTATATCTGTG
AAAATCAGGATTCGATCTCCAGTAAACTGAAGGAATGCTGTGAAAAACCTCTGTTGGAAAAA
TCCCCTGCATTGCCGAAGTGGAAAAATGATGAGATGCCTGCTGACTTGCCTTCATTAGCTGC
TGATTTTGTGAAAGTAAGGATGTTTGCAAAAACCTATGCTGAGGCAAAGGATGTCTTCCTGG
10 GCATGTTTTTGTATGAATATGCAAGAAGGCATCCTGATTACTCTGTCGTGCTGCTGCTGAGA
CTTGCCAAGACATATGAAACCACTCTAGAGAAGTGCTGTGCCGCTGCAGATCCTCATGAATG
CTATGCCAAAGTGTTTCGATGAATTTAAACCTCTTGTGGAAGAGCCTCAGAATTTAATCAAAC
AAAACCTGTGAGCTTTTTTAAGCAGCTTGAGAGTACAAATTCAGAATGCGCTATTAGTTCGT
TACACCAAGAAAGTACCCCAAGTGTCAACTCCAACCTTGTAGAGGTCTCAAGAAACCTAGG
15 AAAAGTGGGCAGCAAATGTTGTAAACATCCTGAAGCAAAAAGAATGCCCTGTGCAGAAGACT
ATCTATCCGTGGTCCCTGAACCAGTTATGTGTGTTGCATGAGAAAACGCCAGTAAGTGACAGA
GTCACAAAATGCTGCACAGAGTCCTTGGTGAACAGGCGACCATGCTTTTCAGCTCTGGAAGT
CGATGAAACATACGTTCCCAAAGAGTTTAATGCTGAAACATTACCTTCCATGCAGATATAT
GCACACTTTCTGAGAAGGAGAGACAAATCAAGAAACAACTGCACTTGTTGAGCTTGTGAAA
20 CACAAGCCCCAAGGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGATTTTCGAGCTTTTGT
AGAGAAGTGCTGCAAGGCTGACGATAAGGAGACCTGCTTTGCCGAGGAGGGTAAAAAACTTG
TTGCTGCAAGTCAAGCTGCCTTAGGCTTATAA (SEQ ID NO: 426).

By “alteration” or “modification” is meant a change in the expression level, structure,
or activity of an analyte, gene or polypeptide as detected by standard art known methods such
25 as those described herein. As used herein, an alteration includes a change (e.g., increase or
decrease) in expression levels. In embodiments, the increase or decrease in expression levels
is by 10%, 25%, 40%, 50% or greater. In some embodiments, an alteration includes an
insertion, deletion, or substitution of a nucleobase or amino acid (by, e.g., genetic engineering).

By “ameliorate” is meant decrease, suppress, attenuate, diminish, arrest, or stabilize
30 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
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 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, rIgG, 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.

Antibodies (immunoglobulins) comprise two heavy chains linked together by disulfide bonds, and two light chains, with each light chain being linked to a respective heavy chain by disulfide bonds in a "Y" shaped configuration. Each heavy chain has at one end a variable domain (VH) followed by a number of constant domains (CH). Each light chain has a variable domain (VL) at one end and a constant domain (CL) at its other end. The variable domain of the light chain (VL) is aligned with the variable domain of the heavy chain (VL), and the light chain constant domain (CL) is aligned with the first constant domain of the heavy chain (CH1). The variable domains of each pair of light and heavy chains form the antigen binding site. The isotype of the heavy chain (gamma, alpha, delta, epsilon or mu) determines the immunoglobulin class (IgG, IgA, IgD, IgE or IgM, respectively). The light chain is either of two isotypes (kappa (κ) or lambda (λ)) found in all antibody classes. The terms "antibody" or "antibodies" include intact antibodies, such as polyclonal antibodies or monoclonal antibodies (mAbs), as well as proteolytic portions or fragments thereof, such as the Fab or F(ab')₂ fragments, that are capable of specifically binding to a target protein. Antibodies may include chimeric antibodies; recombinant and engineered antibodies, and antigen binding fragments thereof. Exemplary functional antibody fragments comprising whole or essentially whole variable regions of both the light and heavy chains are defined as follows: (i) Fv, defined as a genetically engineered fragment consisting of the variable region of the light chain and the variable region of the heavy chain expressed as two chains; (ii) single-chain Fv ("scFv"), a genetically engineered single-chain molecule including the

variable region of the light chain and the variable region of the heavy chain, linked by a suitable polypeptide linker; (iii) Fab, a fragment of an antibody molecule containing a monovalent antigen-binding portion of an antibody molecule, obtained by treating an intact antibody with the enzyme papain to yield the intact light chain and the Fd fragment of the heavy chain, which consists of the variable and CH1 domains thereof; (iv) Fab', a fragment of an antibody molecule containing a monovalent antigen-binding portion of an antibody molecule, obtained by treating an intact antibody with the enzyme pepsin, followed by reduction (two Fab' fragments are generated per antibody molecule); and (v) F(ab')₂, a fragment of an antibody molecule containing a monovalent antigen-binding portion of an antibody molecule, obtained by treating an intact antibody with the enzyme pepsin (i.e., a dimer of Fab' fragments held together by two disulfide bonds).

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 (*e.g.*, Cas9 or Cpf1) in conjunction with a guide polynucleotide (*e.g.*, guide RNA (gRNA)). Representative nucleic acid and protein sequences of base editors include those sequences with about or at least about 85% sequence identity to any base editor sequence provided in the sequence listing, such as those corresponding to SEQ ID NOs: 2-11.

By "BE4 cytidine deaminase (BE4) polypeptide," is meant a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain, a cytidine deaminase domain, and two uracil glycosylase inhibitor domains (UGIs). In embodiments, the napDNAbp is a Cas9n(D10A) polypeptide. Non-limiting examples of cytidine deaminase domains include rAPOBEC, ppAPOBEC, RrA3F, AmAPOBEC1, and SsAPOBEC3B.

By "BE4 cytidine deaminase (BE4) polynucleotide," is meant a polynucleotide encoding a BE4 polypeptide.

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.

The term "base editor system" refers to an intermolecular complex 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 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 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 adenosine base editor (ABE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE) or a cytidine or cytosine base editor (CBE). In some embodiments, the base editor system (*e.g.*, a base editor system comprising a cytidine deaminase) comprises a uracil glycosylase inhibitor or other agent or peptide (*e.g.*, a uracil stabilizing protein such as provided in WO2022015969, the disclosure of which is incorporated herein by reference in its entirety for all purposes) that inhibits the inosine base excision repair system.

The term “Cas9” or “Cas9 domain” refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease.

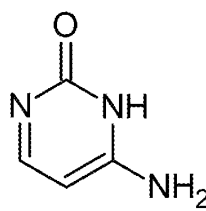
The term “coding sequence” or “protein coding sequence” as used interchangeably herein refers to a segment of a polynucleotide that codes for a protein. Coding sequences can also be referred to as open reading frames. The region or sequence is bounded nearer the 5' end by a start codon and nearer the 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
	TGG → TAG
	TGG → TAA

30

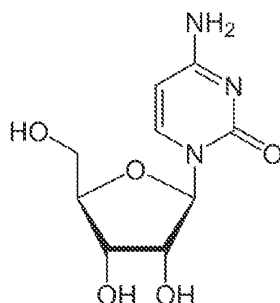
By “complex” is meant a combination of two or more molecules whose interaction relies on inter-molecular forces. Non-limiting examples of inter-molecular forces include covalent and non-covalent interactions. Non-limiting examples of non-covalent interactions include hydrogen bonding, ionic bonding, halogen bonding, hydrophobic bonding, van der
5 Waals interactions (e.g., dipole-dipole interactions, dipole-induced dipole interactions, and London dispersion forces), and π -effects. In an embodiment, a complex comprises polypeptides, polynucleotides, or a combination of one or more polypeptides and one or more polynucleotides. In one embodiment, a complex comprises one or more polypeptides that associate to form a base editor (e.g., base editor comprising a nucleic acid programmable
10 DNA binding protein, such as Cas9, and a deaminase) and a polynucleotide (e.g., a guide RNA). In an embodiment, the complex is held together by hydrogen bonds. It should be appreciated that one or more components of a base editor (e.g., a deaminase, or a nucleic acid programmable DNA binding protein) may associate covalently or non-covalently. As one example, a base editor may include a deaminase covalently linked to a nucleic acid
15 programmable DNA binding protein (e.g., by a peptide bond). Alternatively, a base editor may include a deaminase and a nucleic acid programmable DNA binding protein that associate noncovalently (e.g., where one or more components of the base editor are supplied in trans and associate directly or via another molecule such as a protein or nucleic acid). In an embodiment, one or more components of the complex are held together by hydrogen
20 bonds.

By “cytosine” or “4-Aminopyrimidin-2(1*H*)-one” is meant a purine nucleobase with



the molecular formula $C_4H_5N_3O$, having the structure
to CAS No. 71-30-7.

By “cytidine” is meant a cytosine molecule attached to a ribose sugar via a glycosidic



bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_9H_{13}N_3O_5$.

By “Cytidine Base Editor (CBE)” is meant a base editor comprising a cytidine
5 deaminase.

By “Cytidine Base Editor (CBE) polynucleotide” is meant a polynucleotide encoding a CBE.

By “cytidine deaminase” or “cytosine deaminase” is meant a polypeptide or fragment thereof capable of deaminating cytidine or cytosine. In embodiments, the cytidine or
10 cytosine is present in a polynucleotide. In one embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. The terms “cytidine deaminase” and “cytosine deaminase” are used interchangeably throughout the application. *Petromyza marinus* cytosine deaminase 1 (PmCDA1) (SEQ ID NO: 13-14), Activation-induced cytidine deaminase (AICDA) (SEQ ID NOs: 15-21), and APOBEC (e.g., SEQ ID NOs: 12-61) are
15 exemplary cytidine deaminases. Further exemplary cytidine deaminase (CDA) sequences are provided in the Sequence Listing as SEQ ID NOs: 62-66 and SEQ ID NOs: 67-189. Non-limiting examples of cytidine deaminases include those described in PCT/US20/16288, PCT/US2018/021878, 180802-021804/PCT, PCT/US2018/048969, and PCT/US2016/058344. By “cytosine deaminase activity” is meant catalyzing the deamination
20 of cytosine or cytidine. In one embodiment, a polypeptide having cytosine deaminase activity converts an amino group to a carbonyl group. In an embodiment, a cytosine deaminase converts cytosine to uracil (*i.e.*, C to U) or 5-methylcytosine to thymine (*i.e.*, 5mC to T). In some embodiments, a cytosine deaminase as provided herein has increased cytosine deaminase activity (*e.g.*, at least 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold,
25 80-fold, 90-fold, 100-fold or more) relative to a reference cytosine deaminase.

The term “deaminase” or “deaminase domain,” as used herein, refers to a protein or fragment thereof that catalyzes a deamination reaction.

“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 “disease” is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. Exemplary diseases include autoimmune disorders, such as autoimmune disorders mediated by IgG. Non-limiting examples of autoimmune disorders include myasthenia gravis (gMG), warm autoimmune hemolytic anemia (wAIHA), idiopathic thrombocytopenia purpura (ITP), Grave’s disease, chronic inflammatory demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and hemolytic diseases of fetus and newborn (HDFN).

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 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.*, 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 gene of interest in all cells of a subject, tissue or organ, but only to alter the gene of interest in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an effective amount is sufficient to ameliorate one or more symptoms of a disease.

By “neonatal Fc receptor for IgG (FcRn) polypeptide” or “Fc fragment of IgG receptor and transporter (FCGRT) polypeptide” is meant a protein having at least about 85% amino acid sequence identity to NCBI reference sequence NP_001129491 or a fragment thereof capable of binding albumin. An exemplary FcRn polypeptide sequence is provided below. Throughout the present disclosure, references are made to amino acid positions within the FcRn polypeptide sequence (*e.g.*, E115(138) or E115). Unless indicated otherwise, such references are made with reference to the below sequence, and the position number outside of parenthesis corresponds to the position in the below FcRn sequence without the

first 23 amino acids, which correspond to a signal peptide, included in the numbering, and the position inside the parenthesis corresponds to the position in the below FcRn sequence with the first 23 amino acids included in the numbering. For example, position E115(138) is in **bold-underlined text** in the below amino acid sequence.

5 1 mgvprpqpwa lglllflfpg slgaeshls lylhtavssp apgtpafwvs gwlgpqqyls
 61 ynslrgeaep cgawvwenqv swywekettd lrikeklfle afkalggkqp ytlqgllgce
 121 lgpdntsvpt akfalngeef mnfdlkqgtw ggdwpealai sqrwqqgdka ankeltflff
 181 scphrlrehl ergrgnlew eppsmrlkar psspgfsvlt csafsfyppe lqlrflrnl
 241 aagtgggdfg pnsdgsfhas ssltvksgde hhyccivqha glaqlrvel espakssvlv
 10 301 vgivigvlll taaavggall wrrmrslpa pwislrgddt gvllptpgea qdadlkdvnn
 361 ipata (SEQ ID NO: 427).

By “Fc fragment of IgG receptor and transporter (FcRn; FCGRT) polynucleotide” or “Fc fragment of IgG receptor and transporter (FCGRT) polynucleotide” is meant a nucleic acid molecule encoding an FcRn polypeptide, as well as the introns, exons, 3' untranslated
 15 regions, 5' untranslated regions, and regulatory sequences associated with its expression, or fragments thereof. In embodiments, an FcRn polynucleotide is the genomic sequence, cDNA, mRNA, or gene associated with and/or required for FcRn expression. An exemplary FcRn nucleotide sequence from *Homo sapiens* is provided below. A further exemplary FcRn nucleotide sequence from *Homo sapiens* is provided at Ensembl Accession No.
 20 ENSG00000211893.

1 aggatgtgag agaggaactg gggctctccag tcacgggagc caggagccgg ccagggccgc
 61 aggcaggaag ggagcgaggc tgaagggaac gtctgtctct cagcatgggg gtcccgggc
 121 ctacagccctg ggcgctgggg ctctgtctct ttctccttcc tgggagcctg ggcgcagaaa
 181 gccacctctc cctcctgtac caccttaccg cgggtgtctc gctgccccg gggactcctg
 25 241 ccttctgggt gtccggctgg ctgggcccgc agcagtacct gagctacaat agcctgcccg
 301 ggcagggcga gccctgtgga gcttgggtct gggaaaacca ggtgtcctgg tattgggaga
 361 aagagaccac agatctgagg atcaaggaga agctctttct ggaagcttcc aaagctttgg
 421 ggggaaaagg tccctacact ctgcagggcc tgctgggctg tgaactgggc cctgacaaca
 481 cctcggtgcc caccgccaag ttgcgcctga acggcgagga gttcatgaat ttgcacctca
 30 541 agcagggcac ctggggtggg gactggcccg aggcctggc tatcagtcag cgggtggcagc
 601 agcaggacaa ggcggccaac aaggagctca ccttctgtct attctcctgc ccgcaccgcc
 661 tgccgggagca cctggagagg ggccgcgga acctggagtg gaaggagccc cctccatgc
 721 gcctgaaggc ccgacccagc agccctggct ttccgtgct tacctgcage gccttctct
 781 tctaccctcc ggagctgcaa ctgcggttcc tgcggaatgg gctggccgct ggcaccggcc
 35 841 agggtgactt cgcccccaac agtgacggat ccttccacgc ctctctgtca ctaacagtca
 901 aaagtggcga tgagcaccac tactgtgtgca ttgtgcagca cgcggggctg gcgcagcccc
 961 tcagggtgga gctggaatct ccagccaagt cctccgtgct cgtggtggga atcgtcatcg
 1021 gtgtcttgc actcacggca gcggctgtag gaggagctct gttgtggaga aggatgagga

1081 gtggggtgcc agcccccttg atctcccttc gtggagacga caccgggggtc ctctgcccc
 1141 ccccagggga ggcccaggat gctgatttga aggatgtaaa tgtgattcca gccaccgcct
 1201 gaccatccgc cattccgact gctaaaagcg aatgtagtca ggcccccttc atgctgtgag
 1261 acctctgga aactggcat ctctgagcct ccagaagggg ttctgggcct agttgtcctc
 5 1321 cctctggagc cccgtcctgt ggtctgcctc agtttccct cctaatacat atgctgttt
 1381 tccacctcga taatataaca cgagtttggg cccgaatcag tgtgttctca tcatttttca
 1441 ggcaggggag gtaagggaat aagtcggggg actgaatggc ggctgggcct cggatctctc
 1501 ctacaggtaa c (SEQ ID NO: 428).

By “fragment” is meant a portion of a polypeptide or nucleic acid molecule. This
 10 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. In some embodiments, the fragment is a functional fragment. By
 “guide polynucleotide” is meant a polynucleotide or polynucleotide complex which is
 15 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.

“Hybridization” means hydrogen bonding, which may be Watson-Crick, Hoogsteen
 20 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 “immunoglobulin gamma 1 (IgG1) polypeptide” is meant a protein having at least
 about 85% amino acid sequence identity to GenBank Accession No. CAA75030.1, provided
 25 below, or a fragment thereof having immunomodulatory activity. Exemplary IgG1 amino
 acid sequences from *Homo sapiens* is provided in **FIG. 2A**

>CAA75030.1 immunoglobulin kappa heavy chain [Homo sapiens]
 MEFGLRWVFLVAILKDVQCQDLVESGGGLVQPGGSLRLSCAASGFAYSSFWMHWRQAPGR
 GLVWVSRINPDGRITVYADAVKGRFTISRDNKNTLYLQMNNLRAEDTAVYYCARGTRFLEL
 30 TSRGQMDQWGQGLTLTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNS
 GALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDK
 THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEV
 HNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREP
 QVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYS
 35 KLTVDKSRWQQGNVFCFSVMHEALHNHYTQKSLSLSPGK (SEQ ID NO: 429).

By “immunoglobulin gamma 1 (IgG1) polynucleotide” is meant a nucleic acid molecule encoding an IgG1 polypeptide, as well as the introns, exons, 3' untranslated regions, 5' untranslated regions, and regulatory sequences associated with its expression, or fragments thereof. In embodiments, an IgG1 polynucleotide is the genomic sequence, cDNA, mRNA, or gene associated with and/or required for IgG1 expression. Exemplary IgG1 nucleotide sequences from *Homo sapiens* are provided below (GenBank: Y14735.1:36-1457):

>Y14735.1:36-1457 Homo sapiens mRNA for immunoglobulin kappa heavy chain

ATGGAATTTGGGCTGCGCTGGGTTTTCTTGTGCTATTTTAAAAGATGTCCAGTGTGACGT
 GCAACTGGTGGAGTCCGGGGGAGGCTTAGTTCAGCCTGGGGGGTCCCTGAGACTCTCCTGCG
 10 CAGCCTCTGGATTCGCCTACAGTAGTTTTTGGATGCACTGGGTCCGCCAAGCTCCAGGGAGG
 GGTCTGGTGTGGGTCTCACGTATTAATCCTGATGGGAGAATCACAGTCTACGCGGACGCCGT
 AAAGGGCCGATTACCATCTCCAGAGACAACGCCAAGAACACGCTCTATCTCCAAATGAACA
 ACCTGAGAGCCGAGGACACGGCTGTTTATTACTGTGCAAGAGGGACACGATTTCTGGAGTTG
 ACTTCTAGGGGACAAATGGACCAGTGGGGCCAGGGAACCCCTGGTCACTGTCTCCTCAGCCTC
 15 CACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCCTCCTCCAAGAGCACCTCTGGGGGCACAG
 CGGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGTGTGCTGGAACCTCA
 GGCGCCCTGACCAGCGGCGTGCACACCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTC
 CCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAACG
 TGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAAA
 20 ACTCACACATGCCACCGTGCCAGCACCTGAACTCCTGGGGGGACCGTCAGTCTTCTCTT
 CCCCCCAAACCCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGG
 TGGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGAGAGTG
 CATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAAGCGT
 CCTCACCGTCTTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACA
 25 AAGCCCTCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCA
 CAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTGAGCCTGACCTG
 CCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGG
 AGAACAACTACAAGACCACGCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCTCTACAGC
 AAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCA
 30 TGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGA (SEQ
 ID NO: 430).

By “immunoglobulin gamma 2 (IgG2) polypeptide” is meant a protein having at least about 85% amino acid sequence identity to GenBank Accession No. AAB59393.1, provided below, or a fragment thereof having immunomodulatory activity. Exemplary IgG2 amino

acid sequences from *Homo Sapiens* are provided below, including GenBank Accession No. AH005273.2:216-509,902-937,1056-1382,1480-1802, and in **FIG. 2A**:

>AAB59393.1 immunoglobulin gamma-2 heavy chain, partial [*Homo sapiens*]

STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLY
 5 SLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVKCCVECPPCAPPVAGPSVFLFPPK
 PKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSFLT
 VHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVK
 GFYPSDIAVEWESNGQPENNYKTTTPMLDSGDSFFLYSKLTVDKSRWQQGNVFSQVMHEAL
 HNHYTQKSLSLSPGK (SEQ ID NO: 431).

10 >exemplary IgG2 amino acid sequence

ASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL
 YSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVKCCVECPPCAPPVAGPSVFLFPP
 KPKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSFLT
 VVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLV
 15 KGFYPSDISVEWESNGQPENNYKTTTPMLDSGDSFFLYSKLTVDKSRWQQGNVFSQVMHEA
 LHNHYTQKSLSLSPGK (SEQ ID NO: 432).

By “immunoglobulin gamma 2 (IgG2) polynucleotide” is meant a nucleic acid molecule encoding an IgG2 polypeptide, as well as the introns, exons, 3' untranslated regions, 5' untranslated regions, and regulatory sequences associated with its expression, or fragments thereof. In embodiments, an IgG2 polynucleotide is the genomic sequence, cDNA, mRNA, or gene associated with and/or required for IgG2 expression. An exemplary IgG2 nucleotide sequence from *Homo sapiens* is provided below (GenBank: AH005273.2:216-509,902-937,1056-1382,1480-1802):

>AH005273.2:216-509,902-937,1056-1382,1480-1802 *Homo sapiens* immunoglobulin gamma-2 heavy chain (IgH), immunoglobulin gamma-4 heavy chain (IgH), immunoglobulin epsilon chain constant region (IgH), and immunoglobulin alpha-2 heavy chain (IgH) genes, partial cds

CCTCCACCAAGGGCCCATCGGTCTTCCCCCTGGCGCCCTGCTCCAGGAGCACCTCCGAGAGC
 ACAGCCGCCCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGTGTCTGTGGAA
 30 CTCAGGCGCTCTGACCAGCGCGTGCACACCTTCCAGCTGTCCTACAGTCCTCAGGACTCT
 ACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAACTTCGGCACCCAGACCTACACCTGC
 AACGTAGATCACAAGCCCAGCAACACCAAGGTGGACAAGACAGTTGAGCGCAAATGTTGTGT
 CGAGTGCCACCGTGCCAGCACCTGTGGCAGGACCGTCAGTCTTCTCTTCCCCCAA
 AACCCAAGGACACCTCATGATCTCCCGGACCCCTGAGGTACAGTGCCTGGTGGTGGACGTG

AGCCACGAAGACCCCGAGGTCCAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGC
 CAAGACAAAGCCACGGGAGGAGCAGTTCAACAGCACGTTCCGTGTGGTCAGCGTCTCCACCG
 TTGTGCACCAGGACTGGCTGAACGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGGCCTC
 CCAGCCCCCATCGAGAAAACCATCTCCAAAACCAAAGGGCAGCCCCGAGAACCACAGGTGTA
 5 CACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCA
 AAGGCTTCTACCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAAC
 TACAAGACCACACCTCCCATGCTGGACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCAC
 CGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTC
 TGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGA (SEQ ID NO:
 10 433).

By “increases” is meant a positive alteration of at least 10%, 25%, 50%, 75%, or
 100%, or about 1.5 fold, about 2 fold, about 3-fold, about 4-fold, about 5-fold, about 6-fold,
 about 7-fold, about 8-fold, about 9-fold, about 10-fold, about 15-fold, about 20-fold, about
 25-fold, about 30-fold, about 35-fold, about 40-fold, about 45-fold, about 50-fold, or about
 15 100-fold.

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.

An “intein” is a fragment of a protein that is able to excise itself and join the
 20 remaining fragments (the exteins) with a peptide bond in a process known as protein splicing.

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
 25 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
 30 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 molecule 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.

The term “linker”, as used herein, refers to a molecule that links two moieties. In one embodiment, the term “linker” refers to a covalent linker (*e.g.*, covalent bond) or a non-covalent linker.

By “marker” is meant any analyte, protein or polynucleotide having an alteration in expression, level, structure, or activity that is associated with a disease or disorder. In embodiments, the marker is an IgG polypeptide capable of binding an autoantigen and/or associated with an autoimmune disease or an FcRn polypeptide. The term “mutation” or “alteration” as used herein, refers to a substitution of a residue within a polynucleotide or polypeptide sequence another nucleotide or residue, or a deletion or insertion of one or more nucleotides or residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino

acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The terms “nucleic acid” and “nucleic acid molecule,” as used herein, refer to a
5 compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, “nucleic acid” refers to individual nucleic acid residues (*e.g.*, nucleotides and/or nucleosides). In some
10 embodiments, “nucleic acid” refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, “nucleic acid” encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for
15 example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including
20 non-naturally occurring nucleotides or nucleosides. Furthermore, the terms “nucleic acid,” “DNA,” “RNA,” and/or similar terms include nucleic acid analogs, *e.g.*, analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules,
25 nucleic acids 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-
30 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).

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 (SEQ ID NO: 190), KRPAATKKAGQAKKKK (SEQ ID NO: 191), KKTELQTTNAENKTKKL (SEQ ID NO: 192), KRGINDRNFWRGENGRKTR (SEQ ID NO: 193), RKSGKIAAIVVKRPRK (SEQ ID NO: 194), PKKKRKV (SEQ ID NO: 195), or MDSLLMNRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 196).

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-hydrmethylcytosine. 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. Non-limiting examples of modified nucleobases and/or chemical modifications that a modified nucleobase may include are the following: pseudo-uridine, 5-Methyl-cytosine, 2'-O-methyl-3'-phosphonoacetate, 2'-O-methyl thioPACE (MSP), 2'-O-methyl-PACE (MP), 2'-fluoro RNA (2'-F-RNA), constrained ethyl (S-cEt), 2'-O-methyl ('M'), 2'-O-methyl-3'-phosphorothioate ('MS'), 2'-O-methyl-3'-thiophosphonoacetate ('MSP'), 5-methoxyuridine, phosphorothioate, and N1-Methylpseudouridine.

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 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 is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/Cas Φ (Cas12j/Casphi). Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/Cas Φ , Cpf1, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, homologues thereof, or modified or engineered versions thereof. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, 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. Exemplary nucleic acid programmable DNA binding proteins and nucleic acid sequences encoding nucleic acid programmable DNA binding proteins are provided in the Sequence Listing as SEQ ID NOs: 197-230, and 378.

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).

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

By “subject” or “patient” is meant a mammal, including, but not limited to, a human or non-human mammal. In embodiments, the mammal is a bovine, equine, canine, ovine, rabbit, rodent, nonhuman primate, or feline. In an embodiment, “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 is associated with a disease or disorder or 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. A protein, peptide, or polypeptide can be naturally occurring, recombinant, or synthetic, or any combination thereof.

5 The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins.

 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
10 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%.

15 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. In embodiments, a reference is a cell or subject not contacted with
20 a base editor system provided herein, or a component thereof. In some cases, a reference is a cell or subject administered an agent (e.g., a small molecule drug) that interferes with the activity of FcRn in a subject. In some cases, a reference is an FcRn polypeptide that does not comprise an alteration at an amino acid residue of interest, or that does not contain any of the alterations provided herein (i.e., a wild-type FcRn polypeptide sequence). In various
25 instances, a reference is a cell that has not been altered according to the methods provided herein.

 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
30 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.

5 The term “RNA-programmable nuclease,” and “RNA-guided nuclease” refer to a nuclease that forms a complex 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
10 (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (CsnI) from *Streptococcus pyogenes* (e.g., SEQ ID NO: 197), Cas9 from *Neisseria meningitidis* (NmeCas9; SEQ ID NO: 208), Nme2Cas9 (SEQ ID NO: 209), *Streptococcus constellatus* (ScoCas9), or derivatives thereof (e.g. a sequence with at least about 85% sequence identity to a Cas9, such as Nme2Cas9 or spCas9).

15 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)
20 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
25 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
30 described 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.

By “specifically binds” is meant a nucleic acid molecule, polypeptide, polypeptide/polynucleotide complex, 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.

- 5 By “substantially identical” is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence. In one embodiment, a reference sequence is a wild-type amino acid or nucleic acid sequence. In another embodiment, a reference sequence is any one of the amino acid or nucleic acid sequences described herein. In one embodiment, such a sequence is at least about 60%, 80%, 85%,
 10 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or even 99.99%, 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,
 15 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
 20 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,
 25 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:

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

f) END GAP OPEN: 10; and

g) END GAP EXTEND: 0.5.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a functional fragment thereof.

5 Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having “substantial identity” to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the

10 invention or a functional fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having “substantial identity” to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By “hybridize” is meant pair to form a double-stranded molecule between

15 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).

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

20 trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, *e.g.*, formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more

25 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

30 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.

The term “target site” refers to a sequence within a nucleic acid molecule that is modified. In embodiments, the modification is deamination of a base. The deaminase can be a cytidine or an adenine deaminase. The fusion protein or base editing complex comprising a deaminase may comprise a dCas9-adenosine deaminase fusion protein, a Cas12b-adenosine deaminase fusion, or a base editor disclosed herein.

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 composition as described herein.

By “uracil glycosylase inhibitor” or “UGI” is meant an agent that inhibits the uracil-excision repair system. Base editors comprising a cytidine deaminase convert cytosine to uracil, which is then converted to thymine through DNA replication or repair. In various embodiments, a uracil DNA glycosylase (UGI) prevent base excision repair which changes the U back to a C. In some instances, contacting a cell and/or polynucleotide with a UGI and a base editor prevents base excision repair which changes the U back to a C. An exemplary UGI comprises an amino acid sequence as follows:

>spIP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA
PEYKPWALVIQDSNGENKIKML (SEQ ID NO: 231).

In some embodiments, the agent inhibiting the uracil-excision repair system is a uracil stabilizing protein (USP). See, e.g., WO 2022015969 A1, incorporated herein by reference.

As used herein, the term “vector” refers to a means of introducing a nucleic acid sequence into a cell. Vectors include plasmids, transposons, phages, viruses, liposomes, lipid nanoparticles, and episomes. “Expression vectors” are nucleic acid sequences comprising the nucleotide sequence to be expressed in the recipient cell. Expression vectors contain a polynucleotide sequence as well as additional nucleic acid sequences to promote and/or facilitate the expression of the introduced sequence, such as start, stop, enhancer, promoter, and secretion sequences, into the genome of a mammalian cell. Examples of vectors include nucleic acid vectors, e.g., DNA vectors, such as plasmids, RNA vectors, viruses or other suitable replicons (e.g., viral vectors). A variety of vectors have been developed for the delivery of polynucleotides encoding exogenous proteins into a prokaryotic or eukaryotic cell. Examples of such expression vectors are disclosed in, e.g., WO 1994/11026; incorporated herein by reference. Certain vectors that can be used for the expression of editors, e.g., base editors or prime editors, and/or guide polynucleotides of some aspects and embodiments herein include plasmids that contain regulatory sequences, such as promoter

and enhancer regions, which direct gene transcription. Other useful vectors for expression of antibodies and antibody fragments contain polynucleotide sequences that enhance the rate of translation of these genes or improve the stability or nuclear export of the mRNA that results from gene transcription. These sequence elements include, e.g., 5' and 3' untranslated regions, an internal ribosomal entry site (IRES), and polyadenylation signal site in order to direct efficient transcription of the gene carried on the expression vector. The expression vectors of some aspects and embodiments herein may also contain a polynucleotide encoding a marker for selection of cells that contain such a vector. Examples of a suitable marker include genes that encode resistance to antibiotics, such as ampicillin, chloramphenicol, kanamycin, or nourseothricin.

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.

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. Any

embodiments specified as “comprising” a particular component(s) or element(s) are also contemplated as “consisting of” or “consisting essentially of” the particular component(s) or element(s) in some embodiments. 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.

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

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A and 1B provide a 3D stick structure and a plot taken from *European Journal of Immunology*, 29:2819-2825 (1999), the disclosure of which is incorporated herein by reference in its entirety for all purposes. **FIG. 1A** provides a 3D stick structure of the Fc region of human IgG1. The figure was prepared using the RASMOL program (Roger Sayle, Bioinformatics Research Institute, University of Edinburg, GB). **FIG. 1B** provides a plot showing elimination curves showing FcRn interaction with IgG of recombinant human Fc-hinge derivatives and Fc-papain fragment in mice.

FIGs. 2A and 2B provide a multiple sequence alignment and a ribbon structure of IgG2 bound to FcRn. **FIG. 2A** provides an alignment of IgG1 and IgG2 amino acid sequences, with important binding residues underlined. **FIG. 2B** provides a ribbon structure showing binding of IgG2 to FcRn, where important residues are indicated. The following sequences are depicted in **FIG. 2A** from top-to-bottom:

LGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQY
NSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDEL
TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQG
NVFSCSVMEALHNHYTQKSLSLSPGK (SEQ ID NO: 434) and
VAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHNAKTKPREEQF
NSTFRVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQVYTLPPSREEM

TKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPMLDSDGSFFLYSKLTVDKSRWQQG
NVFSCSVMEALHNHYTQKSLSLSPGK (SEQ ID NO: 435).

FIGs. 3A and 3B provide ribbon structures relating to the FcRn:IgG interface.

FIG. 4 provides a ribbon structure relating to the FcRn:IgG binding site, with
5 important residues indicated.

FIG. 5 provides a ribbon structure of FcRn bound to IgG, where structures of FcRn amino acids important for forming a complex with IgG are depicted using spheres. In **FIG. 5**, amino acids forming part of a hydrophobic pocket helping to position W131, an important residue for IgG binding, are shown as a cluster of amino acids depicted using spheres of the lightest shade of grey. In **FIG. 5**, amino acids corresponding to pH dependent FcRn IgG
10 binding sites are depicted using a cluster of spheres of the darkest shade of grey. In **FIG. 5**, amino acids associated with stabilization of the complex between IgG and FcRn and reduced binding affinity at neutral pH are depicted using spheres of an intermediate shade of grey. Alteration of the amino acids depicted in **FIG. 5** using spheres can reduce binding to and
15 recycling of IgG1, IgG2, IgG3, and/or IgG4 while, in various embodiments, advantageously preserving albumin recycling and FcRn expression. In some instances, alterations to amino acid residues of FcRn are associated with a >50% reduction in circulating IgGs *in vivo*.

FIGs. 6A and 6B provide bar graphs showing base editing rates achieved when HEK293T cells were contacted with base editing systems containing the guide
20 polynucleotides and base editors (i.e., ABE or CBE) indicated on the x-axis. The base editors used were, SpCas9-ABE8.8, spCas9-BE4, VRQR spCas9-ABE8.8, VRQR spCas9-BE4, KKH-saCas9-ABE8.8, KKH-saCas9-BE4, SaABE8.8, SaBE4, and spCas9-ABE. Base editing rates are shown for each particular FcRn alteration or combination of alterations that were observed in base-edited cells. In **FIG. 6A**, bars corresponding to base editing systems
25 that achieved base editing efficiencies of greater than 40% are outlined by shaded boxes in **FIG. 6A**. The base editing system containing an adenosine base editor (ABE) and the guide RNA gRNA1583 achieved a base editing efficiency of over 70% in HEK293T cells and introduced the W131R alteration to FcRn. **FIG. 6B** depicts a subset of the data presented in **FIG. 6A**. The arrow in **FIG. 6B** indicates a bar corresponding to the base editing efficiency
30 measured for the combined amino acid alteration containing E116K and M118I. In **FIG. 6A**, the rightmost four bars correspond to positive control base editor systems. In **FIG. 6B**, the rightmost two bars correspond to positive control base editor systems. In **FIG. 6A**, the amino acid positions listed along the x-axis are numbered from the first amino acid of the FcRn 23 amino acid-long signal peptide.

FIGs. 7A – 7D provide results from surface plasmon resonance (SPR) measurements for binding of albumin or IgG1 to FcRn polypeptides. **FIGs. 7A and 7B** provide bar graphs showing results from surface plasmon resonance (SPR) measurements for binding of albumin or IgG1 to FcRn polypeptides containing the ten alterations indicated on the x-axis. All ten FcRn variants maintained albumin binding. **FIG. 7A** provides a bar graph showing surface plasmon resonance measurements of albumin binding to FcRn polypeptides containing the alterations indicated on the X-axis. **FIG. 7B** provides a bar graph showing surface plasmon resonance measurements of IgG1 binding to FcRn polypeptides containing the alterations indicated on the X-axis. In **FIG. 7B**, arrows indicate amino acid alterations that were associated with a significant reduction in IgG1 binding by FcRn. Four of the FcRn variants evaluated showed reduced IgG binding. **FIG. 7C** shows a comparison of wild-type, M118I, and W131R FcRn binding to IgG. The measurements were performed with FcRn-biotin on the surface. IgG was injected at the indicated concentrations. **FIG. 7D** shows a comparison of wild-type, M118I, and W131R binding to albumin. The measurements were performed with FcRn-biotin on the surface. Albumin was injected at the indicated concentrations.

FIGs. 8A-8C provide a schematic diagram and bar graphs. **FIG. 8A** provides a schematic diagram depicting an experimental schema used to evaluate base editing in a primary human hepatocytes (PHH) co-culture. In **FIG. 8A**, “MC” indicates a media change, “TF” indicates transfection with a base editing system, “NGS” indicates next-generation sequencing, and “RT-qPCR” indicates reverse transcriptase quantitative polymerase chain reaction. Samples were collected for next-generation sequencing at day 10 post-transfection and samples were taken for RT-qPCR measurements at day 13 post-transfection. Cells were transfected using a sub-saturating dose of a base editing system (600 ng total containing 160 ng end-modified guide polynucleotide + 450 ng mRNA encoding the base editor). The gRNA1583 guide, which facilitated creation of the W131(154)R alteration performed well in the PHH co-culture system. **FIG. 8B** provides a bar graph showing base editing efficiencies associated with the particular FcRn alterations indicated on the x-axis and achieved using the base editor systems indicated on the x-axis. **FIG. 8C** provides a bar graph showing levels of exon 5-6 and exon 4-5 of *FCGRT* detected in mRNA isolated from transfected cells. Transcript levels were normalized to transcript levels measured for ACTB. Cells edited using the base editor system containing gRNA1583 and an adenosine base editor showed a decrease of about 30% in FcRn mRNA expression compared to untreated cells and cells edited using the guide sg23. In **FIGs. 8B and 8C**, a base editor system containing the guide

g23 (alternatively referred to as gRNA23) and an ABE base editor was used as a positive control.

FIGs. 9A and 9B provide a schematic diagram and a bar graph relating to spacer-length optimization in HEK293T cells. HEK293T cells were transfected with mRNA encoding an adenosine base editor and the guide RNAs indicated on the x-axis, which contained spacers varying in length from 19 to 23 nucleotide. **FIG. 9A** provides a schematic summarizing an experimental design for evaluating the impact of spacer length on base editing efficiencies. HEK293T cells were seeded at Day 0 and transfected with a base editor system at Day 1. Media was changed at day 2 and genomic DNA from the cells was sequenced 72-hours post transfection using next-generation sequencing. **FIG. 9B** shows base editing efficiencies associated with the indicated FcRn alterations created using the indicated base editing systems. Cells were transfected using a sub-saturating dose of a base editing system (600 ng total containing 160 ng end-modified guide polynucleotide + 450 ng mRNA encoding the base editor). All spacer lengths evaluated showed similar base editing efficiencies for the primary alterations achieved.

DETAILED DESCRIPTION OF THE INVENTION

The invention features compositions and methods for editing, modifying expression, and/or silencing the neonatal Fc receptor (FcRn) gene, *FCGRT*.

The invention is based, at least in part, on the discovery that base editing can be used to alter FcRn polypeptides encoded by cells, such that the polypeptides show reduced binding to IgG while maintaining binding to albumin. Therefore, in various embodiments, the methods and base editing systems provided herein can be used to treat IgG-mediated autoimmune disorders by introducing alterations to FcRn that reduce the binding thereof to IgG, thereby advantageously reducing IgG half-life in a subject in need of treatment, while maintaining the beneficial function of FcRn in albumin cycling.

Accordingly, the disclosure provides improved compositions and methods for treatment of FcRn-mediated autoimmune disorders.

The details of embodiments of the presently-disclosed subject matter are set forth in this document. Modifications to embodiments described in this document, and other embodiments, will be evident to those of ordinary skill in the art after a study of the information provided in this document.

Genome editing involves the molecular manipulation of genetic material by deleting, replacing, or inserting a nucleotide sequence of a target gene, optionally to effect a correction

of a genetic mutation of the gene. In embodiments, genome editing comprises CRISPR systems, base editing, prime editing, and the like.

Clustered regularly interspaced short palindromic repeat (CRISPR) systems are naturally occurring bacterial and archaea defense mechanisms against viruses. CRISPR systems have been adapted for genome editing by introducing double stranded DNA breaks (DSBs) or RNA breaks at user-defined loci in living cells. Porto, et al., *Base editing: advances and therapeutic opportunities*, Nature Reviews 19: 839-59 (2020). CRISPR methods include use of a guide RNA and a nucleic acid programmable DNA binding domain Cas protein, which together introduce a break in the target nucleotide sequence. Cas proteins include Cas9, catalytically inactivated (dead) dCas9, nCas9 (nickase), Cas12, and Cas13. Repair of the break by non-homologous end joining (NHEJ) or homology directed repair (HDR) introduces insertions, deletions, or point mutations at the site of the break. The non-specific nature of the mutation may introduce frame shifts in the target nucleotide sequence.

Base editing allows for the direct conversion of target residues at a specific locus, without introducing DSBs. Base editing directly introduces single-nucleotide modifications into DNA or RNA of living cells. Base editors include those targeting DNA and RNA. DNA base editors comprise a nucleic acid programmable DNA binding domain and cytidine deaminase domains that convert a target C-G to T-A or a target G-C to A-T in a target region of the DNA, e.g., the *FCGRT* gene, or adenosine deaminase domains that convert a target A-T to G-C or a target T-A to C-G in a target region of DNA, e.g., the *FCGRT* gene. In some embodiments a base editor comprising a cytidine deaminase domain further comprises uracil glycosylase inhibitor (UGI). Base editing techniques are described in detail, for example, in Porto, et al. (2020), which is incorporated herein by reference in its entirety. In embodiments, the nucleic acid programmable DNA binding domain comprises a catalytically inactivated (dead) Cas9 (dCas9) or a Cas9 nickase (nCas9).

Prime editing retains CRISPR's target specificity, while incorporating an edited RNA template extending from the guide RNA (prime editing guide RNA, or "pegRNA") and reverse transcriptase fused to the nCas9. See, e.g., Scholefield, et al., *Prime editing: an update on the field*, Gene Therapy 28: 396-401 (2021). nCas9 does not introduce DSBs, but instead nicks the non-complementary strand of DNA upstream of the PAM site. This nickase exposes a DNA overhang having a 3' OH, which binds to the primer binding site (PBS) of the pegRNA. This serves as a primer for the reverse transcriptase, which fills in the 3' overhang by copying the edited sequence of the pegRNA. The 5' overhang is excised and the strands are ligated to complete the edit. Prime editing techniques are described in detail in Scholefield, et al. (2021),

which is incorporated herein by reference in its entirety. In embodiments, the prime editor comprises a nucleic acid programmable DNA binding domain and a reverse transcriptase and the guide RNA is a prime editing guide RNA (pegRNA), wherein the prime editor replaces one or more nucleotides in the *FCGRT* gene with a different nucleotide. In embodiments, the
5 nucleic acid programmable DNA binding domain comprises a catalytically inactivated (dead) Cas9 (dCas9) or a Cas9 nickase (nCas9).

In another embodiment, a method of modifying an FcRn protein in a mammalian cell is provided, the method comprising contacting the cell with a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of
10 an *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene. In embodiments, the genome editor comprises a base editor or a prime editor.

In another embodiment, a method of treating an IgG-mediated autoimmune disorder in
15 a subject in need thereof is provided, the method comprising modifying FcRn protein in a mammalian cell of the subject. In specific embodiments, modifying the FcRn protein comprises genome editing an *FCGRT* gene in the mammalian cell of the subject. Optionally, the genome editing comprises contacting the mammalian cell with a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a
20 portion of the *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene. In specific embodiments, the genome editor comprises a base editor or a prime editor.

The genome editor may be delivered to the mammalian cell of interest via a variety of
25 delivery techniques known in the art. In embodiments, the genome editor is delivered to the mammalian cell via a nanoparticle, a viral vector, or electroporation. Nanoparticles suitable for use in the present compositions and methods include inorganic nanoparticles (e.g., gold), lipid-based particles (e.g., lipid nanoparticles, liposomes, exosomes, cell-derived membrane-bound particles, etc.), peptide nanoparticles, polymer nanoparticles, and the like.

30 Various viral vectors are known in the art and suitable for use in delivering the compositions of the present disclosure. In embodiments, the viral vector is selected from the group consisting of a retrovirus (e.g., HIV, lentivirus), an adenovirus, an adeno-associated virus (AAV), a herpesvirus (e.g., HSV), and a sendai virus.

The compositions and methods disclosed herein modify the nucleic acid encoding an FcRn protein by introducing one or more single nucleotide modifications in the *FCGRT* gene. In embodiments, the modified or variant FcRn protein exhibits reduced ability to bind to an Fc region of an IgG antibody. In further embodiments, the modified or variant FcRn protein
5 comprises at least one amino acid alteration relative to a reference FcRn protein, such as a wild type FcRn protein.

The presently disclosed methods can be carried out *ex vivo*, *in vitro*, or *in vivo*. That is, the compositions disclosed herein may be administered directly to a subject (e.g., intravenously, or locally, by injection, inhalation, etc.), or may be administered to a cell,
10 optionally a cell obtained from a subject. In embodiments, the subject is a human.

Various modifications may be made to the *FCGRT* gene to provide a modified FcRn protein as disclosed herein. In embodiments, the modified FcRn protein differs from a reference FcRn protein at one or more amino acids selected from the group consisting of: leucine (L) at position 112, glutamic acid (E) at position 115, glutamic acid (E) at position 116,
15 tryptophan (W) at position 131, proline (P) at position 132, and glutamic acid (E) at position 133. In other embodiments, the modified FcRn protein comprises one or more mutations as set forth in **FIG. 4**.

Optionally, the genome editor or delivery vehicle is conjugated to or incorporates a targeting moiety that binds to FcRn or albumin. In certain embodiments, the targeting moiety
20 is selected from the group consisting of an Fc domain of IgG, an antibody that specifically binds FcRn, an antibody that specifically binds albumin, a peptide that binds albumin, albumin, or a fragment or derivative thereof.

Additional targeting moieties include, but are not limited to, variant Fc domains; antibodies or other specific binding agents (e.g., engineered scaffold proteins such as affibodies, darpins, or peptides (which may be selected using display technologies such as phage display)) that bind to the extracellular domain of FcRn; albumin or a fragment or variant thereof that retains ability to bind to FcRn. In this approach, albumin (or fragment/variant) binds to the FcRn and the delivery vehicle/active agent is internalized along with the albumin (or fragment/variant). Other targeting moieties include other specific binding agents (e.g.,
25 engineered scaffold proteins such as affibodies or darpins or peptides (which may be selected using display technologies such as phage display)) that bind to albumin but do not substantially prevent binding of albumin to FcRn. The delivery vehicle will be internalized by cells along with albumin when albumin binds to the FcRn.

Also provided herein are compositions comprising a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of the *FCGRT* gene and targets the base genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene. The disclosed compositions may further comprise a delivery vehicle, as described herein, and/or a targeting moiety that binds to FcRn and/or albumin.

In a specific embodiment, a delivery vehicle as disclosed herein comprises a guide RNA and a genome editor or a nucleic acid that encodes a genome editor. In embodiments, the delivery vehicle comprises a targeting moiety that binds to FcRn and/or albumin.

Lipid nanoparticles (LNPs) are spherical nanometer-scale particles comprising an ionizable lipid monolayer shell and a lipid core matrix that can solubilize lipophilic molecules, such as drugs or nucleic acids. Traditional LNPs are taken up by host cells via endocytosis, escape the endosome, and release their cargo into the cytoplasm of the host cell. LNPs are generally regarded as safe, effective, and suitable for industrial manufacture and clinical use in drug delivery.

Embodiments of the presently disclosed LNPs include an Fc region or fragment of an Fc region of an IgG antibody or other targeting moiety embedded or incorporated into the lipid monolayer shell, and enclose within the core a nucleic acid for silencing or modulating expression of FcRn (*FCGRT* gene). When the LNP contacts FcRn on the surface of an epithelial cell, the Fc region or fragment thereof binds FcRn and the LNP fuses or is otherwise internalized with the cell and delivers its payload. A released nucleic acid then silences, modulates, or moderates expression of FcRn, which in turn results in reduced circulation of IgG (but preferably not albumin) in the host and a reduction of autoimmune disorder symptoms and pathologies.

In one embodiment, a solid LNP is provided, comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane. In embodiments, the lipid core of the LNP comprises at least one nucleic acid.

In one embodiment, the IgG or fragment thereof incorporated in the LNP is IgG1 subclass. In a specific embodiment, IgG1 or a fragment thereof has the following amino acid substitutions: aspartic acid at position 265 is substituted for alanine, or proline at position 238 is substituted for alanine.

In another embodiment, the IgG incorporated in the LNP is IgG2 subclass or a fragment thereof.

In another embodiment, the IgG incorporated in the LNP is IgG3 subclass or a fragment thereof.

In another embodiment, the IgG incorporated in the LNP is IgG4 subclass or a fragment thereof.

5 In another embodiment, the IgG incorporated in the LNP recognizes FcRn receptor. In specific embodiments, IgG1 or a fragment thereof has the following amino acid substitutions: aspartic acid at position 265 is substituted for alanine, or proline at position 238 is substituted for alanine.

10 In another embodiment, the IgG is not incorporated in the LNP and recognizes FcRn receptor. In specific embodiment IgG1 or fragment thereof has the following amino acid substitutions: aspartic acid at position 265 is substituted for alanine or proline at position 238 is substituted for alanine. In specific embodiment the IgG and can directly deliver the payload.

In some embodiments an engineered Fc variant has increased affinity for FcRn at basic pH (e.g., a pH typical of the blood, e.g., 7.35-7.45) relative to a naturally occurring Fc region.

15 In embodiments, the nucleic acid incorporated into the lipid core of the LNP is DNA, or RNA. In a specific embodiment, the nucleic acid is a small interfering RNA (siRNA), a micro RNA (miRNA), guide RNA, pegRNA, or a short hairpin RNA (shRNA). In a very specific embodiment, the nucleic acid is an siRNA. In another specific embodiment, the nucleic acid is a guide RNA or a pegRNA. In another embodiment, the nucleic acid encodes a genome editor.

20 In embodiments, the siRNA is functional to modulate expression of one or more genes. In a specific embodiment, the siRNA modulates expression of *FCGRT*, the gene that encodes the neonatal Fc receptor (FcRn).

25 In embodiments, the nucleic acid incorporated in the LNP is a guide RNA which is functional to target a genome editor to edit or modify *FCGRT*, the gene that encodes the neonatal Fc receptor (FcRn). Suitable modifications of the *FCGRT* gene are set forth, for example, in **FIG. 4** of the present disclosure.

30 In particular embodiments, tryptophan residues at positions 51 or 61 and histidine at position 166 are not modified, as these amino acids are responsible for binding and half-life extension of human serum albumin.

Various lipids are suitable for use in the lipid monolayer of the disclosed LNPs. In embodiments, the lipid monolayer membrane is comprised of a lipid selected from the group consisting of lecithin, phosphatidylcholines, phosphatidic acid, phosphatidylethanolamines, phosphatidylglycerols, phosphatidylserines, phosphatidylinositols, cardiolipins, lipid-

polyethyleneglycol conjugates, and combinations thereof. In embodiments, the lipids of the lipid monolayer may be PEGylated, at least in part, in order to facilitate the avoidance of immune clearance of the LNP. In embodiments, the lipid monolayer may further comprise cholesterol as a stabilizer.

5 The lipid core matrix of the disclosed LNPs comprises a cationic lipid suitable for complexing with the nucleic acid in the core. As used herein, the term “cationic lipid” encompasses any of a number of lipid species that carry a net positive charge at physiological pH, which can be determined using any method known to one of skill in the art. Such lipids include, but are not limited to, the cationic lipids of formula (I) disclosed in International
 10 Application No. PCT/US2009/042476, entitled “Methods and Compositions Comprising Novel Cationic Lipids,” which was filed on May 1, 2009, and is herein incorporated by reference in its entirety. These include, but are not limited to, N-methyl-N-(2-(arginoylamino) ethyl)- N, N- Di octadecyl aminium chloride or di stearyl arginyl ammonium chloride] (DSAA), N,N-di-myristoyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl
 15 ammonium chloride (DMGLA), N,N-dimyristoyl-N-methyl-N-2[N2-guanidino-L- lysiny] aminoethyl ammonium chloride, N,N-dimyristoyl-N-methyl-N-2[N'-(N2, N6- di-guanidino-L-lysiny)] aminoethyl ammonium chloride, and N, N-di-stearyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride (DSGLA). Other non-limiting examples of cationic lipids that can be present in the liposome or lipid bilayer of the presently
 20 disclosed lipid nanoparticles include N,N-dioleoyl-N,N-dimethylammonium chloride (DODAC); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTAP); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTMA) or other N- (N,N-l- dialkoxy)-alkyl-N,N,N-trisubstituted ammonium surfactants; N,N-distearyl- N,N-dimethylammonium bromide (DDAB); 3-(N-(N',N'-dimethylaminoethane)- carbamoyl)
 25 cholesterol (DC-Choi) and N-(1,2-dimyristyloxyprop-3-yl)-N,N- dimethyl-N-hydroxyethyl ammonium bromide (DMRIE); 1,3-dioleoyl-3- trimethylammonium-propane, N-(1-(2,3- dioleoyloxy)propyl)-N-(2- (sperminecarboxamido)ethyl)-N,N-dimethy- 1 ammonium trifluoro- acetate (DOSPA); GAP-DLRIE; DMDHP; 3-p[4N-(H8N-diguanidino spermidine)- carbamoyl] cholesterol (BGSC); 3-P[N,N-diguanidinoethyl-aminoethane)-carbamoyl]
 30 cholesterol (BGTC); N,N\N2,N3 Tetra-methyltetrapalmitylspermine (cellfectin); N-t-butyl-N'- tetradecyl-3-tetradecyl-aminopropion-amidine (CLONfectin); dimethyldioctadecyl ammonium bromide (DDAB); 1,3-dioleoyloxy-2-(6-carboxyspermyl)-propyl amide (DOSPER); 4-(2,3-bis-palmitoyloxy-propyl)- 1-methyl- 1H-imidazole (DPIM) N,N,N',N'- tetramethyl-N,N'-bis(2-hydroxyethyl)-2,3 dioleoyloxy- 1 ,4- butanediammonium iodide (Tfx-

50); 1,2 dioleoyl-3-(4'-trimethylammonio) butanol-sn- glycerol (DOBT) or cholesteryl (4'trimethylammonia) butanoate (ChOTB) where the trimethylammonium group is connected via a butanol spacer arm to either the double chain (for DOTB) or cholesteryl group (for ChOTB); DL-1,2-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORI) or DL-
 5 1,2-0-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORIE) or analogs thereof as disclosed in International Application Publication No. WO 93/03709, which is herein incorporated by reference in its entirety; 1,2-dioleoyl-3-succinyl-sn-glycerol choline ester (DOSC); cholesteryl hemisuccinate ester (ChOSC); lipopolyamines such as dioctadecylamidoglycylspermine (DOGS) and dipalmitoyl phosphatidylethanolamylspermine
 10 (DPPES), or the cationic lipids disclosed in U.S. Pat. No. 5,283,185, which is herein incorporated by reference in its entirety; cholesteryl-3P- carboxyl-amido-ethylenetrimethylammonium iodide; 1-dimethylamino-3- trimethylammonio-DL-2-propyl-cholesteryl carboxylate iodide; cholesteryl-3- β - carboxyamidoethyleneamine; cholesteryl-3-P-oxysuccinamido- ethylenetrimethylammonium iodide; 1-dimethylamino-3-trimethylammonio-
 15 DL-2- propyl-cholesteryl-3-P-oxysuccinate iodide; 2-(2-trimethylammonio)-ethylmethylamino ethyl-cholesteryl-3-P-oxysuccinate iodide; 3- β -N-(polyethyleneimine)-carbamoylcholesterol, DC-cholesterol; and N⁴-cholesteryl-spermine HCl salt (GL67).

In embodiments, the lipid core matrix further comprises cholesterol as a stabilizer.

In another embodiment, a pharmaceutical composition is provided, comprising: at least
 20 one LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one nucleic acid; and at least one pharmaceutically-acceptable excipient.

Optionally, the pharmaceutical composition is formulated for local or systemic
 25 administration to a subject. Administration to deliver compounds of the combination therapy systemically or to a desired surface or target can include, but is not limited to, injection, infusion, instillation, and inhalation administration. Injection includes, without limitation, intravenous, intramuscular, intraarterial, intrathecal, intraventricular, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous,
 30 subcuticular, and intraarticular injection and infusion.

Pharmaceutical compositions for injection include aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include, but are not limited to, physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered

saline (PBS). The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. Fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, and sodium chloride may be included in the composition. The resulting solutions can be packaged for use as is, or lyophilized; the lyophilized preparation can later be combined with a sterile solution prior to administration.

In another embodiment, a method of treating an IgG-mediated autoimmune disorder in a subject in need thereof is provided, the method comprising administering to the subject a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one siRNA or guide RNA that moderates expression of or silences an *FCGRT* gene.

IgG-mediated autoimmune disorders include, but are not limited to, myasthenia gravis, warm autoimmune hemolytic anemia (WAIHA), idiopathic thrombocytopenia purpura (ITP), Grave's disease, chronic inflammatory demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and hemolytic diseases of fetus and newborn (HDFN).

In another embodiment, a method of silencing FcRn expression in a cell is provided, the method comprising contacting the cell with a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one siRNA that silences an *FCGRT* gene. In embodiments, the method is *ex vivo*, *in vivo*, or *in vitro*.

FcRn

Immunoglobulin G (IgG) (see, e.g., **FIGs. 1 and 2A**) is the most common type of antibody found in blood circulation and extracellular fluids where it controls infection of body tissues. While IgG can directly bind antigen, the neonatal Fc receptor for IgG (FcRn) also binds receptors on cells to effect an immune response. The family of Fc gamma receptors (FcγR) includes the atypical neonatal Fc receptor (FcRn), encoded by the *FCGRT* gene. FcRn functions to recirculate and maintain IgG and albumin, as well as transport IgG and albumin across polarized cellular barriers, thereby increasing the half-life of IgG and

albumin in circulation. FcRn also interacts with and facilitates antigen presentation of peptides derived from IgG immune complexes (IC).

FcRn was first identified as the receptor that transports maternal IgG antibodies from mother to child facilitating passive humoral immunity in the child from the mother. FcRn binds to the Fc region of monomeric immunoglobulin gamma (see **FIGs. 1B and 2B-5**) and mediates its selective uptake from milk. IgG in the milk is bound at the apical surface of the intestinal epithelium. The resultant FcRn-IgG complexes are transcytosed across the intestinal epithelium and IgG is released from FcRn into blood or tissue fluids. Throughout life, contributes to effective humoral immunity by recycling IgG and extending its half-life in the circulation. Mechanistically, monomeric IgG binding to FcRn in acidic endosomes of endothelial and hematopoietic cells recycles IgG to the cell surface where it is released into the circulation.

Initially, it was believed that FcRn was only present in placental and intestinal tissues during the fetal and newborn stages. However, FcRn is now known to be expressed in many tissues throughout the body, including epithelia, endothelia, and cells of hematopoietic origin. Specifically, FcRn expression in the epithelia has been detected in the intestines, placenta, kidney, and liver.

Mechanistically, monomeric IgG binding to FcRn in acidic endosomes of endothelial and hematopoietic cells recycles IgG to the cell surface where it is released into the circulation. In addition to IgG, FcRn regulates homeostasis of the other most abundant circulating protein albumin/ALB.

FcRn is expressed in many tissues. For example, FcRn is expressed in the liver, hepatocytes, and Muller cells. FcRn is also expressed highly on epithelial, endothelial, and myeloid lineages and performs multiple roles in adaptive immunity. On myeloid cells, FcRn participates in both phagocytosis and antigen presentation together with classical FcγR and complement. In podocytes (kidney), FcRn reabsorbs IgG from the glomerular basement membrane which prevents deposition of immune complexes that might lead to glomerular diseases.

A number of autoimmune disorders are caused by the reaction of IgG to autoantigens, including, for example, myasthenia gravis(gMG), warm autoimmune haemolytic anaemia (wAIHA), idiopathic thrombocytopenia purpura (ITP), Grave's disease, chronic inflammatory demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and haemolytic diseases of fetus and newborn (HDFN). As FcRn functions to maintain IgG levels in circulation, FcRn also extends the half-life of antibodies that give rise to such autoimmune

disorders. Intravenous immunoglobulin (IVIg) is a recently developed therapy that saturates FcRn's IgG recycling capacity and reduces the levels of pathogenic IgG binding to FcRn, thereby facilitating the reduction in levels of IgG autoantibodies.

5 Efgartigimod (ARGX-113, VYVGART) is an IV/SC treatment developed by Argenx to initially treat Myasthenia Gravis (gMG). Efgartigimod is an IgG1 Fc fragment with increased affinity for FcRn. Efgartigimod blocks access to FcRn for IgG and reduces the overall serum half-life thereof. Administration of Efgartigimod (about 10 mg/kg/week administered using one IV infusion) to a subject has been associated with a 50-70% decrease in IgGs in the subject.

10 Various modifications may be made to the *FCGRT* gene to provide a modified FcRn protein as disclosed herein. The modifications impact the serum half-life of IgG in a subject containing FcRn proteins modified according to the methods provided herein. In
embodiments, the modified FcRn protein differs from a reference FcRn protein at one or more amino acids selected from the group consisting of: leucine (L) at position 112, glutamic
15 acid (E) at position 115, glutamic acid (E) at position 116, tryptophan (W) at position 131, proline (P) at position 132, and glutamic acid (E) at position 133. In other embodiments, the modified FcRn protein comprises one or more alterations as set forth in **Table 1** any of **FIGs. 2B, 4-7B, 8B, 8C, and 9B** and/or an alteration at position M118(141) (e.g., M118(141)I).

Table 1. Exemplary target FcRn alterations impacting the FcRn:IgG interface

Position in FcRn	Amino acid to be modified	Codon of amino acid to be modified	Mutated amino acid	Codon of mutated amino acid
112	Leucine (L)	CTG	Phenylalanine	TTC
			Proline	CCT
			Alanine	GCA, GCC, GCG, GCT
115	Glutamic acid (E)	GAG	Lysine	AAG
			Glycine	GGG
			Glutamine	GAA
			Alanine	GCA, GCC, GCG, GCT
116	Glutamic acid (E)	GAG	Lysine	AAG
			Glycine	GGG
			Glutamine	GAA
			Alanine	GCA, GCC, GCG, GCT
119	Asparagine (N)	AAT	Serine	AGT
			Alanine	GCA, GCC, GCG, GCT
122	Leucine (L)	CTC	Proline	CCC
			Phenylalanine	TTC
			Alanine	GCA, GCC, GCG, GCT
126	Threonine (T)	ACC	Alanine	GCC
			Isoleucine	ATC
			Alanine	GCA, GCG, GCT
127	Tryptophan (W)	TGG	Arginine	CGG

Position in FcRn	Amino acid to be modified	Codon of amino acid to be modified		Mutated amino acid	Codon of mutated amino acid
130	Aspartic acid (D)	GAC		Alanine	GCA, GCC, GCG, GCT
				Asparagine	AAC
				Glycine	GGC
131	Tryptophan	TGG		Alanine	GCA, GCC, GCG, GCT
				Arginine	CGG
				Alanine	GCA, GCC, GCG, GCT
132	Proline	CCC		Serine	TCC
				Leucine	CTC
				Alanine	GCA, GCC, GCG, GCT
133	Glutamic acid	GAG		Alanine	GCG
				Glycine	GGG
				Lysine	AAG
135	Leucine	CTG		Alanine	GCA, GCG, GCT
				Proline	CCG
				Alanine	GCA, GCC, GCG, GCT

In some embodiments, the methods and compositions of the present disclosure are used to introduce an alteration to one or more of the amino acids underlined or in bold in the below FcRn amino acid sequence, where **bold** residues are involved in IgG binding, underlined residues are involved in albumin binding, and the **bold-underline-italic** residue

5 corresponds to M118(141):

1 mgvprpqpwa lglllfllpg slgaeshls lylhtavssp apgtpafwvs gwlgpqqyls
 61 ynslrgeaep cgawvwenqv swywekettd lrikeklfle afkalggkqp ytlqgllgce
 121 lgpdnsvpt akfalngeef mnfdlkqgtw ggdwpealai sqrwqqqdka ankeltfllf
 181 scphrlrehl ergrgnlew eppsmrlkar psspgfsvlt csafsfpype lqlrflrngl
 10 241 aagtgggdfg pnsdgsfhas ssltvksgde hhyccivqha glaqlrvel espakssvlv
 301 vgivigvlll taaavggall wrrmrsglpa pwislrghdt gvllptpgea qdadlkdvrv
 361 ipata (SEQ ID NO: 427).

In embodiments, the methods provided herein are used to produce an FcRn containing alterations that modify one or more of the following properties of the FcRn: A) stability of a
 15 complex formed between the FcRn and an IgG (e.g., reduce or increase); B) binding affinity for IgG at neutral pH (e.g., reduce or increase); C) binding affinity for IgG at pH lower or higher than neutral (e.g., reduce or increase); D) positioning of W131 (e.g., to reduce or increase binding to IgG).

In particular embodiments, tryptophan residues at positions 51 or 61 and histidine at
 20 position 166 are not modified, as these amino acids are responsible for binding and half-life extension of human serum albumin.

In another embodiment, a method of silencing FcRn expression in a cell is provided, the method comprising contacting the cell with a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment
 25 thereof embedded therein; and a lipid core matrix enclosed in the lipid monolayer membrane, wherein the lipid core matrix comprises at least one siRNA that silences an *FCGRT* gene. In embodiments, the method is *ex vivo*, *in vivo*, or *in vitro*.

EDITING OF TARGET GENES

30 In some embodiments, to produce the gene edits described herein, cells (e.g., cells from a subject, such as hepatocytes, endothelial cells, epithelial cells, or myeloid cells) are contacted *in vivo* or *in vitro* with one or more guide RNAs and a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a cytidine deaminase or adenosine deaminase. In some embodiments, cells to be edited are
 35 contacted with at least one polynucleotide, wherein said polynucleotide(s) encodes one or

more guide RNAs and a nucleobase editor polypeptide comprising a nucleic acid programmable DNA binding protein (napDNAbp) and a cytidine deaminase. In some embodiments, the gRNA comprises one or more nucleotide analogs. In some instances, the gRNA is added directly to a cell. In some embodiments, these nucleotide analogs can inhibit degradation of the gRNA from cellular processes.

In various instances, it is advantageous for a spacer sequence to include a 5' and/or a 3' "G" nucleotide. In some embodiments, for example, any spacer sequence or guide polynucleotide provided herein comprises or further comprises a 5' "G", where, in some embodiments, the 5' "G" is or is not complementary to a target sequence. In some embodiments, the 5' "G" is added to a spacer sequence that does not already contain a 5' "G." For example, it can be advantageous for a guide RNA to include a 5' terminal "G" when the guide RNA is expressed under the control of a U6 promoter or the like because the 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 (2013) doi: 10.1126/science.1231143). In some embodiments, a 5' terminal "G" is added to a guide polynucleotide that is to be expressed under the control of a promoter, but is optionally not added to the guide polynucleotide if or when the guide polynucleotide is not expressed under the control of a promoter.

In embodiments, a guide polynucleotide comprises a scaffold sequence containing a nucleotide sequence selected from
 GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
 CACCGAGUCGGUGCUUUU (SpCas9 scaffold; SEQ ID NO: 317) and
 GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAUGCCGUGUUUA
 UCUCGUCAACUUGUUGGCGAGAUUUU (SaCas9 scaffold; SEQ ID NO: 436).

Tables 2A and 2B provide exemplary gRNA sequences (e.g., full guide sequences and spacer sequences) suitable for use in embodiments of the disclosure.

Table 2A: Exemplary guide RNA sequences

Guide Number	Guide Name	Guide Polynucleotide Sequence	SEQ ID NO
1	gRNA1560	CAUGAAUUUCGACCCUACAAGCGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	437
2	gRNA1561	AUGAAUUUCGACCCUACAAGCAGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	438
3	gRNA1562	AGGGCACCCUGGGGUGGGGACGCUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	439
4	gRNA1563	UGGGGACUGGCCCGAGGCCCGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	440
5	gRNA1564	CUCGGCCAGUCCCCACCCCGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	441
6	gRNA1565	CACCCAGGUGCCCUGCUUGGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	442
7	gRNA1566	GCCGUUCAGGGCGAAACUUGGGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	443
8	gRNA1567	GUUCAGGGCGAAACUUGGGGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	444
9	gRNA1568	UUCAGGGCGAAACUUGGGGUGUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	445
10	gRNA1569	UCGACCUCAAGCAGGGCACCCGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	446
11	gRNA1570	CGACCUCAAGCAGGGCACCCUGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	447
12	gRNA1571	GACCUCAAGCAGGGCACCCUGGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	448
13	gRNA1572	CAUGAACUCCUCGCCGUUCAGUUUUUAGAGCUAGAAAUAAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	449

Guide Number	Guide Name	Guide Polynucleotide Sequence	SEQ ID NO
14	gRNA1573	GGCGAGGAGUUCAUAGAAUUUGUUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	450
15	gRNA1574	CCACCCAGGUGCCUGCUGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	451
16	gRNA1575	CCAGGUGCCUGCUUUGAGGUGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	452
17	gRNA1576	CUGCUUGAGGUCGAAAUUCAGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	453
18	gRNA1577	GAACUCCUGCCGUAUAGGGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	454
19	gRNA1578	GUUCAUGAAUUUCGACCCUCAGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	455
20	gRNA1579	UGAAUUUCGACCUCAAAGCAGGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	456
21	gRNA1580	GGGCACCUUGGGUGGGGACUGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	457
22	gRNA1581	GGGACUGGCCCGAGGCCUGUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	458
23	gRNA1582	CGAGGCCUGGCUAUCAGUCGUUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	459
24	gRNA1583	GGCCAGUCCCCACCCAGGUUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	460
25	gRNA1584	UCAGGGCGAACUUUGCGGUGGUUUUUAGAGCUAGAAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUAUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUU	461
26	gRNA1587	GCCUGAACGGCGAGGAGUUCGUUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUUGCGGAGAUUUU	462
27	gRNA1588	UUCGACCUCAAGCAGGGCACCUGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUUGCGGAGAUUUU	463
28	gRNA1589	GACUGGCCCGAGGGCCUGGCUUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUUGCGGAGAUUUU	464

Guide Number	Guide Name	Guide Polynucleotide Sequence	SEQ ID NO
29	gRNA1590	GACUGAUAGCCAGGGCCUCGGGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	465
30	gRNA1591	GGCCUCGGGCGAGUCCCAACCGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	466
31	gRNA1592	CCCCACCCAGGUGCCUCGUGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	467
32	gRNA1593	CUCGCCGUUCAGGGCGAACUUGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	468
13		AGGGCACCCUGGGGUGGGGACGUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACU UGAAAAAGUGGCACCGAGUCGGGUGCUUUU	469
28		AGUCCCAACCCAGUGCCCGUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACU UGAAAAAGUGGCACCGAGUCGGGUGCUUUU	470
29		AUGAACUCCUCGCCGUUCAGGUUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACU UGAAAAAGUGGCACCGAGUCGGGUGCUUUU	471
40		CUCGCCGUUCAGGGCGAACUUGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	472
39		GACUGGCCCGAGGGCCUCGGCUGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	473
32		GCCUGAACGGCGAGGAGUUCGUUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	474
27		GGCACCCUGGGGUGGGACUGUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACU UGAAAAAGUGGCACCGAGUCGGGUGCUUUU	475
38		UUCGACCUCAAGCAGGGCACCGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUUCUACUAAAAACAAGGCAAA AUGCCGUGUUUAUCUCGUCAACUUGUGGCGAGAUUUU	476
	gRNA3265	UCAUGAACUCCUCGCCGUUCGUUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACU UGAAAAAGUGGCACCGAGUCGGGUGCUUUU	477
	gRNA3025	GGCAGUCCCAACCCAGGGUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACUUGAAAAA GUGGCACCGAGUCGGGUGCUUUU	478
	gRNA1583	GGGCCAGUCCCAACCCAGGGUUUUAGAGCUAGAAAUAGCAAGUUUAAAAUAAAGGCUAGUCCGUAUUCAAACUUGAAAAA AGUGGCACCGAGUCGGGUGCUUUU	479

Guide Number	Guide Name	Guide Polynucleotide Sequence	SEQ ID NO
	gRNA3026	CGGGCCAGUCCCCACCCAGGGUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCUUUU	480
	gRNA3027	UCGGCCAGUCCCCACCCAGGGUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCUUUU	481
	gRNA3028	CUCGGCCAGUCCCCACCCAGGGUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUU	482
	gRNA3025	GGCCAGUCCCCACCCAGGGUUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCUUUU	483

Table 2B: Exemplary spacer sequence, target amino acid alterations, and base editors corresponding to the guide RNA sequences provided in Table 2A

Guide Number	Guide Name	Base Editor	Spacer Sequence	SEQ ID NO	PAM Sequence	Target Amino Acid Alteration*
1	gRNA1560	spCas9 ABE	CAUGAAUUUCGACCUCUACGC	484	AGG	N142G
2	gRNA1561	spCas9 ABE	AUGAAUUUCGACCUCUACGCA	485	GGG	N142G
3	gRNA1562	spCas9 ABE	AGGCACCUUGGGUGGGGAC	486	TGG	T149A
4	gRNA1563	spCas9 ABE	UGGGACUGGCCCGAGGCCCC	487	TGG	D153G
5	gRNA1564	spCas9 ABE	CUCGGCCAGUCCCCACCC	488	AGG	W154R
6	gRNA1565	spCas9 ABE	CACCCAGGUGCCUUGCUUG	489	AGG	W150R
7	gRNA1566	spCas9 ABE	GCCGUUCAGGGCGAACUUGG	490	CGG	L135P
8	gRNA1567	spCas9 ABE	GUUCAGGGCGAACUUGGCGG	491	TGG	L135P
9	gRNA1568	spCas9 ABE	UUCAGGGCGAACUUGGCGGU	492	GGG	L135P
10	gRNA1569	spCas9 CBE	UCGACCUCUACGAGGGCACC	493	TGG	L145F
11	gRNA1570	spCas9 CBE	CGACCUCUACGAGGGCACC	494	GGG	L145F
12	gRNA1571	spCas9 CBE	GACCUCUACGAGGGCACC	495	GGG	L145F
13	gRNA1572	spCas9 CBE	CAUGAACUCCUCCGCUUCA	496	GGG	E139K
14	gRNA1573	spCas9 VRQR ABE	GGCAGGAGUUCUUGAAUUU	497	CGA	E138G,E139G

Guide Number	Guide Name	Base Editor	Spacer Sequence	SEQ ID NO	PAM Sequence	Target Amino Acid Alteration*
15	gRNA1574	spCas9 VRQR ABE	CCCACCCACAGGUGCCCCUGCU	498	TGA	W150R
16	gRNA1575	spCas9 VRQR ABE	CCAGGUGCCUGCUUGAGGU	499	CGA	W150R
17	gRNA1576	spCas9 VRQR ABE	CUGCUUGAGGUCGAAAUUCA	500	TGA	L145P
18	gRNA1577	spCas9 VRQR CBE	GAACUCCUCCGCGUUCAGGG	501	CGA	E138K,E139K
19	gRNA1578	spCas9 NGC ABE	GUUCAUGAAUUUCGACCUCA	502	AGC	M141V,N142G
20	gRNA1579	spCas9 NGC ABE	UGAAUUUCCAGCCUCAAGCAG	503	GGC	N142G
21	gRNA1580	spCas9 NGC ABE	GGCACCUUGGGGUGGGGACU	504	GGC	T149A
22	gRNA1581	spCas9 NGC ABE	GGGACUGGGCCCGAGGCCCU	505	GGC	D153G
23	gRNA1582	spCas9 NGC ABE	CGAGGCCUUGGCUAUCAGUC	506	AGC	E156G
24	gRNA1583	spCas9 NGC ABE	GGGCCAGUCCCAACCCACAGG	507	TGC	W154R
25	gRNA1584	spCas9 NGC ABE	UCAGGGCGAACUUGGCGGUG	508	GGC	F133S,L135P
26	gRNA1587	saCas9 ABE	GCCUGAACGGCGAGGAGUUC	509	ATGAAT	N136G,E138G
27	gRNA1588	saCas9 CBE	UUCGACCUCAAGCAGGGCACC	510	TGGGGT	L145F
28	gRNA1589	saCas9 KKH ABE	GACUGGCCCGAGGCCUUGGCU	511	ATCAGT	E156G
29	gRNA1590	saCas9 KKH ABE	GACUGAUAGCCAGGGCCUCCGG	512	GCCAGT	L158P,I160T
30	gRNA1591	saCas9 KKH ABE	GGCCUCCGGCCAGUCCCCACC	513	CCAGGT	W154R
31	gRNA1592	saCas9 KKH ABE	CCCCACCCAGGUGCCCCUGCU	514	TGAGGT	W150R
32	gRNA1593	saCas9 KKH ABE	CUCGCCGUUCAGGGCGGAACUU	515	GGCGGT	L135P
13		spCas9 CBE	AGGGACCUUGGGGUGGGGAC	516	TGG	T149I
28		spCas9 NGC CBE	AGUCCCAACCCAGGUGCCCC	517	TGC	G151D,G152K,D153N
29		spCas9 NGC CBE	AUGAACUCCUCCGCGGUUCAG	518	GGC	E139K
40		saCas9 KKH CBE	CUCGCCGUUCAGGGCGGAACUU	519	GGCGGT	G137N,E138K
39		saCas9 KKH CBE	GACUGGCCCGAGGCCUUGGCU	520	ATCAGT	P155F
32		saCas9 KKH ABE	GCCCUGAACGGCGAGGAGUUC	521	ATGAAT	N136G,E138G
27		spCas9 NGC CBE	GGGCACCUUGGGGUGGGGACU	522	GGC	T149I

Guide Number	Guide Name	Base Editor	Spacer Sequence	SEQ ID NO	PAM Sequence	Target Amino Acid Alteration*
38		saCas9 KKH CBE	UUCGACCUCAAGCAGGGCACC	523	TGGGGT	L145F
	gRNA3265	spCas9 ABE	UCAUGAAGUCCUCGCCGUUC	524	AGG	F140P,M141T
	gRNA3025		GGCCAGUCCCCACCCCAGG	525		
	gRNA1583		GGCCAGUCCCCACCCCAGG	526		
	gRNA3026		CGGGCCAGUCCCCACCCCAGG	527		
	gRNA3027		UCGGGCCAGUCCCCACCCCAG G	528		
	gRNA3028		CUCGGGCCAGUCCCCACCCC GG	529		

* The positions of the alterations listed in Table 2B assume that the 23 amino acid signal peptide is included.

NUCLEOBASE EDITORS

Useful in the methods and compositions described herein are nucleobase editors that edit, modify or alter a target nucleotide sequence of a polynucleotide. Nucleobase editors described herein typically include 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 and thereby localize the base editor to the target nucleic acid sequence desired to be edited.

In certain embodiments, the nucleobase editors provided herein comprise one or more features that improve base editing activity. For example, any of the nucleobase editors provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the nucleobase editors 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). 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 opposite the targeted nucleobase. Mutation of the catalytic residue (*e.g.*, D10 to A10) prevents cleavage of the edited (*e.g.*, deaminated) strand containing the targeted residue (*e.g.*, A or C). Such Cas9 variants can 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 nucleobase change on the non-edited strand.

Polynucleotide Programmable Nucleotide Binding Domain

Polynucleotide programmable nucleotide binding domains bind polynucleotides (*e.g.*, RNA, DNA). A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains (*e.g.*, one or more nuclease domains). In some embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain comprises an endonuclease or an exonuclease. An endonuclease can cleave a single strand of a double-stranded nucleic acid or both strands of a double-stranded nucleic acid molecule. In some embodiments, a nuclease domain of a polynucleotide programmable nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide.

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 embodiments, 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 (*e.g.*, a functional portion) of a CRISPR protein (*i.e.* a base editor comprising as a domain all or a portion (*e.g.*, a functional 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.

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, Cse1, Cse2, 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 (*e.g.*, SEQ ID NO: 232), Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ, CAREF, DinG, homologues thereof, or modified versions thereof. 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. A Cas protein (*e.g.*, Cas9, Cas12) or a Cas domain (*e.g.*, Cas9, Cas12) can refer to a

polypeptide or domain 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 Cas polypeptide or Cas domain. Cas (e.g., Cas9, Cas12) can refer to the wild-type or a modified form of the Cas 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 (e.g., a functional portion) of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

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.*, *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., *et al.*, *Nature* 471:602-607(2011); and “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., *et al.*, *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences 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.

High Fidelity Cas9 Domains

Some aspects of the disclosure provide high fidelity Cas9 domains. High fidelity Cas9 domains are known in the art and described, for example, in Kleinstiver, B.P., *et al.* “High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.”

Nature 529, 490-495 (2016); and Slaymaker, I.M., *et al.* “Rationally engineered Cas9 nucleases with improved specificity.” *Science* 351, 84-88 (2015); the entire contents of each of which are incorporated herein by reference. An Exemplary high fidelity Cas9 domain is provided in the Sequence Listing as SEQ ID NO: 233. In some embodiments, high fidelity

5 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-phosphate backbone of DNA have less off-target effects. In some embodiments, the Cas9 domain (*e.g.*, a wild type Cas9

10 domain (SEQ ID NOs: 197 and 200) comprises one or more mutations that decrease the association between the Cas9 domain and the sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and the sugar-phosphate backbone of DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least

15 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 or complexes provided herein comprise one or more of a D10A, 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

20 any amino acid. In some embodiments, the high fidelity Cas9 enzyme is SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, or hyper accurate Cas9 variant (HypaCas9). In some embodiments, the modified Cas9 eSpCas9(1.1) contains alanine substitutions that weaken the interactions between the HNH/RuvC groove and the non-target DNA strand, preventing strand separation and cutting at off-target sites. Similarly, SpCas9-HF1 lowers off-target

25 editing through alanine substitutions that disrupt Cas9’s interactions with the DNA phosphate backbone. HypaCas9 contains mutations (SpCas9 N692A/M694A/Q695A/H698A) in the REC3 domain that increase Cas9 proofreading and target discrimination. All three high fidelity enzymes generate less off-target editing than wildtype Cas9.

30 ***Cas9 Domains with Reduced Exclusivity***

Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a “protospacer adjacent motif (PAM)” or PAM-like motif, which is a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. The presence of an NGG PAM sequence is

required 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 or complexes 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. Exemplary polypeptide sequences for spCas9 proteins capable of binding a PAM sequence are provided in the Sequence Listing as SEQ ID NOs: 197, 201, and 234-237. Accordingly, in some embodiments, any of the fusion proteins or complexes 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); the entire contents of each are hereby incorporated by reference.

20

Nickases

In some embodiments, the polynucleotide programmable nucleotide binding domain comprises a nickase domain. Herein the term “nickase” refers to a polynucleotide programmable nucleotide binding domain comprising a nuclease domain that is capable of cleaving only one strand of the two strands in a duplexed nucleic acid molecule (*e.g.*, DNA). In some embodiments, a nickase can be derived from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by introducing one or more mutations into the active polynucleotide programmable nucleotide binding domain. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can include a D10A mutation and a histidine at position 840. In such embodiments, 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 comprises an H840A mutation, while the amino acid residue at position 10 remains a D. In some embodiments, a nickase can be derived

30

from a fully catalytically active (*e.g.*, natural) form of a polynucleotide programmable nucleotide binding domain by removing all or a portion (*e.g.*, a functional 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
 5 derived from Cas9, the Cas9-derived nickase domain can comprise a deletion of all or a portion (*e.g.*, a functional portion) of the RuvC domain or the HNH domain.

In some embodiments, wild-type Cas9 corresponds to, or comprises the following amino acid sequence:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
 10 TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIE
 GDLNPDNSDVKLFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENLIAQ
 LPGEKKNGLFGNLIALLSLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLNLLAQIGDQ
 YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ
 15 LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
 RKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPIYYVGPLA
 RGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
 LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
 KIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTTLTLE
 20 DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
 SDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGikelGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 25 KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSklVSDFRKD
FQFYKvreINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIakse
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
 YSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIK
 30 LPKYSLFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQ
 KQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHrdKPIREQAENIIH
 LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD
 (SEQ ID NO: 197) (single underline: HNH domain; double underline: RuvC domain).

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, Cas12-derived nickase domain) is the strand that is not edited by the base editor (*i.e.*, the strand that is cleaved by the base editor is opposite to a strand comprising a base to be edited). In other embodiments, a base editor comprising a nickase domain (*e.g.*, Cas9-derived nickase domain, Cas12-derived nickase domain) can cleave the strand of a DNA molecule which is being targeted for editing. In such embodiments, the non-targeted strand is not cleaved.

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an “nCas9” protein (for “nickase” Cas9). 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:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
 30 RLKRTARRRYTRRKNRICYLQEIFSNEAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVKLFI
 QLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIALLSLGL
 TPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
 EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEF

YKFIKPILEKMDGTEELLVKLNRDILLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLK
 DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT
 NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
 VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
 5 LTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
 LKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVV
 DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
 NVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKH
 10 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVBREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNS
 DKLIARKKDWDPKKGFFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNP
 IDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNFYLYLAS
 15 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH RDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQ
 LGGD (SEQ ID NO: 201)

The Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9
 undergoes a conformational change upon target binding that positions the nuclease domains
 20 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.

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

Catalytically Dead Nucleases

Also provided herein are base editors comprising a polynucleotide programmable nucleotide binding domain which is catalytically dead (*i.e.*, incapable of cleaving a target polynucleotide sequence). Herein the terms “catalytically dead” and “nuclease dead” are used interchangeably to refer to a polynucleotide programmable nucleotide binding domain which has one or more mutations and/or deletions resulting in its inability to cleave a strand of a nucleic acid. In some embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain base editor can lack nuclease activity as a result of specific point mutations in one or more nuclease domains. For example, in the case of a base editor comprising a Cas9 domain, the Cas9 can comprise both a D10A mutation and an H840A mutation. Such mutations inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises one or more deletions of all or a portion (e.g., a functional 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 (e.g., a functional portion) of a nuclease domain. dCas9 domains are known in the art and described, for example, in 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.

Additional suitable nuclease-inactive dCas9 domains 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. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, e.g., Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

In some embodiments, 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. 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. In some embodiments, a nuclease-inactive Cas9 domain comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

5 In some embodiments, a variant Cas9 protein can cleave the complementary strand of a guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the RuvC domain. As a non-limiting example, in some embodiments, a variant Cas9 protein has a D10A (aspartate to
10 alanine at amino acid position 10) and can therefore cleave the complementary strand of a double stranded guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence (thus resulting in a single strand break (SSB) instead of a double strand break (DSB) when the variant Cas9 protein cleaves a double stranded target nucleic acid) (see, for example, Jinek *et al.*, *Science*. 2012
15 Aug. 17; 337(6096):816-21).

In some embodiments, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain
20 (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus
25 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).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors W476A and W1126A mutations such that the polypeptide has a reduced ability to
30 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 embodiments, 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 embodiments, the variant Cas9 protein
5 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 embodiments, the variant Cas9 protein harbors H840A, D10A, W476A, and W1126A, mutations such that the
10 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 embodiments, the variant Cas9 protein
15 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 embodiments, the variant Cas9 protein harbors D10A, H840A, P475A,
20 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 embodiments, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A,
25 W476A, N477A, D1125A, W1126A, and D1127A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such embodiments, 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 embodiments, 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
30 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.

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
 5 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.

10 In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises a N579A mutation, or a corresponding mutation in any of the amino acid sequences provided in the Sequence Listing submitted herewith.

15 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 NNGRRV PAM sequence. In some
 20 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
 25 mutations in any of the amino acid sequences provided herein.

In some embodiments, one of the Cas9 domains present in the fusion protein or complexes 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 Cas9 is an SaCas9. Residue A579 of SaCas9 can be mutated from N579 to yield a SaCas9
 30 nickase. Residues K781, K967, and H1014 can be mutated from E781, N967, and R1014 to yield a SaKKH Cas9.

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' was 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, unlike Cas9, 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 that are more similar to types I and III than type II systems. Functional Cpf1 does not require the trans-activating CRISPR RNA (tracrRNA), therefore, only CRISPR (crRNA) is required. This benefits genome editing because Cpf1 is not only smaller than Cas9, but also it has a smaller sgRNA molecule (approximately 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' or 5'-TTN-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 having an overhang of 4 or 5 nucleotides.

In some embodiments, the Cas9 is a Cas9 variant having specificity for an altered PAM sequence. In some embodiments, the Additional Cas9 variants and PAM sequences are described in Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the entirety of which is incorporated herein by reference. In some embodiments, a Cas9 variate have no specific PAM requirements. In some embodiments, a Cas9 variant, *e.g.* a SpCas9 variant has specificity for a NRNH PAM,

wherein R is A or G and H is A, C, or T. In some embodiments, the SpCas9 variant has specificity for a PAM sequence AAA, TAA, CAA, GAA, TAT, GAT, or CAC. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1218, 1219, 1221, 1249, 1256, 1264, 1290, 5 1318, 1317, 1320, 1321, 1323, 1332, 1333, 1335, 1337, or 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1135, 1218, 1219, 1221, 1249, 1320, 1321, 1323, 1332, 1333, 1335, or 1337 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 10 1219, 1221, 1256, 1264, 1290, 1318, 1317, 1320, 1323, 1333 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1131, 1135, 1150, 1156, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1286, 1293, 1320, 1321, 1332, 1335, 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 15 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338, 1349 or a corresponding position thereof. Exemplary amino acid substitutions and PAM specificity of SpCas9 variants are shown in **Tables 3A-3D**.

Table 3A SpCas9 Variants and PAM specificity

PAM	SpCas9 amino acid position												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

Table 3B SpCas9 Variants and PAM specificity

PAM	SpCas9 amino acid position																		
	1114	1134	1135	1137	1139	1151	1180	1188	1211	1219	1221	1256	1264	1290	1318	1317	1320	1323	1333
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N	A	A	R
GAA										V	H						V		K
GAA			N	S						V							V	D	K
GAA			N							V	H		Y				V		K
CAA			N							V	H		Y				V		K
CAA	G		N	S						V	H		Y				V		K
CAA			N					R		V	H						V		K
CAA			N				G		R	V	H		Y				V		K
CAA			N							V	H		Y				V		K
AAA			N				G			V	H	R	Y				V	D	K
CAA	G		N				G			V	H		Y				V	D	K
CAA		L	N				G			V	H		Y			T	V	D	K
TAA	G		N				G			V	H		Y	G	S		V	D	K
TAA	G		N			E	G			V	H		Y		S		V		K
TAA	G		N				G			V	H		Y		S		V	D	K
TAA	G		N				G		R	V	H						V		K
TAA			N				G		R	V	H		Y				V		K
TAA	G		N		A		G			V	H						V		K
TAA	G		N							V	H						V		K

Table 3C SpCas9 Variants and PAM specificity

PAM	SpCas9 amino acid position															
	111	113	113	115	115	118	119	121	121	122	122	124	125	128	129	132
	4	1	5	0	6	0	1	8	9	1	7	9	3	6	3	0
	R	Y	D	E	K	D	K	G	E	Q	A	P	E	N	A	A
SacB.TA T			N				N		V	H						V
SacB.TA T			N					S	V	H		S				
AAT			N					S	V	H	V	S		K	T	
TAT	G		N			G		S	V	H		S	K			
TAT	G		N			G		S	V	H		S				
TAT	G	C	N			G		S	V	H		S				
TAT	G	C	N			G		S	V	H		S				
TAT	G	C	N			G		S	V	H		S				
TAT	G	C	N			G		S	V	H		S				
TAT	G	C	N		E	G		S	V	H		S				
TAT	G	C	N	V		G		S	V	H		S				
TAT		C	N			G		S	V	H		S				
TAT	G	C	N			G		S	V	H		S				

Table 3D SpCas9 Variants and PAM specificity

PAM	SpCas9 amino acid position														
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349	
	R	D	D	D	E	E	N	N	P	D	R	T	S	H	
SacB.CAC			N			V				N	Q	N			
AAC	G		N			V				N	Q	N			
AAC	G		N			V				N	Q	N			
TAC	G		N			V				N	Q	N			
TAC	G		N			V		H		N	Q	N			
TAC	G		N		G	V	D	H		N	Q	N			
TAC	G		N			V				N	Q	N			
TAC	G	G	N	E		V		H		N	Q	N			
TAC	G		N			V		H		N	Q	N			
TAC	G		N			V				N	Q	N	T	R	

Further exemplary Cas9 (e.g., SaCas9) polypeptides with modified PAM recognition are described in Kleinstiver, et al. "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition," *Nature Biotechnology*, 33:1293-1298 (2015) DOI: 10.1038/nbt.3404, the disclosure of which is incorporated herein by reference in its entirety for all purposes. In some embodiments, a Cas9 variant (e.g., a SaCas9 variant) comprising one or more of the alterations E782K, N929R, N968K, and/or R1015H has specificity for, or is associated with increased editing activities relative to a reference polypeptide (e.g., SaCas9) at an NNNRRT or NNHRRT PAM sequence, where N represents any nucleotide, H represents any nucleotide other than G (i.e., "not G"), and R represents a purine. In embodiments, the Cas9 variant (e.g., a SaCas9 variant) comprises the alterations E782K, N968K, and R1015H or the alterations E782K, K929R, and R1015H.

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

In some embodiments, the napDNAbp is a circular permutant (e.g., SEQ ID NO: 238).

The crystal structure of *Alicyclobaccillus acidoterrastris* Cas12b/C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See e.g., Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell*, 2017 Jan. 19; 65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See e.g., Yang *et al.*,

“PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease”, *Cell*, 2016 Dec. 15; 167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with Cas12b/C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between Cas12b/C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins or complexes 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 least 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 least 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.

In some embodiments, a napDNAbp refers to Cas12c. In some embodiments, the Cas12c protein is a Cas12c1 (SEQ ID NO: 239) or a variant of Cas12c1. In some embodiments, the Cas12 protein is a Cas12c2 (SEQ ID NO: 240) or a variant of Cas12c2. In some embodiments, the Cas12 protein is a Cas12c protein from *Oleiphilus* sp. HI0009 (*i.e.*, OspCas12c; SEQ ID NO: 241) or a variant of OspCas12c. These Cas12c molecules have been described in Yan *et al.*, “Functionally Diverse Type V CRISPR-Cas Systems,” *Science*, 2019 Jan. 4; 363: 88-91; the entire contents of which is hereby incorporated by reference. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12c1, Cas12c2, or OspCas12c 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 Cas12c1, Cas12c2, or OspCas12c protein described herein. It should be appreciated that Cas12c1, Cas12c2, or OspCas12c from other
5 bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12g, Cas12h, or Cas12i, which have been described in, for example, Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," *Science*, 2019 Jan. 4; 363: 88-91; the entire contents of each is hereby incorporated by reference. Exemplary Cas12g, Cas12h, and Cas12i polypeptide sequences
10 are provided in the Sequence Listing as SEQ ID NOs: 242-245. By aggregating more than 10 terabytes of sequence data, new classifications of Type V Cas proteins were identified that showed weak similarity to previously characterized Class V protein, including Cas12g, Cas12h, and Cas12i. In some embodiments, the Cas12 protein is a Cas12g or a variant of Cas12g. In some embodiments, the Cas12 protein is a Cas12h or a variant of Cas12h. In
15 some embodiments, the Cas12 protein is a Cas12i or a variant of Cas12i. It should be appreciated that other RNA-guided DNA binding proteins may be used as a napDNAbp, and are within the scope of this disclosure. 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
20 least 99.5% identical to a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12g, Cas12h, or Cas12i 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 Cas12g, Cas12h, or
25 Cas12i protein described herein. It should be appreciated that Cas12g, Cas12h, or Cas12i from other bacterial species may also be used in accordance with the present disclosure. In some embodiments, the Cas12i is a Cas12i1 or a Cas12i2.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins or complexes provided herein may be a
30 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%, 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 present disclosure.

Fusion Proteins or Complexes with Internal Insertions

Provided herein are fusion proteins or complexes 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 a deaminase (*e.g.*, cytidine or adenosine deaminase) or a functional 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. In some embodiments, the cytidine deaminase is an APOBEC deaminase (*e.g.*, APOBEC1). In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10 or TadA*8). In some embodiments, the TadA is a TadA*8 or a TadA*9. TadA sequences (*e.g.*, TadA7.10 or TadA*8) as described herein are suitable deaminases for the above-described fusion proteins or complexes.

In some embodiments, the fusion protein comprises the structure:

NH₂-[N-terminal fragment of a napDNAbp]-[deaminase]-[C-terminal fragment of a napDNAbp]-COOH;

NH₂-[N-terminal fragment of a Cas9]-[adenosine deaminase]-[C-terminal fragment of a Cas9]-COOH;

NH₂-[N-terminal fragment of a Cas12]-[adenosine deaminase]-[C-terminal fragment of a Cas12]-COOH;

NH₂-[N-terminal fragment of a Cas9]-[cytidine deaminase]-[C-terminal fragment of a Cas9]-COOH;

NH₂-[N-terminal fragment of a Cas12]-[cytidine deaminase]-[C-terminal fragment of a Cas12]-COOH;

wherein each instance of “[”-“]” indicates the optional presence of a linker (*i.e.*, the linker is optionally present).

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, 136, or 65 as numbered in a TadA reference sequence.

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

 In some embodiments, the napDNAbp in the fusion protein or complex is a Cas9 polypeptide or a fragment thereof. The Cas9 polypeptide can be a variant Cas9 polypeptide. In some embodiments, the Cas9 polypeptide is a Cas9 nickase (nCas9) polypeptide or a
15 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 or complex can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein or complex 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-
20 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
25 (St1Cas9), or fragments or variants of any of the Cas9 polypeptides described herein.

 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
30 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;
 5 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 optional presence of a linker.

- In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*,
 10 deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10). 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
 15 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;
 20 NH₂-[cytidine deaminase]-[Cas9(TadA*8)]-COOH;
 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 optional presence of a linker.

- 25 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
 30 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, 5 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 10 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 15 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 20 Cas9, or a Cas9 domain lacking partial or complete HNH domain.

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 25 inserted between amino acid positions 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 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 30 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, 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 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:

5 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, 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,

10 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 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

15 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:

20 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

25 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.

30 In some embodiments, a cytidine deaminase (*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 cytidine deaminase 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 cytidine deaminase 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 cytidine deaminase 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.

5 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 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 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 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 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 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 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
5 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
10 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
15 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
20 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
25 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
30 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.

- 5 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 4 below:

Table 4: 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
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term fusion single TadA helix truncated 165-end	ISLAY21
IBE029	TadA-Circular Permutant116 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

10

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 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.

15

20

In some embodiments, the Cas9 polypeptide lacks one or more domains selected from 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 nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some
5 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-terminal and a C-terminal fragment of a napDNAbp. In some embodiments, the fusion
10 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-
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
15 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
20 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
25 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 deaminase 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.

30 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 or complex 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 or complex 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
5 Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) of the fusion protein or complex deaminates no more than two nucleobases within the range of an R-loop. In some embodiments, the deaminase of the fusion protein or complex deaminates no more than three nucleobases
10 within the range of the R-loop. In some embodiments, the deaminase of the fusion protein or complex 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
15 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,
20 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 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
25 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 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 or complex described herein can effect target deamination in an
30 editing window different from canonical base editing. In some embodiments, a target nucleobase is 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 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, 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 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 upstream of the PAM sequence.

The fusion protein or complex can comprise more than one heterologous polypeptide. For example, the fusion protein or complex 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 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 (SEQ ID NO: 246), (GGGGS)_n (SEQ ID NO: 247), (G)_n, (EAAAK)_n (SEQ ID NO: 248), (GGS)_n, SGSETPGTSESATPES (SEQ ID NO: 249). In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 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 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 or complex is a Cas12 polypeptide, *e.g.*, Cas12b/C2c1, or a functional fragment thereof capable of associating with a nucleic acid (*e.g.*, a gRNA) that guides the Cas12 to a specific nucleic acid sequence. 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 GGSGGS (SEQ ID NO: 250) or

GSSGSETPGTSESATPESSG (SEQ ID NO: 251). In other embodiments, the linker is a rigid linker. In other embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC (SEQ ID NO: 252) or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC (SEQ ID NO: 253).

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:

- 5 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
 10 optional presence of a 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.*, TadA*7.10). In some embodiments, the TadA is a
 TadA*8. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase is
 15 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
 20 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.

25 In some embodiments, the “-” used in the general architecture above indicates the
 optional presence of a 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
 30 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 acidiphilus* Cas12b (SEQ ID NO: 254). In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b (SEQ ID NO: 255), *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 (SEQ ID NO: 256), *Bacillus* sp. V3-13 Cas12b (SEQ ID NO: 257), 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 embodiments, the Cas12 polypeptide contains BvCas12b (V4), which in some embodiments is expressed as 5' mRNA Cap---5' UTR---bhCas12b---STOP sequence --- 3' UTR --- 120polyA tail (SEQ ID NOs: 258-260).

In other embodiments, the catalytic domain is inserted between amino acid positions 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 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, 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 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 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 domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein or complex 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 MAPKKKRKVGIIHGVPAA (SEQ ID NO: 261). In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC (SEQ ID NO: 262). 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 or complex further contains a tag (*e.g.*, an influenza hemagglutinin tag).

In some embodiments, the fusion protein or complex comprises a napDNAbp domain (*e.g.*, Cas12-derived domain) with an internally fused nucleobase editing domain (*e.g.*, all or a portion (*e.g.*, a functional 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 5 below.

Table 5: 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

By way of nonlimiting example, an adenosine deaminase (*e.g.*, TadA*8.13) may be inserted into a BhCas12b to produce a fusion protein (*e.g.*, TadA*8.13-BhCas12b) that effectively edits a nucleic acid sequence.

- 5 In some embodiments, the base editing system described herein is an ABE with TadA inserted into a Cas9. Polypeptide sequences of relevant ABEs with TadA inserted into a Cas9 are provided in the attached Sequence Listing as SEQ ID NOs: 263-308.

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

- 10 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.

A to G Editing

- 15 In some embodiments, a base editor described herein comprises an adenosine deaminase domain. 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
- 20 in deoxyribonucleic acid (DNA). In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine
- 25 residue (*e.g.*, inosine), which can improve the activity or efficiency of the base editor.

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
5 deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of adenosine deaminase acting on RNA (ADAR, e.g., ADAR1 or ADAR2) or tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA
10 polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion (e.g., a functional 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 (e.g., a functional portion) of an ADAT from *Escherichia coli* (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y,
15 E155V, L84F, H123Y, I156F, or a corresponding mutation in another adenosine deaminase. Exemplary ADAT homolog polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 1 and 309-315.

The adenosine deaminase can be derived from any suitable organism (e.g., *E. coli*). In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments,
20 the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Shewanella putrefaciens*, *Haemophilus influenzae*, *Caulobacter crescentus*, or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations
25 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 correspond to any of the mutations described herein (e.g., any of the mutations identified in ecTadA) can be
30 generated accordingly.

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 a TadA reference sequence, such as TadA*7.10 (SEQ ID NO: 1)) 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). In some embodiments, the TadA reference sequence is TadA*7.10 (SEQ ID NO: 1). 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 a 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 a TadA reference sequence or another adenosine deaminase.

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

In some embodiments, the adenosine deaminase comprises an A106X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

5 In some embodiments, the adenosine deaminase comprises a E155X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a E155D, E155G, or E155V mutation in a TadA reference sequence, or a
10 corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a D147X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase
15 comprises a D147Y, mutation in a 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), where X indicates any amino acid other than the
20 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.

It should also be appreciated that any of the mutations provided herein may be made individually or in any combination in ecTadA or another adenosine deaminase. For example,
25 an adenosine deaminase may contain a D108N, a A106V, a E155V, and/or a D147Y mutation in a 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1), or corresponding mutations in another
30 adenosine deaminase: 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 may be made in an adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a combination of mutations in a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or corresponding mutations in another adenosine deaminase: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N.

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 a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or E85G, M94L, 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 a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

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

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 a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C,

Q154H or Q154R, E155G or E155V or E155D, K161Q, Q163H, and/or T166P mutation in a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

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

reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, 5 or six mutations selected from the group consisting of H8Y, D108N, N127S, D147Y, R152C, and Q154H in a 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 a 10 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 a TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the 15 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 a 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, 20 D108N, N127S, D147Y, and E155V in a 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 a TadA reference sequence, or a corresponding mutation or mutations in another adenosine 25 deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of the or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some 30 embodiments, the adenosine deaminase comprises a A106V and D108N mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and

Q154H mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and E155V mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y, and E155V mutation in a TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of S2X, H8X, I49X, L84X, H123X, N127X, I156X, and/or K160X mutation in a 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 a 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 in an 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 a 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H123Y mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an I156X mutation in a 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 I156F

mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

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 a TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2X, I49X, A106X, D108X, D147X, and E155X in a TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, A106X, D108X, N127X, and K160X in a TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

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

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D,

A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein corresponding to TadA reference sequence, or one or more
5 corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an E25X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M,
10 E25D, E25A, E25R, E25V, E25S, or E25Y mutation in a 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 a 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
15 adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation in a 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase,
20 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 a 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 a
25 TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A142N, A142D, A142G, mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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

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 a TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N, and/or K161T mutation in a 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H36L mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

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

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

In some embodiments, the adenosine deaminase comprises an R51X mutation in a 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an S146X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase,
5 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 a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an K157X mutation in a
10 TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a K157N mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

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

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

In some embodiments, the adenosine deaminase comprises an W23X mutation in a 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
30 adenosine deaminase. In some embodiments, the adenosine deaminase comprises a W23R or W23L mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R152X mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase,

where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a R152P or R52H mutation in a TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

- 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_I156F),
(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F), (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
5 (E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F), (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),
10 (R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(A106V_D108N_A142N_D147Y_E155V),
(R26G_A106V_D108N_A142N_D147Y_E155V),
(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
15 (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
(E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
(A106V_R107K_D108N_A142N_D147Y_E155V),
(A106V_D108N_A142N_A143G_D147Y_E155V),
(A106V_D108N_A142N_A143L_D147Y_E155V),
20 (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
25 (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
30 (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),

- (L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
5 (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_K157
N), (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
10 (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),
15 (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),
20 (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),
25 (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),
30 (H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N),

(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
5 (W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F_K157N),
10 (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),
15 (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),
20 (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).

25 In some embodiments, the TadA deaminase is a TadA variant. In some embodiments, the TadA variant is TadA*7.10. In particular embodiments, the fusion proteins or complexes comprise a single TadA*7.10 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises TadA*7.10 and TadA(wt), which are capable of forming heterodimers. In one embodiment, a fusion protein of the invention comprises a wild-type
30 TadA linked to TadA*7.10, which is linked to Cas9 nickase.

In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, the adenosine deaminase comprises an alteration in the following sequence:
TadA*7.10

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1)

In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or
5 166. In particular embodiments, TadA*7.10 comprises one or more of the following
alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. In other
embodiments, a variant of 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 +
10 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 some embodiments, a variant of TadA*7.10 comprises one or more of alterations
selected from the group of L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K,
15 and/or D167N. In some embodiments, a variant of TadA*7.10 comprises V82G, Y147T/D,
Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In other
embodiments, a variant of TadA*7.10 comprises a combination of alterations selected from
the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G +
Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G +
20 Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S +
N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G +
Y147D + F149Y + Q154S + N157K + D167N.

In some embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) comprises a
deletion. In some embodiments, an adenosine deaminase variant comprises a deletion of the
25 C terminus. 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding
mutation in another TadA.

In other embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) is a monomer
30 comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S,
T166R, and/or Q154R, relative to a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO:
1)), 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 a
 5 TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), 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 one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to a TadA
 10 reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), 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 +
 15 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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (e.g., TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a
 20 corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises 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 a
 25 TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.
 30

In some embodiments, an adenosine deaminase variant (e.g., MSP828) is a monomer comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to a TadA reference sequence (e.g.,

TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant (*e.g.*, MSP828) is a monomer comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA variant) is a monomer comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), 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: 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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a
 5 corresponding mutation in another TadA.

In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (e.g., TadA*7.10) each having one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a
 10 corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase variant domains (e.g., MSP828) each having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the
 15 adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (e.g., TadA*7.10) each having a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K;
 20 I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (e.g., TadA*8) comprising
 25 one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of 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 +
 30 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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase 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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (e.g., TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (e.g., MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (e.g., TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In particular embodiments, an adenosine deaminase heterodimer comprises a TadA*8 domain and an adenosine deaminase domain selected from *Staphylococcus aureus* (*S. aureus*) TadA, *Bacillus subtilis* (*B. subtilis*) TadA, *Salmonella typhimurium* (*S. typhimurium*) TadA, *Shewanella putrefaciens* (*S. putrefaciens*) TadA, *Haemophilus influenzae* F3031 (*H. influenzae*) TadA, *Caulobacter crescentus* (*C. crescentus*) TadA, *Geobacter sulfurreducens* (*G. sulfurreducens*) TadA, or TadA*7.10.

In some embodiments, an adenosine deaminase is a TadA*8. 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 (SEQ ID NO: 316)

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, 6, 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, 6, 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 comprising an adenosine deaminase variant (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of 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 a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (e.g., TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (e.g., MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (e.g., TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to a TadA reference sequence (e.g., TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in **Table 6**. **Table 6** 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 6** 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 6. Select TadA*8 Variants

		TadA amino acid number									
	TadA	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						
PACE	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
	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 some embodiments, the TadA variant is a variant as shown in **Table 6.1**. **Table 6.1** 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. In some

5 embodiments, the TadA variant is MSP605, MSP680, MSP823, MSP824, MSP825, MSP827, MSP828, or MSP829. In some embodiments, the TadA variant is MSP828. In some embodiments, the TadA variant is MSP829.

Table 6.1. TadA Variants

Variant	TadA Amino Acid Number							
	36	76	82	147	149	154	157	167
TadA-7.10	L	I	V	Y	F	Q	N	D
MSP605			G	T		S		
MSP680		Y	G	T		S		
MSP823	H		G	T		S	K	
MSP824			G	D	Y	S		N
MSP825	H		G	D	Y	S	K	N
MSP827	H	Y	G	T		S	K	
MSP828		Y	G	D	Y	S		N
MSP829	H	Y	G	D	Y	S	K	N

10

In one embodiment, a fusion protein or complex of the invention 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 or complexes comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the

15 fusion protein or complex comprises TadA*8 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.

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 GAVLVLNRRV IGEGWNRAIG ⁵⁰
 20 LHDPTAHAEI MALRQGGLVM QNYRLIDATL **Y**TFEPCVMC AGAMHSRIG ¹⁰⁰
 RVVFGVRNAK TGAAGSLMDV LHYPGMNRV EITEGILADE CAALLC**Y**FFR ¹⁵⁰
 MPRQVFNAQK KAQSS**T**D (SEQ ID NO: 1)

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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), 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 a TadA reference sequence (*e.g.*, TadA*7.10 (SEQ ID NO: 1)), or a corresponding mutation in another TadA.

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 or complex of the invention 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 or complexes 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.

In particular embodiments, the fusion proteins or complexes comprise a single (*e.g.*, provided as a monomer) TadA*8. In some embodiments, the TadA*8 is linked to a Cas9 nickase. In some embodiments, the fusion proteins or complexes of the invention comprise as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA*8. In other embodiments, the fusion proteins or complexes of the invention comprise as a heterodimer of a TadA*7.10 linked to a TadA*8. 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 and a TadA(wt). In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8 and TadA*7.10. In some embodiments, the base editor is ABE8 comprising a heterodimer of a TadA*8. In some embodiments, the TadA*8 is selected from **Table 6, 12, or 13**. In some embodiments, the ABE8 is selected from **Table 12, 13, or 15**.

In some embodiments, the adenosine deaminase is a TadA*9 variant. In some embodiments, the adenosine deaminase is a TadA*9 variant selected from the variants described below and with reference to the following sequence (termed TadA*7.10):

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
 MPRQVFNAQK KAQSSTD (SEQ ID NO: 1)

In some embodiments, an adenosine deaminase comprises one or more of the following alterations: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K. The one or more alternations are shown in the sequence above in underlining and bold font.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: V82S + Q154R + Y147R; V82S + Q154R + Y123H; V82S + Q154R + Y147R + Y123H; Q154R + Y147R + Y123H + I76Y + V82S; V82S + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; Q154R + Y147R + Y123H + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; V82S + Q154R + Y147R; Q154R + Y147R + Y123H + I76Y; Q154R + Y147R + Y123H + I76Y + V82S; I76Y_V82S_Y123H_Y147R_Q154R; Y147R + Q154R + H123H; and V82S + Q154R.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: E25F + V82S + Y123H, T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S + Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; and M70V + V82S + M94V + Y123H + Y147R + Q154R.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: Q71M + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; and V82S + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S +

Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V + V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; and M70V + Q71M + N72K + V82S + Y123H + Y147R + Q154R. In some embodiments, the adenosine deaminase is expressed as a monomer. In other embodiments, the adenosine deaminase is expressed as a heterodimer. In some embodiments, the deaminase or other polypeptide sequence lacks a methionine, for example when included as a component of a fusion protein.

This can alter the numbering of positions. However, the skilled person will understand that such corresponding mutations refer to the same mutation, *e.g.*, Y73S and Y72S and D139M and D138M.

In some embodiments, the TadA*9 variant comprises the alterations described in **Table 16** as described herein. In some embodiments, the TadA*9 variant is a monomer. In some embodiments, the TadA*9 variant is a heterodimer with a wild-type TadA adenosine deaminase. In some embodiments, the TadA*9 variant is a heterodimer with another TadA variant (*e.g.*, TadA*8, TadA*9). Additional details of TadA*9 adenosine deaminases are described in International PCT Application No. PCT/US2020/049975, which is incorporated herein by reference for its entirety.

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 a 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/US2017/045381 (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.

C to T Editing

In some embodiments, a base editor disclosed herein comprises a fusion protein or complex 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.

5 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
10 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
15 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
20 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, the base editor can comprise a uracil stabilizing protein as described herein. In another example, a base editor can incorporate a translesion polymerase to improve the efficiency of C-to-G
25 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).

 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
30 deaminase catalyzes a C nucleobase that is positioned in the context of a single-stranded portion of a polynucleotide. In some embodiments, the entire polynucleotide comprising a target C can be single-stranded. For example, a cytidine deaminase incorporated into the base editor can deaminate a target C in a single-stranded RNA polynucleotide. In other embodiments, a base editor comprising a cytidine deaminase domain can act on a double-

stranded polynucleotide, but the target C can be positioned in a portion of the polynucleotide which at the time of the deamination reaction is in a single-stranded state. For example, in embodiments where the NAGPB domain comprises a Cas9 domain, several nucleotides can be left unpaired during formation of the Cas9-gRNA-target DNA complex, resulting in
5 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 comprises all or a portion (e.g., a functional portion) of an apolipoprotein B mRNA editing complex (APOBEC) family
10 deaminase. 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,
15 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 (e.g., a functional portion) of an APOBEC1 deaminase. In some
20 embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC2 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of is an APOBEC3 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of an APOBEC3A deaminase. In some
25 embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3B deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3C deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3D deaminase. In some
30 embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3E deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3F deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC3G deaminase. In some
embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a

functional portion) of APOBEC3H deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of APOBEC4 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional portion) of activation-induced deaminase (AID).

5 In some embodiments a deaminase incorporated into a base editor comprises all or a portion (e.g., a functional 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, orangutan, alligator, pig, cow, dog, rat, or mouse. In some embodiments,

10 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 derived from an orangutan polypeptide (e.g., a *Pongo pygmaeus* (Orangutan) APOBEC). In some embodiments, the deaminase domain of the base editor is derived from a golden snub-nosed monkey polypeptide (e.g., a *Rhinopithecus roxellana* (golden snub-nosed monkey) APOBEC3F

15 (A3F)). In some embodiments, the deaminase domain of the base editor is derived from an American Alligator polypeptide (e.g., an *Alligator mississippiensis* (American alligator) APOBEC1). In some embodiments, the deaminase domain of the base editor is derived from a pig polypeptide (e.g., a *Sus scrofa* (pig) APOBEC3B). In some embodiments, the deaminase domain of the base editor is human APOBEC1. In some embodiments, the

20 deaminase domain of the base editor is pmCDA1.

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

25 localization sequence, without nuclear export signal, cytoplasmic localizing signal).

Some aspects of the present disclosure are based on the recognition that modulating the deaminase domain catalytic activity of any of the fusion proteins or complexes described herein, for example by making point mutations in the deaminase domain, affect the processivity of the fusion proteins (e.g., base editors) or complexes. For example, mutations

30 that reduce, but do not eliminate, the catalytic activity of a deaminase domain within a base editing fusion protein or complexes 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 comprises 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 comprises 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 or complexes 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises an APOBEC deaminase comprising a W90Y, R126E, and R132E mutation of rAPOBEC1, or one or more corresponding mutations in another APOBEC deaminase.

15 In some embodiments, an APOBEC deaminase incorporated into a base editor comprises 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 or complexes 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 comprises 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 (e.g., a functional portion) of an APOBEC1 deaminase.

In some embodiments, the fusion proteins or complexes of the invention comprise one or more cytidine deaminase domains. 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).

In some embodiments, the cytidine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%

identical to any one of the cytidine deaminase amino acid sequences set forth herein. It should be appreciated that cytidine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). Some embodiments provide a polynucleotide molecule encoding the cytidine deaminase nucleobase editor polypeptide of
5 any previous aspect or as delineated herein. In some embodiments, the polynucleotide is codon optimized.

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,
10 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 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
15 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 embodiments, a fusion protein of the invention comprises two or more nucleic acid
20 editing domains.

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
25 by reference.

Guide Polynucleotides

A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target
30 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.

CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems, correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA.

Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, and then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g., Jinek M., 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. *See e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti, J.J. et al., Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E. et al., Nature 471:602-607(2011); and "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M. et al, Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).*

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.

In an embodiment, a guide polynucleotide described herein can be RNA or DNA. In one embodiment, the guide polynucleotide is a gRNA. An RNA/Cas complex can assist in "guiding" a Cas protein to a target DNA. Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both

RNAs. However, single guide RNAs (“sgRNA”, or simply “gRNA”) can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.,* Jinek M. *et al.*, Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference.

5 In some embodiments, the guide polynucleotide is at least one single guide RNA (“sgRNA” or “gRNA”). In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via for example complementary base pairing (*e.g.*, a dual guide polynucleotide, dual gRNA). For example, a guide polynucleotide can comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR
10 RNA (tracrRNA) or can comprise one or more trans-activating CRISPR RNA (tracrRNA).

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.

15 A guide polynucleotide may include natural or non-natural (or unnatural) nucleotides (*e.g.*, peptide nucleic acid or nucleotide analogs). In some cases, the targeting region of a guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20
20 nucleotides in length.

In some embodiments, the base editor provided herein utilizes one or more guide polynucleotide (*e.g.*, multiple gRNA). In some embodiments, a single guide polynucleotide is utilized for different base editors described herein. For example, a single guide polynucleotide can be utilized for a cytidine base editor and an adenosine base editor.

25 In some embodiments, the methods described herein can utilize an engineered Cas protein. A guide RNA (gRNA) is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Exemplary gRNA scaffold sequences are provided in the sequence listing as SEQ ID NOs: 317-327. Thus, a skilled artisan can change the genomic target of
30 the Cas protein specificity is partially determined by how specific the gRNA targeting sequence is for the genomic target compared to the rest of the genome.

In other embodiments, a guide polynucleotide comprises both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide

polynucleotide can be a single guide RNA (sgRNA or gRNA). Herein the term guide polynucleotide sequence contemplates any single, dual or multi-molecule nucleic acid capable of interacting with and directing a base editor to a target polynucleotide sequence.

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

The guide polynucleotides can be synthesized chemically, synthesized enzymatically, or a combination thereof. For example, the gRNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the gRNA can be synthesized *in vitro* by operably linking DNA encoding the gRNA 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 gRNA comprises two separate molecules (*e.g.*, crRNA and

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

5 A guide polynucleotide may be expressed, for example, by a DNA that encodes the gRNA, *e.g.*, a DNA vector comprising a sequence encoding the gRNA. The gRNA may be encoded alone or together with an encoded base editor. 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 gRNA 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
10 sequence and a second vector containing the gRNA 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 gRNA). An RNA can be transcribed from a synthetic DNA molecule, *e.g.*, a gBlocks® gene fragment. A gRNA molecule can be transcribed *in vitro*.

15 A gRNA 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 does not form a secondary structure or bind a target site. A first region of each gRNA can also be different such that each gRNA guides a fusion protein or complex to a specific target site. Further,
20 second and third regions of each gRNA can be identical in all gRNAs.

A first region of a gRNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the gRNA can base pair with the target site. In some cases, a first region of a gRNA comprises from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or from
25 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 gRNA 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 gRNA can be or can be about 19, 20, or 21
30 nucleotides in length.

A gRNA or a guide polynucleotide can also comprise a second region that forms a secondary structure. For example, a secondary structure formed by a gRNA can comprise a stem (or hairpin) and a loop. A length of a loop and a stem can vary. For example, a loop can range from or from about 3 to 10 nucleotides in length, and a stem can range from or

from about 6 to 20 base pairs in length. A stem can comprise one or more bulges of 1 to 10 or about 10 nucleotides. The overall length of a second region can range from or from about 16 to 60 nucleotides in length. For example, a loop can be or can be about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

5 A gRNA 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 gRNA. Further, the length of a third region can vary. A third region can be more than or more than about 4 nucleotides in length. For example, the
10 length of a third region can range from or from about 5 to 60 nucleotides in length.

A gRNA 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. In some embodiments, a composition comprises multiple gRNAs that all target the same exon or multiple gRNAs that target different exons. An exon and/or an intron
15 of a gene can be targeted.

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

Methods for selecting, designing, and validating guide polynucleotides, *e.g.*, gRNAs and targeting sequences are described herein and known to those skilled in the art. For
25 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 single strand DNA 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
30 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.

- 5 As a non-limiting example, target DNA hybridizing sequences in crRNAs of a gRNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design is 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 gRNAs, *e.g.*, crRNAs, are 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.

A gRNA can then be introduced into a cell or embryo as an RNA molecule or a non-RNA nucleic acid molecule, *e.g.*, DNA molecule. In one embodiment, a DNA encoding a

gRNA is operably linked to promoter control sequence for expression of the gRNA 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 gRNA include, but are not limited to, px330 vectors and px333 vectors. In some cases, a plasmid vector (*e.g.*, px333 vector) comprises at least two gRNA-encoding DNA sequences. Further, a vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, GFP or antibiotic resistance genes such as puromycin), origins of replication, and the like. A DNA molecule encoding a gRNA can also be linear. A DNA molecule encoding a gRNA or a guide polynucleotide can also be circular.

In some embodiments, a reporter system is used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system comprises a reporter gene based assay where base editing activity leads to expression of the reporter gene. For example, a reporter system may include a reporter gene comprising a deactivated start codon, *e.g.*, a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art. Non-limiting examples of reporter genes include gene encoding green fluorescence protein (GFP), red fluorescence protein (RFP), luciferase, secreted alkaline phosphatase (SEAP), or any other gene whose expression are detectable and apparent to those skilled in the art. The reporter system can be used to test many different gRNAs, *e.g.*, in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase will target. sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific base editing protein, *e.g.*, a Cas9 deaminase fusion protein or complex. 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 comprises at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

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

Modified Polynucleotides

To enhance expression, stability, and/or genomic/base editing efficiency, and/or reduce possible toxicity, the base editor-coding sequence (*e.g.*, mRNA) and/or the guide polynucleotide (*e.g.*, gRNA) can be modified to include one or more modified nucleotides and/or chemical modifications, *e.g.* using pseudo-uridine, 5-Methyl-cytosine, 2'-*O*-methyl-3'-phosphonoacetate, 2'-*O*-methyl thioPACE (MSP), 2'-*O*-methyl-PACE (MP), 2'-fluoro RNA (2'-F-RNA), =constrained ethyl (S-cEt), 2'-*O*-methyl ('M'), 2'-*O*-methyl-3'-phosphorothioate ('MS'), 2'-*O*-methyl-3'-thiophosphonoacetate ('MSP'), 5-methoxyuridine, phosphorothioate, and N1-Methylpseudouridine. Chemically protected gRNAs can enhance stability and editing efficiency *in vivo* and *ex vivo*. Methods for using chemically modified mRNAs and guide RNAs are known in the art and described, for example, by Jiang et al., Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nat Commun 11, 1979 (2020). doi.org/10.1038/s41467-020-15892-8, Callum et al., N1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells, *Nucleic Acids Research*, Volume 48, Issue 6, 06 April 2020, Page e35, and Andries et al., Journal of Controlled Release, Volume 217, 10 November 2015, Pages 337-344, each of which is incorporated herein by reference in its entirety.

In a particular embodiment, the chemical modifications are 2'-*O*-methyl (2'-OMe) modifications. The modified guide RNAs may improve saCas9 efficacy and also specificity. The effect of an individual modification varies based on the position and combination of chemical modifications used as well as the inter- and intramolecular interactions with other modified nucleotides. By way of example, S-cEt has been used to improve oligonucleotide intramolecular folding.

In some embodiments, the guide polynucleotide comprises one or more modified nucleotides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some

embodiments, the guide polynucleotide comprises four modified nucleosides at the 5' end and four modified nucleosides at the 3' end of the guide. In some embodiments, the modified nucleoside comprises a 2' O-methyl or a phosphorothioate.

In some embodiments, the guide comprises at least about 50%-75% modified nucleotides. In some embodiments, the guide comprises at least about 85% or more modified nucleotides. In some embodiments, at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified. In some embodiments, at least about 3-5 contiguous nucleotides at each of the 5' and 3' termini of the gRNA are modified. In some embodiments, at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 50% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 100 of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, at least about 50% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, the guide comprises a variable length spacer. In some embodiments, the guide comprises a 20-40 nucleotide spacer. In some embodiments, the guide comprises a spacer comprising at least about 20-25 nucleotides or at least about 30-35 nucleotides. In some embodiments, the spacer comprises modified nucleotides. In some embodiments, the guide comprises two or more of the following:

- at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified;
- at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified;
- at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified;
- at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified;
- a variable length spacer; and
- a spacer comprising modified nucleotides.

In embodiments, the gRNA contains numerous modified nucleotides and/or chemical modifications ("heavy mods"). Such heavy mods can increase base editing ~2 fold *in vivo* or

in vitro. For such modifications, mN = 2'-OMe; Ns = phosphorothioate (PS), where "N" represents the any nucleotide, as would be understood by one having skill in the art. In some cases, a nucleotide (N) may contain two modifications, for example, both a 2'-OMe and a PS modification. For example, a nucleotide with a phosphorothioate and 2' OMe is denoted as

5 "mNs;" when there are two modifications next to each other, the notation is "mNsmNs.

In some embodiments of the modified gRNA, the gRNA comprises one or more chemical modifications selected from the group consisting of 2'-O-methyl (2'-OMe), phosphorothioate (PS), 2'-O-methyl thioPACE (MSP), 2'-O-methyl-PACE (MP), 2'-O-methyl thioPACE (MSP), 2'-fluoro RNA (2'-F-RNA), and constrained ethyl (S-cEt). In

10 embodiments, the gRNA comprises 2'-O-methyl or phosphorothioate modifications. In an embodiment, the gRNA comprises 2'-O-methyl and phosphorothioate modifications. In an embodiment, the modifications increase base editing by at least about 2 fold.

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

15 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

20 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.

25 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, 2'-O-methyl thioPACE (MSP), 2'-O-methyl-PACE (MP), and

30 constrained ethyl (S-cEt), 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'-
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
10 reactivity, base-pairing interactions, or any combination thereof.

A guide polynucleotide can be transferred into a cell by transfecting the cell with an isolated gRNA or a plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A gRNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery. A gRNA or a guide polynucleotide can be
15 isolated. For example, a gRNA can be transfected in the form of an isolated RNA into a cell or organism. A gRNA can be prepared by *in vitro* transcription using any *in vitro* transcription system known in the art. A gRNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a gRNA.

A modification can also be a phosphorothioate substitute. In some cases, a natural
20 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,
25 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
30 attack by endonucleases.

In some embodiments, the guide RNA is designed such that base editing results in disruption of a splice site (*i.e.*, a splice acceptor (SA) or a splice donor (SD)). In some embodiments, the guide RNA is designed such that the base editing results in a premature STOP codon.

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. 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, NGTT, 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 base editor provided herein can comprise a CRISPR protein-derived domain that is capable of binding a nucleotide sequence that contains a canonical or non-canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion (e.g., a functional portion) of CRISPR proteins that have different PAM specificities.

For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the “N” in “NGG” is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. A PAM can be CRISPR protein-specific and can be different between different base editors comprising different CRISPR protein-derived domains. A PAM can be 5' or 3' of a target sequence. A PAM can be upstream or downstream of a target sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. Often, a PAM is between 2-6 nucleotides in length.

In some embodiments, the PAM is an “NRN” PAM where the “N” in “NRN” is adenine (A), thymine (T), guanine (G), or cytosine (C), and the R is adenine (A) or guanine (G); or the PAM is an “NYN” PAM, wherein the “N” in NYN is adenine (A), thymine (T), guanine (G), or cytosine (C), and the Y is cytidine (C) or thymine (T), for example, as described in R.T. Walton *et al.*, 2020, *Science*, 10.1126/science.aba8853 (2020), the entire contents of which are incorporated herein by reference.

Several PAM variants are described in **Table 7** below.

Table 7. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
CpfI	5' (TTTV)
SpyMac	5'-NAA-3'

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 Cas9 variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed “MQKFRAER”) of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the Cas9 variant contains one or more amino acid substitutions selected from D1135V, G1218R, R1335Q, and T1337R (collectively termed VRQR) of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the Cas9 variant contains one or more amino acid substitutions selected from D1135V, G1218R, R1335E, and T1337R (collectively termed VRER) of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the Cas9 variant contains one or more amino acid substitutions selected from E782K, N968K, and R1015H (collectively termed KHH) of saCas9 (SEQ ID NO: 218). In some embodiments, the Cas9 variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219S, R1335E, and T1337R (collectively termed “MQKSER”) of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the Cas9 variant includes one or more

amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219S, R1335E, and T1337R (collectively termed “MQKSER”) of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9.

In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the Cas9 variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or 1219 of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the NGT PAM Cas9 variant is created through targeted mutations at one or more residues 1219, 1335, 1337, 1218 of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9. In some embodiments, the NGT PAM Cas9 variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335 of spCas9 (SEQ ID No: 197, or a corresponding mutation in another Cas9. In some embodiments, the NGT PAM Cas9 variant is selected from the set of targeted mutations provided in **Tables 8A** and **8B** below.

15

Table 8A: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218 of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9

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 8B: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335 of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9

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	N1286 Q	I1331 F			

5 In some embodiments, the NGT PAM Cas9 variant is selected from variant 5, 7, 28, 31, or 36 in **Table 8A** and **Table 8B**. In some embodiments, the variants have improved NGT PAM recognition.

10 In some embodiments, the NGT PAM Cas9 variants have mutations at residues 1219, 1335, 1337, and/or 1218. In some embodiments, the NGT PAM Cas9 variant is selected with mutations for improved recognition from the variants provided in **Table 9** below.

Table 9: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 1218 of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9

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 Cas9 variant is selected from the variants
 5 provided in **Table 10** below.

Table 10. NGT PAM variants, where the amino acid residue locations are referenced to of spCas9 (SEQ ID No: 197), or a corresponding mutation in another Cas9

	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

10 In some embodiments the NGTN Cas9 variant is variant 1. In some embodiments, the NGTN Cas9 variant is variant 2. In some embodiments, the NGTN Cas9 variant is variant 3. In some embodiments, the NGTN Cas9 variant is variant 4. In some embodiments, the NGTN variant is variant 5. In some embodiments, the NGTN Cas9 variant is variant 6.

In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9).
 15 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 some embodiments, the SpCas9 comprises a D9A mutation, or a corresponding mutation in
 20 any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d

domain, or the SpCas9n domain can bind to a nucleic acid sequence having an NGG, a NGA, or a NGCG PAM sequence.

In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135E, R1335Q, and T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135E, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a G1218X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 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 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 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, 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 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 *Neisseria meningitidis* (5'-NNNNGATT) 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, 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 exemplary SpCas9 proteins capable of binding a PAM sequence follow:

In some embodiments, engineered SpCas9 variants are capable of recognizing protospacer adjacent motif (PAM) sequences flanked by a 3' H (non-G PAM) (*see* **Tables 3A-3D**). In some embodiments, the SpCas9 variants recognize NRNH PAMs (where R is A or G and H is A, C or T). In some embodiments, the non-G PAM is NRRH, NRTH, or NRCH (*see e.g.*, Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the contents of which is incorporated herein by reference in its entirety).

In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some embodiments, the recombinant Cas9 domain is a SpyMacCas9 domain. In some embodiments, the SpyMacCas9 domain is a nuclease active SpyMacCas9, a nuclease inactive SpyMacCas9 (SpyMacCas9d), or a SpyMacCas9 nickase (SpyMacCas9n). In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a

nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpyMacCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a NAA PAM sequence.

The sequence of an exemplary Cas9 A homolog of Spy Cas9 in *Streptococcus macacae* with native 5'-NAAN-3' PAM specificity is known in the art and described, for example, by Chatterjee, *et al.*, "A Cas9 with PAM recognition for adenine dinucleotides", *Nature Communications*, vol. 11, article no. 2474 (2020), and is in the Sequence Listing as SEQ ID NO: 237.

In some embodiments, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations relative to a reference Cas9 sequence (e.g., spCas9 (SEQ ID No: 197)), or to a corresponding mutation in another Cas9, 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 embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations relative to a reference Cas9 sequence (e.g., spCas9 (SEQ ID No: 197)), or to a corresponding mutation in another Cas9, 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, 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 relative to a reference Cas9 sequence (e.g., spCas9 (SEQ ID No: 197)), or to a corresponding mutation in another Cas9, 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 embodiments, 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) relative to a reference Cas9 sequence (e.g., spCas9 (SEQ ID No: 197)), or to a corresponding mutation in another Cas9. Also, mutations other than alanine substitutions are suitable.

In some embodiments, a CRISPR protein-derived domain of a base editor comprises all or a portion (e.g., a functional portion) of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM sequence. Such sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, “Engineered CRISPR-Cas9 nucleases with altered PAM specificities” *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, “Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition” *Nature Biotechnology* 33, 1293-1298 (2015); 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 or Complexes Comprising a NapDNAbp and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins or complexes comprising a Cas9 domain or other nucleic acid programmable DNA binding protein (e.g., Cas12) and one or more cytidine deaminase or adenosine deaminase 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/or adenosine deaminases provided herein. The domains of the base editors disclosed herein can be arranged in any order.

In some embodiments, the fusion protein comprises the following domains A-C, A-D, or A-E:

NH₂-[A-B-C]-COOH;
 NH₂-[A-B-C-D]-COOH; or
 NH₂-[A-B-C-D-E]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,
 a cytidine deaminase domain or an active fragment thereof, and

wherein B or B and D, each comprises one or more domains having nucleic acid sequence specific binding activity.

In some embodiments, the fusion protein comprises the following structure:

NH₂-[A_n-B_o-C_n]-COOH;

5 NH₂-[A_n-B_o-C_n-D_o]-COOH; or

NH₂-[A_n-B_o-C_p-D_o-E_q]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,

a cytidine deaminase domain or an active fragment thereof, and

10 wherein n is an integer: 1, 2, 3, 4, or 5, wherein p is an integer: 0, 1, 2, 3, 4, or 5; wherein q is an integer 0, 1, 2, 3, 4, or 5; and wherein B or B and D each comprises a domain having nucleic acid sequence specific binding activity; and wherein o is an integer: 1, 2, 3, 4, or 5.

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

15 NH₂-[adenosine deaminase]-[Cas9 domain]-COOH;

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

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

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

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

20 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.

25 In some embodiments, any of the Cas12 domains or Cas12 proteins provided herein may be fused with any of the cytidine or adenosine deaminases provided herein. For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH₂-[adenosine deaminase]-[Cas12 domain]-COOH;

30 NH₂-[Cas12 domain]-[adenosine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas12 domain]-COOH;

NH₂-[Cas12 domain]-[cytidine deaminase]-COOH;

NH₂-[cytidine deaminase]-[Cas12 domain]-[adenosine deaminase]-COOH;

NH₂-[adenosine deaminase]-[Cas12 domain]-[cytidine deaminase]-COOH;

NH₂-[adenosine deaminase]-[cytidine deaminase]-[Cas12 domain]-COOH;
 NH₂-[cytidine deaminase]-[adenosine deaminase]-[Cas12 domain]-COOH;
 NH₂-[Cas12 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or
 NH₂-[Cas12 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

- 5 In some embodiments, the adenosine deaminase is a TadA*8. Exemplary fusion protein structures include the following:

NH₂-[TadA*8]-[Cas9 domain]-COOH;
 NH₂-[Cas9 domain]-[TadA*8]-COOH;
 NH₂-[TadA*8]-[Cas12 domain]-COOH; or

- 10 NH₂-[Cas12 domain]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein or complex comprises a TadA*8 and a cytidine deaminase and/or an adenosine 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-[TadA*8]-[Cas9/Cas12]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH;
 20 NH₂-[TadA*8]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or
 NH₂-[cytidine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*9 and a cytidine deaminase and/or an adenosine deaminase. Exemplary fusion protein structures include the following:

- 25 NH₂-[TadA*9]-[Cas9/Cas12]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH;
 NH₂-[TadA*9]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or
 NH₂-[cytidine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH.

- 30 In some embodiments, the fusion protein comprises a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises a cytidine deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas 12 polypeptide.

In some embodiments, the fusion proteins or complexes comprising a cytidine deaminase or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12 domain) do not include a linker sequence. In some embodiments, a linker is present between the cytidine or adenosine deaminase and the napDNAbp. In some embodiments, the “-” used in the general architecture above indicates the optional presence of a linker. In some embodiments, cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

It should be appreciated that the fusion proteins or complexes of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein or complex 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 or complexes. 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 or complex comprises one or more His tags.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/US2017/045381, PCT/US2019/044935, and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

Fusion Proteins or Complexes Comprising a Nuclear Localization Sequence (NLS)

In some embodiments, the fusion proteins or complexes 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, the NLS is fused to the N-terminus or the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus or N-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of

the Cas12 domain. In some embodiments, the NLS is fused to the N-terminus or C-terminus of the cytidine or adenosine deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises the amino acid sequence

PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 328), KRTADGSEFESPKKKRKV (SEQ ID NO: 190), KRPAATKKAGQAKKKK (SEQ ID NO: 191), KKTELQTTNAENKTKKL (SEQ ID NO: 192), KRGINDRNFWRGNGRKRTR (SEQ ID NO: 193), RKSGKIAAIVVKRPRKPKKKRKV (SEQ ID NO: 329), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 196).

In some embodiments, the fusion proteins or complexes comprising a cytidine or 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 or adenosine deaminase, Cas9 domain or NLS) are present. 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 below indicates the optional presence of a 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 herein.

In some embodiments, the general architecture of exemplary napDNAbp (*e.g.*, Cas9 or Cas12) fusion proteins with a cytidine or adenosine deaminase and a napDNAbp (*e.g.*, 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]-[napDNAbp domain]-COOH;
 NH₂-NLS [napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;

- NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS [napDNAbp domain]-[adenosine deaminase]-COOH;
 NH₂-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 5 NH₂-NLS-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-COOH;
 NH₂-NLS-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-COOH;
 NH₂-NLS-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-COOH;
 NH₂-NLS-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-COOH;
 10 NH₂-NLS-[napDNAbp domain]-[cytidine deaminase]-[adenosine deaminase]-COOH;
 NH₂-[cytidine deaminase]-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;
 NH₂-[adenosine deaminase]-[napDNAbp domain]-[cytidine deaminase]-NLS-COOH;
 NH₂-[adenosine deaminase] [cytidine deaminase]-[napDNAbp domain]-NLS-COOH;
 NH₂-[cytidine deaminase]-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;
 15 NH₂-[napDNAbp domain]-[adenosine deaminase]-[cytidine deaminase]-NLS-COOH; or
 NH₂-[napDNAbp 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. A bipartite NLS comprises two basic amino acid clusters, which are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while
 20 monopartite NLSs are not). The NLS of nucleoplasmin, KR [PAATKKAGQA] KKKK (SEQ ID NO: 191), 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 (SEQ ID NO: 328)

- 25 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 amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs at or near the carboxy-terminus, or any combination thereof (*e.g.*, one or more NLS at the amino-terminus and one or more NLS
 30 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 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.

5

Additional Domains

A base editor described herein can include any domain which helps to facilitate the nucleobase editing, modification or altering of a nucleobase of a polynucleotide. In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9), a nucleobase editing domain (*e.g.*, deaminase domain), and one or more additional domains. In some embodiments, the additional domain can facilitate enzymatic or catalytic functions of the base editor, binding functions of the base editor, or be inhibitors of cellular machinery (*e.g.*, enzymes) that could interfere with the desired base editing result. In some embodiments, a base editor comprises 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 comprises an uracil glycosylase inhibitor (UGI) domain. In some embodiments, 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 embodiments, uracil DNA glycosylase (UDG) can catalyze removal of U from DNA in cells, which can initiate base excision repair (BER), mostly resulting in reversion of the U:G pair to a C:G pair. In such embodiments, 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 or complex comprising a UGI domain and/or a uracil stabilizing protein (USP) domain.

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

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
 5 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
 10 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, substitutions in any domain does not change the length of the base editor.

15 Non-limiting examples of such base editors, where the length of all the domains is the same as the wild type domains, can include:

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-
 20 COOH;
 NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing
 25 domain]-[UGI]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-[UGI]-COOH;
 NH2-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-[UGI]-COOH;
 30 NH2-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-[UGI]-COOH;
 NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;
 NH2-[UGI]-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-COOH;

NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH; or

NH2-[UGI]-[nucleobase editing domain]-[APOBEC1]-[nucleobase editing domain]-COOH.

5 BASE EDITOR SYSTEM

Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a deaminase domain) for editing the nucleobase; and (2) a
 10 guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system is a cytidine base editor (CBE) or an adenosine base editor (ABE). In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA or RNA binding domain. In some embodiments, the nucleobase editing domain is a
 15 deaminase domain. In some embodiments, a deaminase domain can be a cytidine deaminase or an cytosine deaminase. In some embodiments, a deaminase domain can be an adenine deaminase or an adenosine deaminase. In some embodiments, the adenosine base editor can deaminate adenine in DNA. In some embodiments, the base editor is capable of deaminating a cytidine in DNA.

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

Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/US2017/045381 (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
 30 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:caao4774 (2017), the entire contents of which are hereby incorporated by reference.

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

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

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.

The components of a base editor system (*e.g.*, a deaminase domain, a guide RNA, and/or a polynucleotide programmable nucleotide binding domain) 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, optionally where the polynucleotide programmable nucleotide binding domain is complexed with a polynucleotide (*e.g.*, a guide RNA). In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or

linked to a deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can target a deaminase domain to a target nucleotide sequence by non-covalently interacting with or associating with the deaminase domain. For example, in some embodiments, the nucleobase editing component (*e.g.*, the deaminase component) comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with a corresponding heterologous portion, antigen, or domain that is part of a polynucleotide programmable nucleotide binding domain and/or a guide polynucleotide (*e.g.*, a guide RNA) complexed therewith. In some embodiments, the polynucleotide programmable nucleotide binding domain, and/or a guide polynucleotide (*e.g.*, a guide RNA) complexed therewith, comprises an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with a corresponding heterologous portion, antigen, or domain that is part of a nucleobase editing domain (*e.g.*, the deaminase component). 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 polypeptide linker. In some embodiments, the additional heterologous portion is capable of binding to a polynucleotide linker. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bcl-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain (*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibrin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1

domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (e.g., a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlg1-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In embodiments, an additional heterologous portion comprises a polynucleotide (*e.g.*, an RNA motif), such as an MS2 phage operator stem-loop (*e.g.* an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 operator stem-loop, an SfMu phage Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 380, 382, 384, 386-388, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 379, 381, 383, 385, or fragments thereof.

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) comprises 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 heterologous portion or segment (*e.g.*, a polynucleotide motif), or antigen 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. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bcl-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain (*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibrin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1 domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (*e.g.*, a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlg1-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In some embodiments, an additional heterologous portion comprises a polynucleotide (*e.g.*, an RNA motif), such as an MS2 phage operator stem-loop (*e.g.* an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 operator stem-loop, an SfMu phage Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID

NOs: 380, 382, 384, 386-388, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 379, 381, 383, 385, or fragments thereof.

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

capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the guide polynucleotide (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or

5 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

10 embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. An additional heterologous portion may be a protein domain. In some embodiments, an additional heterologous portion comprises a polypeptide, such as a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N (N22p), a 2G12 IgG homodimer domain, an ABI, an antibody (*e.g.* an antibody that binds a

15 component of the base editor system or a heterologous portion thereof) or fragment thereof (*e.g.* heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, an Fab2, miniantibodies, and/or ZIP antibodies), a barnase-barstar dimer domain, a Bel-xL domain, a Calcineurin A (CAN) domain, a Cardiac phospholamban

20 transmembrane pentamer domain, a collagen domain, a Com RNA binding protein domain (*e.g.* SfMu Com coat protein domain, and SfMu Com binding protein domain), a Cyclophilin-Fas fusion protein (CyP-Fas) domain, a Fab domain, an Fe domain, a fibrin foldon domain, an FK506 binding protein (FKBP) domain, an FKBP binding domain (FRB) domain of mTOR, a foldon domain, a fragment X domain, a GAI domain, a GID1

25 domain, a Glycophorin A transmembrane domain, a GyrB domain, a Halo tag, an HIV Gp41 trimerisation domain, an HPV45 oncoprotein E7 C-terminal dimer domain, a hydrophobic polypeptide, a K Homology (KH) domain, a Ku protein domain (*e.g.*, a Ku heterodimer), a leucine zipper, a LOV domain, a mitochondrial antiviral-signaling protein CARD filament domain, an MS2 coat protein domain (MCP), a non-natural RNA aptamer ligand that binds a

30 corresponding RNA motif/aptamer, a parathyroid hormone dimerization domain, a PP7 coat protein (PCP) domain, a PSD95-Dlg1-zo-1 (PDZ) domain, a PYL domain, a SNAP tag, a SpyCatcher moiety, a SpyTag moiety, a streptavidin domain, a streptavidin-binding protein domain, a streptavidin binding protein (SBP) domain, a telomerase Sm7 protein domain (*e.g.* Sm7 homoheptamer or a monomeric Sm-like protein), and/or fragments thereof. In

embodiments, an additional heterologous portion comprises a polynucleotide (e.g., an RNA motif), such as an MS2 phage operator stem-loop (e.g. an MS2, an MS2 C-5 mutant, or an MS2 F-5 mutant), a non-natural RNA motif, a PP7 operator stem-loop, an SfMu phage Com stem-loop, a steril alpha motif, a telomerase Ku binding motif, a telomerase Sm7 binding motif, and/or fragments thereof. Non-limiting examples of additional heterologous portions include polypeptides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 380, 382, 384, 386-388, or fragments thereof. Non-limiting examples of additional heterologous portions include polynucleotides with at least about 85% sequence identity to any one or more of SEQ ID NOs: 379, 381, 383, 385, or fragments thereof.

In some instances, components of the base editing system are associated with one another through the interaction of leucine zipper domains (e.g., SEQ ID NOs: 387 and 388). In some cases, components of the base editing system are associated with one another through polypeptide domains (e.g., FokI domains) that associate to form protein complexes containing about, at least about, or no more than about 1, 2 (i.e., dimerize), 3, 4, 5, 6, 7, 8, 9, 10 polypeptide domain units, optionally the polypeptide domains may include alterations that reduce or eliminate an activity thereof.

In some instances, components of the base editing system are associated with one another through the interaction of multimeric antibodies or fragments thereof (e.g., IgG, IgD, IgA, IgM, IgE, a heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), an immunoglobulin Fc region, a heavy chain domain 3 (CH3) of IgG or IgA, a heavy chain domain 4 (CH4) of IgM or IgE, an Fab, and an Fab2). In some instances, the antibodies are dimeric, trimeric, or tetrameric. In embodiments, the dimeric antibodies bind a polypeptide or polynucleotide component of the base editing system.

In some cases, components of the base editing system are associated with one another through the interaction of a polynucleotide-binding protein domain(s) with a polynucleotide(s). In some instances, components of the base editing system are associated with one another through the interaction of one or more polynucleotide-binding protein domains with polynucleotides that are self complementary and/or complementary to one another so that complementary binding of the polynucleotides to one another brings into association their respective bound polynucleotide-binding protein domain(s).

In some instances, components of the base editing system are associated with one another through the interaction of a polypeptide domain(s) with a small molecule(s) (e.g., chemical inducers of dimerization (CIDs), also known as “dimerizers”). Non-limiting examples of CIDs include those disclosed in Amara, *et al.*, “A versatile synthetic dimerizer

for the regulation of protein-protein interactions,” *PNAS*, 94:10618-10623 (1997); and Voß, *et al.* “Chemically induced dimerization: reversible and spatiotemporal control of protein function in cells,” *Current Opinion in Chemical Biology*, 28:194-201 (2015), the disclosures of each of which are incorporated herein by reference in their entireties for all purposes.

- 5 Non-limiting examples of polypeptides that can dimerize and their corresponding dimerizing agents are provided in **Table 10.1** below.

Table 10.1. Chemically induced dimerization systems.

Dimerizing Polypeptides		Dimerizing agent
FKBP	FKBP	FK1012
FKBP	Calcineurin A (CNA)	FK506
FKBP	CyP-Fas	FKCsA
FKBP	FRB (FKBP-rapamycin-binding) domain of mTOR	Rapamycin
GyrB	GyrB	Coumermycin
GAI	GID1 (gibberellin insensitive dwarf 1)	Gibberellin
ABI	PYL	Absciscic acid
ABI	PYRMandi	Mandipropamid
SNAP-tag	HaloTag	HaXS
eDHFR	HaloTag	TMP-HTag
Bcl-xL	Fab (AZ1)	ABT-737

- 10 In embodiments, the additional heterologous portion is part of a guide RNA molecule. In some instances, the additional heterologous portion contains or is an RNA motif. The RNA motif may be positioned at the 5' or 3' end of the guide RNA molecule or various positions of a guide RNA molecule. In embodiments, the RNA motif is positioned within the guide RNA to reduce steric hindrance, optionally where such hindrance is associated with
- 15 other bulky loops of an RNA scaffold. In some instances, it is advantageous to link the RNA motif is linked to other portions of the guide RNA by way of a linker, where the linker can be about, at least about, or no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides in length. Optionally, the linker contains a GC-rich nucleotide sequence. The guide RNA can contain 1, 2, 3, 4, 5, or more copies of the RNA motif, optionally where they are positioned
- 20 consecutively, and/or optionally where they are each separated from one another by a linker(s). The RNA motif may include any one or more of the polynucleotide modifications described herein. Non-limiting examples of suitable modifications to the RNA motif include 2' deoxy-2-aminopurine, 2'-ribose-2-aminopurine, phosphorothioate mods, 2'-O-methyl mods,

2'-Fluoro mods and LNA mods. Advantageously, the modifications help to increase stability and promote stronger bonds/folding structure of a hairpin(s) formed by the RNA motif.

In some embodiments, the RNA motif is modified to include an extension. In
embodiments, the extension contains about, at least about, or no more than about 2, 3, 4, 5,
5 10, 15, 20, or 25 nucleotides. In some instances, the extension results in an alteration in the
length of a stem formed by the RNA motif (e.g., a lengthening or a shortening). It can be
advantageous for a stem formed by the RNA motif to be about, at least about, or no more
than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100
nucleotides in length. In various embodiments, the extension increases flexibility of the RNA
10 motif and/or increases binding with a corresponding RNA motif.

In some embodiments, the base editor inhibits base excision repair (BER) of the
edited strand. In some embodiments, the base editor protects or binds the non-edited strand.
In some embodiments, the base editor comprises UGI activity or USP activity. In some
embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In
15 some embodiments, the base editor comprises nickase activity. In some embodiments, the
intended edit of base pair is upstream of a PAM site. In some embodiments, the intended edit
of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides
upstream of the PAM site. In some embodiments, the intended edit of base-pair is
downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4,
20 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.
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
25 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 base editing fusion proteins or complexes 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 embodiments, a target can be within
30 a 4 base region. In some embodiments, 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:eao4774 (2017), the entire contents of which are hereby incorporated by reference.

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. 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.

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

can include transposase activity, integrase activity, recombinase activity, ligase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ribosylation activity, deribosylation activity, myristoylation activity, demyristoylation activity, polymerase activity, helicase activity, or nuclease activity, SUMOylation activity, deSUMOylation activity, or any combination of the above. In some embodiments, the Cas9 protein is fused to a histone demethylase, a transcriptional activator or a deaminase.

Further suitable fusion partners include, but are not limited to boundary elements (e.g., CTCF), proteins and fragments thereof that provide periphery recruitment (e.g., Lamin A, Lamin B, etc.), and protein docking elements (e.g., FKBP/FRB, Pill/Ablyl, etc.).

A domain may be detected or labeled with an epitope tag, a reporter protein, other binding domains. 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, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

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 or complexes. 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, FAsH 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 or complex comprises one or more His tags.

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 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 (SEQ ID NO: 330)-XTEN-(SGGS)₂ (SEQ ID NO: 330)) 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 I156F).

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 **Table 11** below. 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 **Table 11** below. 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 11** below.

Table 11. 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

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	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 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 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).

- 5 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 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
- 10 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 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
- 15 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 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-
- 20 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 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 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,
- 30 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), 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 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 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, 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 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-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 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 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 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, 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 12** below.

Table 12: Adenosine Base Editor 8 (ABE8) 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)

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, 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).

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 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 13** 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 13**, 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 13: Additional Adenosine Base Editor 8 Variants. In the table, “monomer” indicates an ABE comprising a single TadA*7.10 comprising the indicated alterations and “heterodimer” indicates an ABE comprising a TadA*7.10 comprising the indicated alterations fused to an E. coli TadA adenosine deaminase.

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. pyogenes* 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. pyogenes* 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. pyogenes* 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. pyogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in **Table 14** below.

Table 14. Genotypes of ABEs

	2	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
	3																					
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

As shown in Table 15 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 15** below.

Table 15. Residue Identity in Evolved TadA

	2 3	3 6	4 8	5 1	7 6	8 2	8 4	10 6	10 8	12 3	14 6	14 7	15 2	15 4	15 5	15 6	15 7	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:

>ABE8.1_Y147T_CP5_NGC PAM_monomer

5 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALR
 QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
 RVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD**SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVAK
 10 **VEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR**
MLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEF
SKRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTK
EVL DATLIHQSI TGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
 15 **YLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD**

STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENPINASGVDA
 KAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALLSLGLTPNFKSNFDAEDAKLQLSKDITYD
 DDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
 ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
 5 KQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
 MTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVK
 YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDREFNASLG
 TYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRY
 TGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSL
 10 HEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
 EEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
 DDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR
 EINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEF
 15 ESPKKKRKV (SEQ ID NO: 331)

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, and the underlined sequence denotes a bipartite nuclear localization sequence. Other ABE8 sequences are provided in the attached sequence listing (SEQ ID NOs: 332-354).

20 In some embodiments, the base editor is a ninth generation ABE (ABE9). In some
 embodiments, the ABE9 contains a TadA*9 variant. ABE9 base editors include an adenosine
 deaminase variant comprising an amino acid sequence, which contains alterations relative to an
 ABE 7*10 reference sequence, as described herein. Exemplary ABE9 variants are listed in
Table 16. Details of ABE9 base editors are described in International PCT Application
 25 No. PCT/US2020/049975, which is incorporated herein by reference for its entirety.

30 **Table 16. Adenosine Base Editor 9 (ABE9) Variants.** In the table, “monomer” indicates an
 ABE comprising a single TadA*7.10 comprising the indicated alterations and
 “heterodimer” indicates an ABE comprising a TadA*7.10 comprising the indicated
 alterations fused to an E. coli TadA adenosine deaminase.

ABE9 Description	Alterations
ABE9.1 _monomer	E25F, V82S, Y123H, T133K, Y147R, Q154R

ABE9.2 monomer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.3 monomer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.4 monomer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.5 monomer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.6 monomer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.7 monomer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.8 monomer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.9 monomer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.10 monomer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.11 monomer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.12 monomer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.13 monomer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.14 monomer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.15 heterodimer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.16 heterodimer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.17 heterodimer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.18 heterodimer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.19 heterodimer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.2 heterodimer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.21 heterodimer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.22 heterodimer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.23 heterodimer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.24 heterodimer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.25 heterodimer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.26 heterodimer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.27 heterodimer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.28 heterodimer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.29 monomer	E25F I76Y V82S Y123H Y147R Q154R
ABE9.30 monomer	I76Y V82T Y123H Y147R Q154R
ABE9.31 monomer	N38G I76Y V82S Y123H Y147R Q154R
ABE9.32 monomer	N38G I76Y V82T Y123H Y147R Q154R
ABE9.33 monomer	R23H I76Y V82S Y123H Y147R Q154R
ABE9.34 monomer	P54C I76Y V82S Y123H Y147R Q154R
ABE9.35 monomer	R21N I76Y V82S Y123H Y147R Q154R
ABE9.36 monomer	I76Y V82S Y123H D138M Y147R Q154R
ABE9.37 monomer	Y72S I76Y V82S Y123H Y147R Q154R
ABE9.38 heterodimer	E25F I76Y V82S Y123H Y147R Q154R
ABE9.39 heterodimer	I76Y V82T Y123H Y147R Q154R
ABE9.40 heterodimer	N38G I76Y V82S Y123H Y147R Q154R
ABE9.41 heterodimer	N38G I76Y V82T Y123H Y147R Q154R
ABE9.42 heterodimer	R23H I76Y V82S Y123H Y147R Q154R
ABE9.43 heterodimer	P54C I76Y V82S Y123H Y147R Q154R
ABE9.44 heterodimer	R21N I76Y V82S Y123H Y147R Q154R
ABE9.45 heterodimer	I76Y V82S Y123H D138M Y147R Q154R
ABE9.46 heterodimer	Y72S I76Y V82S Y123H Y147R Q154R

ABE9.47_monomer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.48_monomer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.49_monomer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.50_monomer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.51_monomer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.52_monomer	M70V,Q71M,N72K,V82S, Y123H, Y147R, Q154R
ABE9.53_heterodimer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.54_heterodimer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.55_heterodimer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.56_heterodimer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.57_heterodimer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.58_heterodimer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R

In some embodiments, the base editor includes an adenosine deaminase variant comprising an amino acid sequence, which contains alterations relative to an ABE 7*10 reference sequence, as described herein. The term “monomer” as used in **Table 16.1** refers to a monomeric form of TadA*7.10 comprising the alterations described. The term “heterodimer” as used in **Table 16.1** refers to the specified wild-type *E. coli* TadA adenosine deaminase fused to a TadA*7.10 comprising the alterations as described.

Table 16.1. Adenosine Deaminase Base Editor Variants

ABE	Adenosine Deaminase	Adenosine Deaminase Description
ABE-605m	MSP605	monomer_TadA*7.10 + V82G + Y147T + Q154S
ABE-680m	MSP680	monomer_TadA*7.10 + I76Y + V82G + Y147T + Q154S
ABE-823m	MSP823	monomer_TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K
ABE-824m	MSP824	monomer_TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N
ABE-825m	MSP825	monomer_TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N
ABE-827m	MSP827	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K
ABE-828m	MSP828	monomer_TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N
ABE-829m	MSP829	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N

ABE-605d	MSP605	heterodimer_(WT)+(TadA*7.10 + V82G + Y147T + Q154S)
ABE-680d	MSP680	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147T + Q154S)
ABE-823d	MSP823	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K)
ABE-824d	MSP824	heterodimer_(WT)+(TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N)
ABE-825d	MSP825	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N)
ABE-827d	MSP827	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K)
ABE-828d	MSP828	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N)
ABE-829d	MSP829	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N)

In some embodiments, the base editor comprises a domain comprising all or a portion (e.g., a functional portion) of a uracil glycosylase inhibitor (UGI) or a uracil stabilizing protein (USP) domain. In some embodiments, the base editor comprises a domain comprising all or a portion (e.g., a functional portion) of a nucleic acid polymerase. In some embodiments, a base editor comprises as a domain all or a portion (e.g., a functional portion) of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion (e.g., a functional 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 embodiments, 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). 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 comprises multiple domains. For example, the base editor comprising a polynucleotide programmable nucleotide binding domain

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

10 Different domains (e.g., adjacent domains) of the base editor disclosed herein can be connected to each other with or without the use of one or more linker domains (e.g., an XTEN linker domain). In some 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

15 (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, 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.,

20 polyethylene, polyethylene glycol, polyamide, polyester, etc.). In certain embodiments, a linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, a linker comprises an aminoalkanoic acid (e.g., glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.). In some embodiments, a linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain

25 embodiments, a linker is based on a carbocyclic moiety (e.g., cyclopentane, cyclohexane). In other embodiments, a linker comprises a polyethylene glycol moiety (PEG). In certain embodiments, a linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. A linker can include functionalized moieties to facilitate attachment of a nucleophile (e.g., thiol, amino) from the peptide to the linker. Any electrophile

30 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, *etc.*).

5

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
10 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
15 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
20 (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
25 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.

Typically, a linker is positioned between, or flanked by, two groups, molecules, or other
30 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
 5 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, any of the fusion proteins provided herein, comprise a cytidine or
 10 adenosine deaminase and a Cas9 domain that are fused to each other via a linker. Various linker lengths and flexibilities between the cytidine or adenosine deaminase and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers of the form (GGGS)_n (SEQ ID NO: 246), (GGGGS)_n (SEQ ID NO: 247), and (G)_n to more rigid linkers of the form (EAAAK)_n (SEQ ID NO: 248), (SGGS)_n (SEQ ID NO: 355), SGSETPGTSESATPES (SEQ ID NO: 249) (*see, e.g.*,
 15 Guilinger JP, *et al.* 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 cytidine or adenosine deaminase 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
 20 (GGS)_n motif, wherein *n* is 1, 3, or 7. In some embodiments, cytidine deaminase or adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 249), which can also be referred to as the XTEN linker.

In some embodiments, the domains of the base editor are fused via a linker that
 25 comprises the amino acid sequence of:
 SGGSSGSETPGTSESATPESGGGS (SEQ ID NO: 356),
 SGGSSGGSSGSETPGTSESATPESGGSSGGGS (SEQ ID NO: 357), or
 GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEG
 SAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 358).

30 In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 249), which may also be referred to

as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGs. In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 359). In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESGGSSGGSSGGSSGGSS (SEQ ID NO: 360). In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESGGSSGGSSGGSSGGSSGSETPGTSESATPESGGSSGGSS (SEQ ID NO: 361). In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGSEPATs (SEQ ID NO: 362).

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 (SEQ ID NO: 363), PAPAPA (SEQ ID NO: 364), PAPAPAP (SEQ ID NO: 365), PAPAPAPA (SEQ ID NO: 366), P(AP)₄ (SEQ ID NO: 367), P(AP)₇ (SEQ ID NO: 368), P(AP)₁₀ (SEQ ID NO: 369) (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.

In another embodiment, the base editor system comprises a component (protein) that interacts non-covalently with a deaminase (DNA deaminase), *e.g.*, an adenosine or a cytidine deaminase, and transiently attracts the adenosine or cytidine deaminase to the target nucleobase in a target polynucleotide sequence for specific editing, with minimal or reduced bystander or target-adjacent effects. Such a non-covalent system and method involving deaminase-interacting proteins serves to attract a DNA deaminase to a particular genomic target nucleobase and decouples the events of on-target and target-adjacent editing, thus enhancing the achievement of more precise single base substitution mutations. In an embodiment, the deaminase-interacting protein binds to the deaminase (*e.g.*, adenosine deaminase or cytidine deaminase) without blocking or interfering with the active (catalytic) site of the deaminase from engaging the target nucleobase (*e.g.*, adenosine or cytidine, respectively). Such a system, termed “MagnEdit,” involves interacting proteins tethered to a Cas9 and gRNA complex and can attract a co-

expressed adenosine or cytidine deaminase (either exogenous or endogenous) to edit a specific genomic target site, and is described in McCann, J. *et al.*, 2020, “MagnEdit – interacting factors that recruit DNA-editing enzymes to single base targets,” Life-Science-Alliance, Vol. 3, No. 4 (e201900606), (doi 10.26508/Isa.201900606), the contents of which are incorporated by
5 reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

In another embodiment, a system called “SunTag,” involves non-covalently interacting components used for recruiting protein (*e.g.*, adenosine deaminase or cytidine deaminase) components, or multiple copies thereof, of base editors to polynucleotide target sites to achieve
10 base editing at the site with reduced adjacent target editing, for example, as described in Tanenbaum, M.E. *et al.*, “A protein tagging system for signal amplification in gene expression and fluorescence imaging,” Cell. 2014 October 23; 159(3): 635–646.doi:10.1016/j.cell.2014.09.039; and in Huang, Y.-H. *et al.*, 2017, “DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A,” Genome Biol 18: 176. doi:10.1186/s13059-
15 017-1306-z, the contents of each of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

Nucleic Acid Programmable DNA Binding Proteins with Guide RNAs

20 Provided herein are compositions and methods for base editing in cells. Further provided herein are compositions comprising a guide polynucleic acid sequence, *e.g.* a guide RNA sequence, or a combination of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more guide RNAs as provided herein. In some embodiments, a composition for base editing as provided herein further comprises a polynucleotide that encodes a base editor, *e.g.* a C-base
25 editor or an A-base editor. For example, a composition for base editing may comprise a mRNA sequence encoding a BE, a BE4, an ABE, and a combination of one or more guide RNAs as provided. A composition for base editing may comprise a base editor polypeptide and a combination of one or more of any guide RNAs provided herein. Such a composition may be used to effect base editing in a cell through different delivery approaches, for example,
30 electroporation, nucleofection, viral transduction or transfection. In some embodiments, the

composition for base editing comprises an mRNA sequence that encodes a base editor and a combination of one or more guide RNA sequences provided herein for electroporation.

Some aspects of this disclosure provide systems comprising any of the fusion proteins or complexes provided herein, and a guide RNA bound to a nucleic acid programmable DNA binding protein (napDNAbp) domain (*e.g.*, a Cas9 (*e.g.*, a dCas9, a nuclease active Cas9, or a Cas9 nickase) or Cas12) of the fusion protein or complex. These complexes are also termed ribonucleoproteins (RNPs). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is an RNA 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 7** 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 or complexes 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. In some embodiments, the 3' end of the target

sequence is immediately adjacent to an *e.g.*, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, or YTN PAM site.

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 differ, *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 or complexes disclosed herein, to a target site, *e.g.*, a site comprising a mutation or other site of interest to be edited, it is typically necessary to co-express the fusion protein or complex together with a guide RNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for napDNAbp (*e.g.*, Cas9 or Cas12) binding, and a guide sequence, which confers sequence specificity to the napDNAbp:nucleic acid editing enzyme/domain fusion protein or complex. 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 napDNAbp:nucleic acid editing enzyme/domain fusion proteins or complexes to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins or complexes to specific target sequences are provided herein.

Distinct portions of sgRNA are predicted to form various features that interact with Cas9 (*e.g.*, SpyCas9) and/or the DNA target. Six conserved modules have been identified within native crRNA:tracrRNA duplexes and single guide RNAs (sgRNAs) that direct Cas9 endonuclease activity (see Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339). The six modules include the spacer

responsible for DNA targeting, the upper stem, bulge, lower stem formed by the CRISPR repeat:tracrRNA duplex, the nexus, and hairpins from the 3' end of the tracrRNA. The upper and lower stems interact with Cas9 mainly through sequence-independent interactions with the phosphate backbone. In some embodiments, the upper stem is dispensable. In some
5 embodiments, the conserved uracil nucleotide sequence at the base of the lower stem is dispensable. The bulge participates in specific side-chain interactions with the Rec1 domain of Cas9. The nucleobase of U44 interacts with the side chains of Tyr 325 and His 328, while G43 interacts with Tyr 329. The nexus forms the core of the sgRNA:Cas9 interactions and lies at the intersection between the sgRNA and both Cas9 and the target DNA. The nucleobases of A51
10 and A52 interact with the side chain of Phe 1105; U56 interacts with Arg 457 and Asn 459; the nucleobase of U59 inserts into a hydrophobic pocket defined by side chains of Arg 74, Asn 77, Pro 475, Leu 455, Phe 446, and Ile 448; C60 interacts with Leu 455, Ala 456, and Asn 459, and C61 interacts with the side chain of Arg 70, which in turn interacts with C15. In some
15 embodiments, one or more of these mutations are made in the bulge and/or the nexus of a sgRNA for a Cas9 (*e.g.*, spyCas9) to optimize sgRNA:Cas9 interactions.

Moreover, the tracrRNA nexus and hairpins are critical for Cas9 pairing and can be swapped to cross orthogonality barriers separating disparate Cas9 proteins, which is instrumental for further harnessing of orthogonal Cas9 proteins. In some embodiments, the nexus and hairpins are swapped to target orthogonal Cas9 proteins. In some embodiments, a sgRNA is
20 dispensed of the upper stem, hairpin 1, and/or the sequence flexibility of the lower stem to design a guide RNA that is more compact and conformationally stable. In some embodiments, the modules are modified to optimize multiplex editing using a single Cas9 with various chimeric guides or by concurrently using orthogonal systems with different combinations of chimeric sgRNAs. Details regarding guide functional modules and methods thereof are described, for
25 example, in Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339, the contents of which is incorporated by reference herein in its entirety.

The domains of the base editor disclosed herein can be arranged in any order. Non-limiting examples of a base editor comprising a fusion protein comprising *e.g.*, a polynucleotide-
30 programmable nucleotide-binding domain (*e.g.*, Cas9 or Cas12) and a deaminase domain (*e.g.*, cytidine or adenosine deaminase) can be arranged as follows:

- NH2-[nucleobase editing domain]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-Linker2-[UGI]-COOH;
 NH2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;
 5 NH2-[adenosine deaminase]-Linker1-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-COOH;
 NH2-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 NH2-[deaminase]-[inosine BER inhibitor]-[nucleobase editing domain]-COOH;
 NH2-[inosine BER inhibitor]-[deaminase]-[nucleobase editing domain]-COOH;
 10 NH2-[nucleobase editing domain]-[deaminase]-[inosine BER inhibitor]-COOH;
 NH2-[nucleobase editing domain]-[inosine BER inhibitor]-[deaminase]-COOH;
 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;
 15 NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH;
 20 NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 25 NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;
 30 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

NH₂-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH; or

NH₂-[inosine BER inhibitor]NH₂-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH.

5 In some embodiments, the base editing fusion proteins or complexes 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 embodiments, a target can be within a 4-base region. In some embodiments, 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
10 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
15 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,
20 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
25 napDNAbp domain. In some embodiments, an NLS of the base editor is localized C-terminal to a napDNAbp domain.

Non-limiting examples of protein domains which can be included in the fusion protein or complex include a deaminase domain (*e.g.*, adenosine deaminase or cytidine deaminase), a uracil glycosylase inhibitor (UGI) domain, epitope tags, reporter gene sequences, and/or protein
30 domains having one or more of the activities described herein.

A domain may be detected or labeled with an epitope tag, a reporter protein, other binding domains. 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),
5 horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose
10 binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

Methods of Using Fusion Proteins or Complexes Comprising a Cytidine or Adenosine Deaminase and a Cas9 Domain

15 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 or complexes provided herein, and with at least one guide RNA described herein.

In some embodiments, a fusion protein or complex of the invention is used for editing a
20 target gene of interest. In particular, a cytidine deaminase or adenosine deaminase nucleobase editor described herein is capable of making multiple mutations within a target sequence. These mutations may affect the function of the target. For example, when a cytidine deaminase or adenosine deaminase nucleobase editor is used to target a regulatory region the function of the regulatory region is altered and the expression of the downstream protein is reduced or
25 eliminated.

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
30 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 or complexes comprising a Cas9 domain and a cytidine or adenosine deaminase, as 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 or complex together with a guide RNA, *e.g.*, an sgRNA. 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 or complex.

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 or complexes to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins or complexes to specific target sequences are provided herein.

20

Multiplex Editing

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 or polynucleotide sequences. In some embodiments, the plurality of nucleobase pairs is located in the same gene or in one or more genes, wherein at least one gene is located in a different locus. In some embodiments, the multiplex editing comprises one or more guide polynucleotides. In some embodiments, the multiplex editing comprises one or more base editor systems. In some embodiments, the multiplex editing comprises one or more base editor systems with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the multiplex editing comprises one or more guide polynucleotides with a single base editor system. In some embodiments, the multiplex editing comprises at least one guide polynucleotide that does or does not require a

30

PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing comprises 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

5 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 combination of methods using any 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.

10 In some embodiments, the plurality of nucleobase pairs are in one more genes. In 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 at least one protein non-coding region, or in at least one protein
15 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 comprises one or more base editor systems. In some embodiments, the base editor system comprises one or more base editor systems in conjunction with a single guide polynucleotide or a plurality of guide polynucleotides.
20 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 or with at least one guide polynucleotide that requires a PAM sequence to target binding to a target polynucleotide sequence, or with a mix of at least one guide
25 polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that does require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editors provided herein. It should also be
30 appreciated that the editing can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes comprises one of ABE7, ABE8, and/or ABE9 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE8 base editor variants described herein has higher multiplex editing efficiency compared to the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one 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% higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one 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 4.0 fold, at least 4.5 fold, at least 5.0 fold, at least 5.5 fold, or at least 6.0 fold higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, use of a base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes described herein does not comprise a risk or occurrence of chromosomal translocations.

30

Base Editor Efficiency

In some embodiments, the purpose of the methods provided herein is to alter a gene and/or gene product via gene editing. The nucleobase editing proteins provided herein can be used for gene editing-based human therapeutics in vitro or in vivo. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, *e.g.*, the fusion proteins or complexes 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 edit a nucleotide from A to G or C to T.

Advantageously, base editing systems as provided herein 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 as CRISPR may do. In some embodiments, the present disclosure provides base editors that efficiently generate an intended mutation, such as a STOP codon, 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.*, adenosine base editor or cytidine base editor) bound to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation. In some embodiments, the intended mutation is in a gene associated with a target antigen associated with a disease or disorder, *e.g.*, an autoimmune disease. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation (*e.g.*, SNP) in a gene associated with a target antigen associated with a disease or disorder, *e.g.* an autoimmune disease. 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 (*e.g.*, regulatory region or element). In some embodiments, the intended mutation is a cytosine (C) to thymine (T) point mutation (*e.g.*, SNP) in a gene associated with a target antigen associated with a disease or disorder, *e.g.*, an autoimmune disease. 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 (*e.g.*, regulatory region or element). 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.

The base editors of the invention advantageously modify a specific nucleotide base encoding a protein without generating a significant proportion of indels. An “indel”, 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
5 some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.*, mutate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.*, mutate or methylate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid.
10 In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, methylations) versus indels. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, mutations) versus indels.

In some embodiments, the base editors provided herein are capable of generating a ratio
15 of intended mutations to indels (*i.e.*, intended point mutations:unintended point mutations) that is greater than 1:1. In some embodiments, the base editors provided herein are capable of generating a ratio of intended mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1,
20 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 may be determined using any suitable method.

In some embodiments, the base editors provided herein can limit formation of indels in a
25 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 can limit 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 than 4%, less than 4.5%, less than 5%, less than 6%, less than
30 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, a number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) to a base editor.

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 considerable number of unintended 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 generate the intended mutation. In some embodiments, the intended mutation is a 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, 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). 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 mutations:unintended 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 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. It should be appreciated that the characteristics of the base editors described herein may be applied to any of the fusion proteins or complexes, or methods of using the fusion proteins or complexes provided herein.

Base editing is often referred to as a “modification”, such as, a genetic modification, a gene modification and modification of the nucleic acid sequence and is clearly understandable based on the context that the modification is a base editing modification. A base editing modification is therefore a modification at the nucleotide base level, for example as a result of

the deaminase activity discussed throughout the disclosure, which then results in a change in the gene sequence, and may affect the gene product. In essence therefore, the gene editing modification described herein may result in a modification of the gene, structurally and/or functionally, wherein the expression of the gene product may be modified, for example, the expression of the gene is knocked out; or conversely, enhanced, or, in some circumstances, the gene function or activity may be modified. Using the methods disclosed herein, in some embodiments a base editing efficiency may be determined as the knockdown efficiency of the gene in which the base editing is performed, wherein the base editing is intended to knockdown the expression of the gene. A knockdown level may be validated quantitatively by determining the expression level by any detection assay, such as assay for protein expression level, for example, by flow cytometry; assay for detecting RNA expression such as quantitative RT-PCR, northern blot analysis, or any other suitable assay such as pyrosequencing; and may be validated qualitatively by nucleotide sequencing reactions. In some embodiments a base editing efficiency may be determined by sequencing the genome of the cells on which base editing has been performed to detect alterations in a target sequence as described herein.

In some embodiments, the modification, *e.g.*, single base edit results in at least 10% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 10% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 20% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 30% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 40% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 50% reduction of the gene targeted expression. In some embodiments, the base editing efficiency may result in at least 60% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 70% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 80% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 90% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 91% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 92% reduction of the targeted gene expression. In some embodiments, the base editing efficiency

may result in at least 93% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 94% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 95% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 96% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 97% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 98% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in at least 99% reduction of the targeted gene expression. In some embodiments, the base editing efficiency may result in knockout (100% knockdown of the gene expression) of the gene that is targeted.

In some embodiments the base editing produces an alteration in a target gene that may reduce expression of the target gene by no more than 5%. In some embodiments the base editing produces an alteration in a target gene that may reduce expression of the target gene by no more than 10%. In some embodiments the base editing produces an alteration in a target gene that may reduce expression of the target gene by no more than 20%. In some embodiments the base editing produces an alteration that may reduce expression of a target gene by no more than 30%. In some embodiments the base editing produces an alteration that may reduce expression of a target gene by no more than 40%. In some embodiments the base editing produces an alteration that may reduce expression of a target gene by no more than 50%. In some embodiments a target gene encodes a gene product, e.g., a protein, that has at least two activities. In some embodiments an alteration of the target gene reduces at least one undesired activity of an encoded gene product while preserving sufficient expression so that the encoded gene product can effectively perform one or more other activities.

In some embodiments, any of the 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 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.

In some embodiments, targeted modifications, *e.g.*, single base editing, are used simultaneously to target at least 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 different endogenous sequences for base editing with different guide RNAs. In some embodiments, targeted modifications, *e.g.* single base editing, are used to sequentially target at least 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 different endogenous gene sequences for base editing with different guide RNAs.

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 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 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% indel formation in the target polynucleotide sequence. In some embodiments, any of the 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 the 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 the 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 the 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 the 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.

10 In some embodiments, any of the 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 the 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.

25 The invention 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 30 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

5 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 system comprising an ABE7 base editor, *e.g.*,

10 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, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least

15 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 editor systems provided herein result in less than 70%, less than 65%, less than 60%, less than 55%, 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%,

20 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

25 0.01% bystander editing of one or more nucleotides (*e.g.*, an off-target nucleotide).

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-specific editing or guide independent editing of a target base (*e.g.*, A or C) in an unintended or

30 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 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%,
5 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 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
10 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 compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at
15 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% base editing efficiency. In some embodiments, the base editing efficiency may be measured by calculating the percentage of edited nucleobases
20 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 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
25 edited nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has 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%,
30 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
5 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 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
10 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
15 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
20 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
25 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
30 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
5 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
10 390%, 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,
15 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
20 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-
25 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
30 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
5 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
10 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
15 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,
20 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.

In some embodiments, any of the 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%,
25 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
30 0.01% off-target editing in the target polynucleotide sequence.

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
5 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.

In some embodiments, a single gene delivery event (*e.g.*, by transduction, transfection, electroporation or any other method) can be used to target base editing of 5 sequences within a
10 cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 6 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 7 sequences within a cell's genome. In some embodiments, a single electroporation event can be used to target base editing of 8 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target
15 base editing of 9 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 10 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 20 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 30 sequences within a cell's genome. In some embodiments, a
20 single gene delivery event can be used to target base editing of 40 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 50 sequences within a cell's genome.

In some embodiments, the method described herein, for example, the base editing methods has minimum to no off-target effects. In some embodiments, the method described
25 herein, for example, the base editing methods, has minimal to no chromosomal translocations.

In some embodiments, the base editing method described herein results in at least 50% of a cell population that have been successfully edited (*i.e.*, cells that have been successfully engineered). In some embodiments, the base editing method described herein results in at least 55% of a cell population that have been successfully edited. In some embodiments, the base
30 editing method described herein results in at least 60% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at

least 65% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 70% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 75% of a cell population that have been successfully edited. In some embodiments, the

5 base editing method described herein results in at least 80% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 85% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 90% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at

10 least 95% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of a cell population that have been successfully edited.

In some embodiments, the percent of viable cells in a cell population following a base editing intervention is greater than at least 60%, 70%, 80%, or 90% of the starting cell

15 population at the time of the base editing event. In some embodiments, the percent of viable cells in a cell population following editing is about 70%. In some embodiments, the percent of viable cells in a cell population following editing is about 75%. In some embodiments, the percent of viable cells in a cell population following editing is about 80%. In some embodiments, the percent of viable cells in a cell population as described above is about 85%. In

20 some embodiments, the percent of viable cells in a cell population as described above is about 90%, or about 91%, 92%, 93%, 94% 95%, 96%, 97%, 98%, 99%, or 100% of the cells in the population at the time of the base editing event. In some embodiments an engineered cell population can be further expanded in vitro by about 2 fold, about 3-fold, about 4-fold, about 5-fold, about 6-fold, about 7-fold, about 8-fold, about 9-fold, about 10-fold, about 15-fold, about

25 20-fold, about 25-fold, about 30-fold, about 35-fold, about 40-fold, about 45-fold, about 50-fold, or about 100-fold.

In embodiments, the cell population is a population of cells contacted with a base editor, complex, or base editor system of the present disclosure.

The number of intended mutations and indels can be determined using any suitable

30 method, for example, as described in International PCT Application Nos. PCT/US2017/045381 (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 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 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.

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

Details of base editor efficiency are described in International PCT Application Nos. PCT/US2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields

C:G-to-T:A base editors with higher efficiency and product purity” Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference. In some embodiments, editing of a plurality of nucleobase pairs in one or more genes using the methods provided herein results in formation of at least one intended mutation. In some embodiments, said formation of said at least one intended mutation results in the disruption the normal function of a gene. In some embodiments, said formation of said at least one intended mutation results decreases or eliminates the expression of a protein encoded by a gene. It should be appreciated that multiplex editing can be accomplished using any method or combination of methods provided herein.

10

DELIVERY SYSTEM

The suitability of nucleobase editors to target one or more nucleotides in a polynucleotide sequence (*e.g.*, a FcRn polynucleotide sequence) may be evaluated as described herein. In one embodiment, a single cell type of interest is transfected, transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a base editing system described herein together with a small amount of a vector encoding a reporter (*e.g.*, GFP). These cells can be any cell line known in the art, including any hepatocyte cell line (*e.g.*, primary human hepatocytes), endothelial cell line, epithelial cell line, or myeloid lineage cell line. Alternatively, primary cells (*e.g.*, human) may be used. Cells may also be obtained from a subject or individual, such as from tissue biopsy, surgery, blood, plasma, serum, or other biological fluid. Such cells may be relevant to the eventual cell target.

Delivery may be performed using a viral vector. In one embodiment, transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation. Following transfection, expression of a reporter (*e.g.*, GFP) can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different nucleobase editors to determine which combinations of editors give the greatest activity. The system can comprise one or more different vectors. In one embodiment, the base editor is codon optimized for expression of the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell. The activity of the nucleobase editor may be assessed as described herein, *i.e.*, by sequencing the genome of the cells to detect alterations in a target sequence. For Sanger

30

sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, minipreped and sequenced with a single primer. Sequencing may also be performed using next generation sequencing (NGS) techniques. When using next generation sequencing, amplicons may be 300-500 bp with the intended cut site placed asymmetrically. Following PCR, next
5 generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, *e.g.*, for use in high throughput sequencing (for example on an Illumina MiSeq). The fusion proteins or complexes that induce the greatest levels of target specific alterations in initial tests can be selected for further evaluation.

In particular embodiments, the nucleobase editors are used to target polynucleotides of
10 interest. In one embodiment, a nucleobase editor of the invention is delivered to cells (*e.g.*, hepatocytes, endothelial cells, epithelial cells, myeloid cells, or progenitors thereof) in conjunction with one or more guide RNAs that are used to target one or more nucleic acid sequences of interest within the genome of a cell, thereby altering the target gene(s) (*e.g.*, a gene encoding an FcRn polypeptide). In some embodiments, a base editor is targeted by one or more
15 guide RNAs to introduce one or more edits to the sequence of one or more genes of interest (*e.g.*, a gene encoding an FcRn polypeptide).

In some embodiments, the host cell is selected from a bacterial cell, plant cell, insect cell, human cell, or mammalian cell. In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell. In some embodiments, the cell is *in vitro*. In
20 some embodiments, the cell is *in vivo*.

Nucleic Acid-Based Delivery of Base Editor Systems

Nucleic acid molecules encoding a base editor system according to the present disclosure can be administered to subjects or delivered into cells *in vitro* or *in vivo* by art-known methods or
25 as described herein. For example, a base editor system comprising a deaminase (*e.g.*, cytidine or adenine deaminase) can be delivered by vectors (*e.g.*, viral or non-viral vectors), or by naked DNA, DNA complexes, lipid nanoparticles, or a combination of the aforementioned compositions.

Nanoparticles, which can be organic or inorganic, are useful for delivering a base editor
30 system or component thereof. Nanoparticles are well known in the art and any suitable nanoparticle can be used to deliver a base editor system or component thereof, or a nucleic acid

molecule encoding such components. In one example, organic (*e.g.* lipid and/or polymer) nanoparticles are 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 17** (below).

5

Table 17. Lipids used for gene transfer.

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
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyloxy-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
Ethylidimyristoylphosphatidylcholine	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

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

Table 18 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 18. Polymers used for gene transfer.

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

5

Table 19 summarizes delivery methods for a polynucleotide encoding a fusion protein or complex described herein.

Table 19. Delivery methods.

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical	(e.g., electroporation, particle gun, Calcium Phosphate transfection)	YES	Transient	NO	Nucleic Acids and Proteins
Viral	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modification	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
Non-Viral	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 base editor system components or nucleic acids encoding such components, for example, a polynucleotide programmable nucleotide binding domain (e.g., Cas9) such as, for example, Cas9 or variants thereof, and a gRNA targeting a nucleic acid sequence of interest, may be accomplished by delivering the ribonucleoprotein

- (RNP) to cells. The RNP comprises a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9), in complex with the targeting gRNA. RNPs or polynucleotides described herein 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, which is incorporated by reference in its entirety. 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).
- Nucleic acid molecules encoding a base editor system can be delivered directly to cells (*e.g.*, hepatocytes or other cells in the liver, endothelial cells, epithelial cells, and myeloid cells, or precursors thereof) as naked DNA or RNA by means of transfection or electroporation, for example, or can be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells. Vectors encoding base editor systems and/or their components can also be used. In particular embodiments, a polynucleotide, *e.g.* a mRNA encoding a base editor system or a functional component thereof, may be co-electroporated with one or more guide RNAs as described herein.
- Nucleic acid vectors can comprise one or more sequences encoding a domain of a fusion protein or complex described herein. A vector can also encode a protein component of a base editor system operably linked to a nuclear localization signal, nucleolar localization signal, or mitochondrial localization signal. As one example, a vector can include a Cas9 coding sequence that includes one or more nuclear localization sequences (*e.g.*, a nuclear localization sequence from SV40), and one or more deaminases.
- The vector can also include any suitable number of regulatory/control elements, *e.g.*, promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, or internal ribosome entry sites (IRES). These elements are well known in the art.

Vectors according to this disclosure include recombinant viral vectors. Exemplary viral vectors are set forth herein above. Other viral vectors known in the art can also be used. In addition, viral particles can be used to deliver base editor system components in nucleic acid and/or protein form. For example, “empty” viral particles can be assembled to contain a base editor system or component as cargo. Viral vectors and viral particles can also be engineered to incorporate targeting ligands to alter target tissue specificity.

Vectors described herein may comprise regulatory elements to drive expression of a base editor system or component thereof. Such vectors include adeno-associated viruses with inverted long terminal repeats (AAV ITR). The use of AAV-ITR 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 can be used to reduce potential toxicity due to over expression.

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

In some embodiments, a base editor system 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.

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).

In particular embodiments, a fusion protein or complex of the invention is encoded by a polynucleotide present in a viral vector (*e.g.*, adeno-associated virus (AAV), AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10, and variants thereof), or a suitable capsid protein of any viral vector. Thus, in some aspects, the disclosure relates to the viral delivery of a fusion protein or complex. Examples of viral vectors include retroviral vectors (*e.g.* Maloney murine leukemia virus, MML-V), adenoviral vectors (*e.g.* AD100), lentiviral vectors (HIV and FIV-based vectors), herpesvirus vectors (*e.g.* HSV-2).

In some aspects, the methods described herein for editing specific genes in a cell can be used to genetically modify the cell.

Viral Vectors

A base editor described herein can be delivered with a viral vector. In some embodiments, a base editor disclosed herein can be encoded on a nucleic acid that is contained in a viral vector. In some embodiments, 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 embodiments, 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.

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

Viral vectors can include lentivirus (*e.g.*, HIV and FIV-based vectors), Adenovirus (*e.g.*, AD100), Retrovirus (*e.g.*, Maloney murine leukemia virus, MML-V), herpesvirus vectors (*e.g.*,

HSV-2), and Adeno-associated viruses (AAVs), or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No. 8,404,658 (formulations, doses for AAV) and U.S. Patent No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For 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.

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

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
5 efficient packaging and delivery into a target cell via a retroviral vector. In some embodiments, 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.

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,
10 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.
15 For example, Adeno-associated virus (“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
20 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.

In applications where transient expression is preferred, adenoviral based systems can be
25 used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. 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
30 (See, *e.g.*, West *et al.*, Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). 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).

5 In some embodiments, AAV vectors are used to transduce a cell of interest with a polynucleotide encoding a base editor or base editor system as provided herein. 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).
10 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
15 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 or complex of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail
20 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.

 Viral vectors can be selected based on the application. For example, for in vivo gene
25 delivery, AAV can be advantageous over other viral vectors. In some embodiments, 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 embodiments, AAV allows low probability of causing insertional mutagenesis because it doesn't integrate into the host genome. Adenoviruses are commonly used as vaccines because of the strong
30 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.

AAV has a packaging capacity of about 4.5 Kb or 4.75 Kb including two 145 base inverted terminal repeats (ITRs). 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.

5

10 In some embodiments, the disclosed base editors are 4.5 kb or less in length.

An AAV can be AAV1, AAV2, AAV5, AAV6 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)).

15

In some embodiments, lentiviral vectors are used to transduce a cell of interest with a polynucleotide encoding a base editor or base editor system as provided herein. 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.

20

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

25

30

reagent). After 6 hours, the media is changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

Lentivirus can be purified as follows. Viral supernatants are harvested after 48 hours. Supernatants are first cleared of debris and filtered through a 0.45 µm low protein binding (PVDF) filter. They are then spun in an ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets are resuspended in 50 µl of DMEM overnight at 4° C. They are then aliquoted and immediately frozen at -80°C.

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 via a subretinal injection. In another embodiment, use of self-inactivating lentiviral vectors are 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 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 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.

5 Other suitable inteins will be apparent to a person of skill in the art.

A fragment of a fusion protein or complex 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.

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

Non-Viral Platforms for Gene Transfer

25 Non-viral platforms for introducing a heterologous polynucleotide into a cell of interest are known in the art.

For example, the disclosure provides a method of inserting a heterologous polynucleotide into the genome of a cell using a Cas9 or Cas12 (e.g., Cas12b) ribonucleoprotein complex (RNP)-DNA template complex where an RNP including a Cas9 or Cas12 nuclease domain and a guide RNA, wherein the guide RNA specifically hybridizes to a target region of the genome of the cell, and wherein the Cas9 nuclease domain cleaves the target region to create an insertion

site in the genome of the cell. A DNA template is then used to introduce a heterologous polynucleotide. In embodiments, the DNA template is a double-stranded or single-stranded DNA template, wherein the size of the DNA template is about 200 nucleotides or is greater than about 200 nucleotides, wherein the 5' and 3' ends of the DNA template comprise nucleotide sequences that are homologous to genomic sequences flanking the insertion site. In some embodiments, the DNA template is a single-stranded circular DNA template. In embodiments, the molar ratio of RNP to DNA template in the complex is from about 3:1 to about 100:1.

In some embodiments, the DNA template is a linear DNA template. In some examples, the DNA template is a single-stranded DNA template. In certain embodiments, the single-stranded DNA template is a pure single-stranded DNA template. In some embodiments, the single stranded DNA template is a single-stranded oligodeoxynucleotide (ssODN).

In some embodiments, the nucleic acid sequence is inserted into the genome of the cell via non-viral delivery. In non-viral delivery methods, the nucleic acid can be naked DNA, or in a non-viral plasmid or vector.

In some embodiments, the nucleic acid is inserted into the cell by introducing into the cell, (a) a targeted nuclease that cleaves a target region to create an insertion site in the genome of the T cell; and (b) the nucleic acid sequence, wherein the nucleic acid sequence is incorporated into the insertion site by HDR.

In some cases, the nucleic acid sequence is introduced into the cell as a linear DNA template. In some cases, the nucleic acid sequence is introduced into the cell as a double-stranded DNA template. In some cases, the DNA template is a single-stranded DNA template. In some cases, the single-stranded DNA template is a pure single-stranded DNA template. As used herein, by "pure single-stranded DNA" is meant single-stranded DNA that substantially lacks the other or opposite strand of DNA. By "substantially lacks" is meant that the pure single-stranded DNA lacks at least 100-fold more of one strand than another strand of DNA. In some cases, the DNA template is a double-stranded or single-stranded plasmid or mini-circle.

In other embodiments, a single-stranded DNA (ssDNA) can produce efficient HDR with minimal off-target integration. In one embodiment, an ssDNA phage is used to efficiently and inexpensively produce long circular ssDNA (cssDNA) donors. These cssDNA donors serve as efficient HDR templates when used with Cas9 or Cas12 (e.g., Cas12a, Cas12b), with integration frequencies superior to linear ssDNA (lssDNA) donors.

Methods for integrating such templates are known in the art and described, for example, in US Patent Publications No. 20190388469, 20210388362, 20210207174, 20210353678, 20200362355, and 20210228631, which are incorporated herein by reference. See also, Roth, T.L et al., Reprogramming human T cell function and specificity with non-viral genome
5 targeting. Nat. Lett. 559, 405–409 (2018); Ferenczi et al., Nat Commun 12, 6751 (2021). doi.org/10.1038/s41467-021-27004-1; Zhang et al., Homology-based repair induced by CRISPR-Cas nucleases in mammalian embryo genome editing. Protein Cell (2021).

Inteins

10 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.

15 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 folding of the intein domain.

20 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 series of acyl-transfer reactions that result in the cleavage of two peptide bonds at the intein-extein junctions and the
25 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 conserved threonine and histidine (referred to as the TXXH motif), along with a commonly found aspartate, which results
30 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 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
5 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, the split intein is selected from Gp41.1, IMPDH.1, NrdJ.1 and Gp41.8 (Carvajal-Vallejos, Patricia *et al.* “Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources.” J. Biol. Chem., vol. 287,34
10 (2012)).

Non-limiting examples of inteins include any intein or intein-pair known in the art, which include 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), and DnaE. Non-limiting examples of
15 pairs of inteins that may be used in 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 in the Sequence Listing at SEQ ID NOs: 370-377. Inteins suitable for use in embodiments of the
20 present disclosure and methods for use thereof are described in U.S. Patent No. 10,526,401, International Patent Application Publication No. WO 2013/045632, and in U.S. Patent Application Publication No. US 2020/0055900, the full disclosures of which are incorporated herein by reference in their entireties by reference for all purposes.

Further non-limiting examples of amino acid and nucleotide sequences for N-inteins and C-inteins suitable for use as intein pairs include those with at least 85% sequence identity to an
25 amino acid or nucleotide sequence listed in the following **Tables 20A-20C**, or a fragments thereof that function as part of a split intein pair.

Table 20A. Exemplary amino acid and nucleotide sequences for N-Inteins.

N-Intein	Amino Acid or Nucleotide Sequence	SEQ ID NO
Cfa(GEP) (nucleotide sequence)	TGCCTGAGCTACGATACCGAGATCCTGACCGTGGAATACGGCTT CCTGCCTATCGGCAAGATCGTCGAGGAACGGATCGAGTGCACAG TGTACACCGTGGATAAGAATGGCTTCGTGTACACCCAGCCTATC GCTCAGTGGCACAACAGAGGCGAGCAAGAGGTGTTTCGAGTACTG CCTGGAAGATGGCAGCATCATCCGGGCCACCAAGGACCACAAGT TTATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTT GAGAGAGGCCCTGGACCTGAAACAGGTGGACGGACTGCCT	389
Cfa(GEP) (amino acid sequence)	CLSYDTEILTVEYGFLLPIGKIVEERIECTVYTVDKNGFVYTQPI AQWHNRGEQEVFEYCLEDSIIIRATKDHKFMTTDGQMLPIDEIF ERGLDLKQVDGLP	390
Gp41.1 (nucleotide sequence)	TGTCTGGACCTCAAGACCCAAGTGCAGACACCTCAGGGCATGAA AGAGATTAGCAATATCCAGGTGGGCGACCTGGTCCTGAGCAACA CCGGCTACAACGAGGTGCTGAACGTGTTCCCTAAGTCCAAGAAG AAATCTTATAAGATCACCTGGAAGATGGCAAGGAAATCATCTG CAGCGAGGAACACCTGTTCCCCACCCAGACCGGCGAGATGAACA TCAGCGGCGGACTGAAGGAGGGCATGTGCCTGTACGTGAAGGAG	391
Gp41.1 (amino acid sequence)	CLDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKK KSYKITLEDGKEIICSEEHLEFPTQTGEMNISGGLKEGMCLYVKE	392
Gp41.8 (nucleotide sequence)	TGCCTGAGCCTGGACACCATGGTGGTGACAAACGGCAAGGCCAT CGAGATCAGAGATGTGAAGGTGGGAGATTGGCTGGAAAGCGAAT GTGGCCAGTGCAGGTACAGAGGTGCTGCCTATCATCAAGCAG CCTGTCTTTGAGATTGTGCTGAAAAGCGGAAAAAAGATCCGGGT GTCCGCTAATCACAAGTTCCCCACCAAGGACGGCCTCAAGACCA TCAACAGCGGCCTGAAGGTGGGCGACTTCCTGAGAAGCAGAGCC AAG	393
Gp41.8 (amino acid sequence)	CLSLDTMVVTNGKAIEIRDVKVGDWLESECGPVQVTEVLPIIKQ PVFEIVLKSGKKIRVSANHKFPTKDGLKTINSLKVGDFLRSRA K	394
IMPDH.1 (nucleotide sequence)	TGTTTTGTGCCTGGCACCTGGTGAACACAGAGAATGGCCTGAA GAAAATCGAGGAAATCAAGGTGGGCGACAAGGTGTTTCAGCCATA CAGGCAAGCTGCAGGAGGTGGTGGACACCCTGATCTTCGACCGG GACGAGGAAATCATCTCTATCAACGGCATTGATTGCACCAAGAA CCACGAGTTCTACGTGATCGATAAGGAAAACGCTAATAGAGTGA ACGAGGACAACATCCACCTCTTCGCCAGATGGGTCCACGCCGAG GAACTGGATATGAAAAGCACCTGCTGATCGAGCTGGAA	395

N-Intein	Amino Acid or Nucleotide Sequence	SEQ ID NO
IMPDH.1 (amino acid sequence)	CFVPGTLVNTENGLKKIEEIKVGDKVFSHTGKLQEVVDTLIFDR DEEIIISINGIDCTKNHEFYVIDKENANRVNEDNIHLFARWVHAE ELDMKKHLLIELE	396
NrdJ.1 (nucleotide sequence)	TGCCTGGTGGGCTCTAGCGAGATTATCACAAGAACTACGGCAA GACCACCATCAAGGAAGTGGTCGAGATCTTCGACAACGACAAGA ATATCCAGGTGCTGGCCTTCAACACCCACACCGATAATATCGAG TGGGCCCCATCAAGGCCGCTCAGCTGACCAGACCTAACGCCGA GCTGGTTGAACTGGAAATCGACACCCTGCACGGCGTGAAAACAA TCCGGTGCACCCCTGACCACCCCGTGTACACCAAGAACAGAGGC TACGTGCGGGCCGACGAGCTGACAGATGATGACGAGCTCGTGGT GGCTATC	397
NrdJ.1 (amino acid sequence)	CLVGSSEIITRNYGKTTIKEVVEIFDNDKNIQVLA FNTHTDNIE WAPIKAAQLTRPNAELVELEIDTLHG VKTIRCTPDHPVYTKNRG YVRADELTDDELVVAI	398
Npu (nucleotide sequence)	TGCCTGAGCTACGAGACAGAGATCCTGACCGTGGAATATGGCCT GCTGCCAATCGGAAAGATCGTGGAAGCGGATCGAGTGCACCG TCTACAGCGTGGACAACAACGGAAATATCTATACACAGCCTGTG GCCCAATGGCAGACCGGGCGAACAGGAGGTGTTTGAGTACTG CCTGGAAGATGGTTCTCTGATTAGAGCCACCAAGGACCACAAGT TCATGACCGTCGACGGCCAGATGCTGCCCATCGACGAAATCTTC GAGCGGGAACCTCGACCTGATGAGAGTGGATAACCTGCCCAAT	399
Npu (amino acid sequence)	CLSYETEILTVEYGLLP I G K I V E K R I E C T V Y S V D N N G N I Y T Q P V A Q W H D R G E Q E V F E Y C L E D G S L I R A T K D H K F M T V D G Q M L P I D E I F E R E L D L M R V D N L P N	400
Cfa N-intein	CLSYDTEILTVEYGLFP I G K I V E E R I E C T V Y T V D K N G F V Y T Q P I A Q W H N R G E Q E V F E Y C L E D G S I I R A T K D H K F M T T D G Q M L P I D E I F E R G L D L K Q V D G L P	401
Npu N-intein	CLSYETEILTVEYGLLP I G K I V E K R I E C T V Y S V D N N G N I Y T Q P V A Q W H D R G E Q E V F E Y C L E D G S L I R A T K D H K F M T V D G Q M L P I D E I F E R E L D L M R V D N L P N	402

Table 20B. Further exemplary amino acid and nucleotide sequences for N-Inteins.

N-Intein-SC	Amino Acid or Nucleotide Sequence	SEQ ID NO
Gp41.1 (nucleotide sequence)	ACAAGAAGCGGATACTGTCTGGACCTCAAGACCCAAGTGCAGACA CCTCAGGGCATGAAAGAGATTAGCAATATCCAGGTGGGCGACCTG GTCCTGAGCAACACCGGCTACAACGAGGTGCTGAACGTGTTCCCT AAGTCCAAGAAGAAATCTTATAAGATCACCTGGAAGATGGCAAG GAAATCATCTGCAGCGAGGAACACCTGTTCCCCACCCAGACCGGC GAGATGAACATCAGCGGCGGACTGAAGGAGGGCATGTGCCTGTAC GTGAAGGAG	403
Gp41.1 (amino acid sequence)	TRSGYCLDLKTQVQTPQGMKEISNIQVGDVLVLSNTGYNEVLNVFP KSKKKSYPKITLEDGKEIICSEHLEFPTQTGEMNISGGLKEGMCLY VKE	404
Gp41.8 (nucleotide sequence)	TCTCAGCTGAACCGGTGCCTGAGCCTGGACACCATGGTGGTGACA AACGGCAAGGCCATCGAGATCAGAGATGTGAAGGTGGGAGATTGG CTGGAAAGCGAATGTGGCCAGTGCAGGTTACAGAGGTGCTGCCT ATCATCAAGCAGCCTGTCTTTGAGATTGTGCTGAAAAGCGGAAAA AAGATCCGGGTGTCCGCTAATCACAAGTTCCCCACCAAGGACGGC CTCAAGACCATCAACAGCGGCCTGAAGGTGGGCGACTTCCTGAGA AGCAGAGCCAAG	405
Gp41.8 (amino acid sequence)	SQLNRCLSLDTMVVTNGKAIEIRDVKVGDWLESECGPVQVTEVLP IIKQPVFEIVLKSGKKIRVSANHKFPTKDGLKTINSGLKVGDFLR SRAK	406
IMPDH.1 (nucleotide sequence)	GGCATCGGCGGAGGATGTTTTGTGCCTGGCACCCCTGGTGAACACA GAGAATGGCCTGAAGAAAATCGAGGAAATCAAGGTGGGCGACAAG GTGTTTCAGCCATACAGGCAAGCTGCAGGAGGTGGTGGACACCCTG ATCTTCGACCGGGACGAGGAAATCATCTCTATCAACGGCATTGAT TGCACCAAGAACCACGAGTTCTACGTGATCGATAAGGAAAACGCT AATAGAGTGAACGAGGACAACATCCACCTCTTCGCCAGATGGGTC CACGCCGAGGAAGTGGATATGAAAAAGCACCTGCTGATCGAGCTG GAA	407
IMPDH.1 (amino acid sequence)	GIGGGCFVPGLVNTENGLKKIEEIKVGDVFSHTGKLQEVVDTL IFDRDEEIIISINGIDCTKNHEFYVIDKENANRVNEDNIHLFARWV HAEELDMKKHLLIELE	408

N-Intein-SC	Amino Acid or Nucleotide Sequence	SEQ ID NO
NrdJ.1 (nucleotide sequence)	GGAACAAACCCATGTTGCCTGGTGGGCTCTAGCGAGATTATCACA AGAACTACGGCAAGACCACCATCAAGGAAGTGGTCGAGATCTTC GACAACGACAAGAATATCCAGGTGCTGGCCTTCAACACCCACACC GATAATATCGAGTGGGCCCCCTATCAAGGCCGCTCAGCTGACCAGA CCTAACGCCGAGCTGGTTGAACTGGAAATCGACACCCTGCACGGC GTGAAAACAATCCGGTGCACCCCTGACCACCCCGTGTACACCAAG AACAGAGGCTACGTGCGGGCCGACGAGCTGACAGATGATGACGAG CTCGTGGTGGCTATC	409
NrdJ.1 (amino acid sequence)	GTNPCCLVGSSEIIITRNYGKTTIKEVVEIFDNDKNIQVLA FNHT DNIEWAPIKAAQLTRPNAELVELEIDTLHGVKTIRCTPDHPVYTK NRGYVRADELTDDELVVAI	410

Table 20C. Exemplary amino acid and nucleotide sequences for C-Inteins.

C-Intein	Amino Acid or Nucleotide Sequence	SEQ ID NO
Cfa(GEP) (nucleotide sequence)	GTCAAGATCATCAGCAGAAAGAGCCTGGGCACCCAGAACGTGTA CGATATCGGAGTGGGCGAGCCCCACAACCTTTCTGCTCAAGAATG GCCTGGTGGCCAGCAAC	411
Cfa(GEP) (amino acid sequence)	VKIIISRKSLGTQNVYDIGVGEPHNLLKNGLVASN	412
Gp41.1 (nucleotide sequence)	ATGATGCTGAAAAAGATCCTGAAGATCGAGGAACTGGATGAGAG AGAGCTGATCGACATCGAAGTGTCTGGCAATCACCTGTTCTACG CCAACGACATCCTGACCCACAACAGC	413
Gp41.1 (amino acid sequence)	MMLKKILKIEELDERELIDIEVSGNHLFYANDILTHNS	414
Gp41.8 (nucleotide sequence)	ATGTGCGAAATCTTCGAGAACGAGATTGATTGGGACGAAATCGC CTCTATCGAGTACGTGGGCGTGGAAAGAGACAATCGACATCAACG TGACCAACGACAGACTGTTTTTCGCCAATGGCATCCTGACCCAC AACAGC	415
Gp41.8 (amino acid sequence)	MCEIFENEIDWDEIASIEYVGVEETIDINVTNDRLFFANGILTH NS	416
IMPDH.1 (nucleotide sequence)	ATGAAATTCAAGCTGAAGGAAATCACCAGCATCGAGACAAAGCA CTACAAGGGCAAGGTGCACGATCTGACCGTGAACCAGGACCACA GCTACAACGTCAGAGGCACCGTGGTGCATAATTCT	417
IMPDH.1 (amino acid sequence)	MKFKLKEITSIETKHYKGKVHDLTVNQDHSYNVRGTVVHNS	418

C-Intein	Amino Acid or Nucleotide Sequence	SEQ ID NO
NrdJ.1 (nucleotide sequence)	ATGGAAGCCAAGACCTACATCGGCAAGCTGAAATCTAGAAAGATCGTGTCCAACGAGGATACATACGACATCCAGACCAGCACCCACAATTTCTTCGCCAACGACATCCTGGTGCACAACAGC	419
NrdJ.1 (amino acid sequence)	MEAKTYIGKLKSRKIVSNEDTYDIQTSTHNEFFANDILVHNS	420
Npu (nucleotide sequence)	ATGATCAAGATCGCCACAAGAAAGTACCTGGGCAAGCAGAACGTGTACGACATCGGCGTGGAGAGAGACCACAACCTTCGCCCTGAAGACGGCTTTATCGCCTCTAAT	421
Npu (amino acid sequence)	MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN	422
Cfa C-intein (amino acid sequence)	MVKIISRKSLGTQNVYDIGVGEPHNFLKNGLVASN	423
Npu C-intein (amino acid sequence)	IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN	424

Intein-N and intein-C may be fused to the N-terminal portion of a 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

5 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 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. In embodiments, a base editor is encoded by two polynucleotides, where one polynucleotide encodes a fragment of the

10 base editor fused to an intein-N and another polynucleotide encodes a fragment of the base editor fused to an intein-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 inteins are known in the art and described, for example by WO2014004336,

15 WO2017132580, WO2013045632A1, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

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 or complex 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, 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. 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. In some embodiments, an N-terminal fragment of a nucleic acid programmable DNA binding protein (napDNAbp) domain (*e.g.*, Cas9) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. In some embodiments, an N-terminal fragment of a deaminase domain (*e.g.*, adenosine or cytidine deaminase) 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. 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, inteins are utilized to join fragments or portions of a cytidine or adenosine base editor protein that is grafted onto an AAV capsid protein. 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.

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 capital letters in the sequence below (called the “Cas9 reference sequence”).

```

1  mdkkysigld igtntsvgwav itdeykvpsk kfkvlgnldr hsiiknliga llfdsgetae
61 atrlkrtrarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlndpnsd

```

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 vklnredllr kqrtfdngsi phqihlgelh
 5 421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**ArgnS** rf**Aw**m**TrkSe** e**Ti**Tpwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkvy tegmrkpaf1
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**CfdS**vei sgvedrfr**AS** lgtyhdllki
 601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfdkvmkq lkrtrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgds1
 10 721 hehianlags paikkilqt vkvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltres dknrgksdnv pseevkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vagildsrnm tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrv
 15 1021 miaksegeig katakyffys nimffktei tlangeirkr plietngetg eivwdkgrdf
 1081 atvrkvlsmv qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
 1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
 1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
 1261 qhkhyldiei eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltlnlga
 20 1321 paafkyfddt idrkrytstk evldatlihq sitglyetri dlsq1ggd

(SEQ ID NO: 197)

Pharmaceutical Compositions

In some aspects, the present invention provides a pharmaceutical composition comprising
 25 any of the polynucleotides, vectors, editors, e.g., base editors, editor systems, e.g., base editor
 systems, guide polynucleotides, fusion proteins, complexes, fusion protein-guide polynucleotide
 complexes, LNPs, or cells described herein.

The pharmaceutical compositions of the present invention can be prepared in accordance
 with known techniques. See, e.g., Remington, The Science And Practice of Pharmacy (21st ed.
 30 2005). In general, the polynucleotides, vectors, editors, editor systems, guide polynucleotides,
 fusion proteins, complexes, or the fusion protein-guide polynucleotide complexes, LNPs, cells,
 or population thereof is admixed with a suitable carrier prior to administration or storage, and in
 some embodiments, the pharmaceutical composition further comprises a pharmaceutically
 acceptable carrier. Suitable pharmaceutically acceptable carriers generally comprise inert
 35 substances that aid in administering the pharmaceutical composition to a subject, aid in

processing the pharmaceutical compositions into deliverable preparations, or aid in storing the pharmaceutical composition prior to administration. Pharmaceutically acceptable carriers can include agents that can stabilize, optimize or otherwise alter the form, consistency, viscosity, pH, pharmacokinetics, solubility of the formulation. Such agents include buffering agents, wetting agents, emulsifying agents, diluents, encapsulating agents, and skin penetration enhancers. For example, carriers can include, but are not limited to, saline, buffered saline, dextrose, arginine, sucrose, water, glycerol, ethanol, sorbitol, dextran, sodium carboxymethyl cellulose, and combinations thereof.

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.

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,
5 *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
10 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
15 concentration sufficient to modulate the osmotic properties of the formulation.

Pharmaceutical compositions of the present invention can include at least one additional therapeutic agent useful in the treatment of disease. For example, some embodiments of the pharmaceutical composition described herein further comprises a chemotherapeutic agent. In some embodiments, the pharmaceutical composition further comprises a cytokine peptide or a
20 nucleic acid sequence encoding a cytokine peptide. In some embodiments, a pharmaceutical composition further comprises an immunosuppressive agent. In some embodiments, the pharmaceutical compositions can be administered separately from an additional therapeutic agent.

For any composition to be administered to an animal or human, and for any particular
25 method of administration, one can determine: toxicity, such as by determining the lethal dose (LD) and LD50 in a suitable animal model (*e.g.*, a rodent such as a mouse); and, the dosage of the composition(s), concentration of components therein, and the timing of administering the composition(s), which elicit a suitable response. Such determinations do not require undue experimentation from the knowledge of the skilled artisan, this disclosure and the documents
30 cited herein.

In some embodiments, the pharmaceutical composition is formulated for delivery to a subject. 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, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

In some embodiments, the pharmaceutical composition described herein is administered locally to a site of interest (*e.g.*, a liver). The site may be, *e.g.*, a diseased site or a site where a target gene is abundantly expressed. 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-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, *e.g.*, U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

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

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

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 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 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.

In some embodiments, any of the fusion proteins or nucleic acids encoding them, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins, nucleic acids, or complexes provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease (*e.g.*, Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. In some embodiments, pharmaceutical compositions comprise a lipid nanoparticle and a pharmaceutically acceptable excipient. In some embodiments, the lipid nanoparticle contains a gRNA, a base editor, a complex, a base editor system, or a component thereof of the present disclosure, and/or one or more polynucleotides encoding the same. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with any of the

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

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

20 Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical
25 formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro
30 (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and

known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011/053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease.

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

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

15 In some embodiments, compositions in accordance with the present disclosure can be used for treatment of any of a variety of diseases, disorders, and/or conditions.

Methods of Treatment

20 Some aspects of the present invention provide methods of treating a subject in need, the method comprising administering to a subject in need an effective therapeutic amount of a pharmaceutical composition as described herein. More specifically, the methods of treatment include administering to a subject in need thereof one or more pharmaceutical compositions comprising one or more agents provided herein. In other embodiments, the methods of the invention comprise expressing or introducing into a cell *in vivo* or *in vitro* a base editor

25 polypeptide and one or more guide RNAs capable of targeting a nucleic acid molecule encoding at least one polypeptide.

30 In another embodiment, a method of treating an IgG-mediated autoimmune disorder in a subject in need thereof is provided, the method comprising administering to the subject a LNP comprising: a lipid monolayer membrane comprising at least one Fc region of an IgG antibody or a functional fragment thereof embedded therein; and a lipid core matrix enclosed in the lipid

monolayer membrane, wherein the lipid core matrix comprises at least one siRNA or guide RNA that moderates expression of or silences an *FCGRT* gene.

One of ordinary skill in the art would recognize that multiple administrations of the pharmaceutical compositions contemplated in particular embodiments may be required to affect the desired therapy. For example, a composition may be administered to the subject 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more times over a span of 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 1 year, 2 years, 5 years, 10 years, or more. In any of such methods, the methods may comprise administering to the subject an effective amount of an editor system, e.g., a base editor system or polynucleotide encoding such system. In some embodiments a single administration is sufficient to produce a desired effect. In some embodiments between 1 and 5 administrations may be used.

Administration of the pharmaceutical compositions contemplated herein may be carried out using conventional techniques including, but not limited to, infusion, transfusion, or parenterally. In some embodiments, parenteral administration includes infusing or injecting intravascularly, intravenously, intramuscularly, intraarterially, intrathecally, intratumorally, intradermally, intraperitoneally, transtracheally, subcutaneously, subcuticularly, intraarticularly, subcapsularly, subarachnoidly and intrasternally.

In some embodiments, a composition described herein (e.g., base editor system) is administered in a dosage that is about 0.5-30 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-20 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.5-10 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 0.04 mg, about 0.08 mg, about 0.16 mg, about 0.32 mg, about 0.64 mg, about 1.25 mg, about 1.28 mg, about 1.92 mg, about 2.5 mg, about 3.56 mg, about 3.75 mg, about 5.0 mg, about 7.12 mg, about 7.5 mg, about 10 mg, about 14.24 mg, about 15 mg, about 20 mg, or about 30 mg per kilogram body weight of the human subject. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered two times a week. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the

human subject and the composition is administered two times a week. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a week. In another embodiment, the amount of the
5 composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a week. In another embodiment, the amount of the composition administered is about 1.92 mg, about 3.75 mg, about 7.5 mg, about 15.0 mg, or about 30.0 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a
10 seven day period. In another embodiment, the composition is administered intravenously once a day, seven times in a seven day period. In another embodiment, the amount of the composition administered is about 1.28 mg, about 2.56 mg, about 5.0 mg, about 10 mg, or about 20 mg per kilogram body weight of the human subject and the composition is administered once a day three, five or seven times in a seven day period. In another embodiment, the composition is
15 administered intravenously once a day, seven times in a seven day period.

In some embodiments, the composition is administered over a period of 0.25 h, 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h, 8 h, 9 h, 10 h, 11 h, or 12 h. In another embodiment, the composition is administered over a period of 0.25-2 h. In another embodiment, the composition is gradually administered over a period of 1 h. In another embodiment, the composition is gradually
20 administered over a period of 2 h.

Kits

The invention provides kits for the treatment of an autoimmune disorder in a subject. In some embodiments, the kit includes a genome editor system or a polynucleotide encoding a
25 genome editor system and a guide RNA. In some embodiments, for example, the kit includes a base editor system or a polynucleotide encoding a base editor system, wherein the base editor system comprises a nucleic acid programmable DNA binding protein (napDNAbp), a deaminase, and a guide RNA. In some embodiments, the napDNAbp is Cas9 or Cas12. In some
embodiments, the polynucleotide encoding the genome editor, e.g., base editor, is a mRNA
30 sequence. In some embodiments, the deaminase is a cytidine deaminase or an adenosine deaminase.

The kits may further comprise written instructions for using a genome editor, genome editor system, base editor, base editor system and/or edited cell as described herein. 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
5 (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 comprises instructions in the form of a label or separate insert (package insert) for suitable operational parameters. In yet another embodiment, the kit comprises one or more containers with appropriate positive and negative controls or control samples, to be used as standard(s) for detection, calibration, or
10 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.

15 The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal
20 Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the
25 invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the
30 inventors regard as their invention.

EXAMPLES

EXAMPLE 1: Base editing to alter an FcRn polynucleotide in HEK293T cells

Experiments were undertaken to develop base editor systems capable of altering an FcRn polynucleotide sequence in HEK293T cells. The polynucleotide alterations resulted in alterations to the encoded FcRn polypeptide sequence. Without intending to be bound by theory, the polypeptide alterations, which were selected based upon structural insights (see **2B-5**) fell into three categories (see **Table 21**): A) amino acid residues involved in stabilization of the complex of FcRn with IgG and/or binding affinity with IgG at neutral pH; B) amino acid residues corresponding to pH dependent FcRn IgG binding sites; and C) amino acid residues associated with a hydrophobic pocket that helps position W131(154) to facilitate binding to IgG.

Table 21. Target FcRn alterations

FcRn Residues	IgG1 Counter Residues	Predicted Function
E115(138)	H310	Stabilization of the complex and binding affinity at neutral PH
D130(153)	H435	
L135(158)	Y436	
L112(135)	I253	Known PH dependent FcRn IgG binding sites
W131(154)	I253	
E116(139)	H311	
P132(155)	NA	
E133(156)	I253	
N119(142)	NA	A hydrophobic pocket help positioning W131(154)
L122(145)	NA	
T126(145)	NA	
W127(150)	NA	

Amino acid residue numbers in parentheses included 23 AA signal peptide

Base editor systems each containing a guide polynucleotide and a base editor were designed to facilitate base edits to the FcRn gene corresponding to the FcRn alterations at the

positions listed in **Table 21** and correspond to guide numbers 1-40 listed in **Tables 2A and 2B**.

Base editing efficiencies were measured in HEK293T cells edited using base editor systems containing the following guides (see **Tables 2A and 2B**) and their corresponding base editors listed in **Table 2B**: gRNA1560

- 5 gRNA1561, gRNA1562, gRNA1563, gRNA1564, gRNA1565, gRNA1566, gRNA1567, gRNA1568, gRNA1569, gRNA1570, gRNA1571, gRNA1572, gRNA1573, gRNA1574, gRNA1575, gRNA1576, gRNA1577, gRNA1578, gRNA1579, gRNA1580, gRNA1581, gRNA1582, gRNA1583, gRNA1584, gRNA1587, gRNA1588, gRNA1589, gRNA1590, gRNA1591, gRNA1592, and gRNA1593. The base editing efficiencies measured for each base editing system are shown in **FIGs. 6A and 6B**, where the base editing efficiencies for each amino acid alteration or set of alterations observed in the edited HEK293T cells was evaluated separately. **Table 22** provides a summary of main and bystander amino acid alterations associated with base editor systems that were evaluated.
- 10

15 **Table 22. Summary of main and bystander amino acid alterations**

gRNA	Editor	Main Alteration(s)	Bystander Alteration(s)
gRNA1572; gRNA1577	CBE	E116K; E115K+E116K (~20%)	E116Q, M118I (~3%)
gRNA1563; gRNA1581	ABE	D130G	
gRNA1567	ABE	L112P	F110L
gRNA1583	ABE	W131R (~74%)	D130N, D130H
gRNA1561; gRNA1579	ABE	N119S	N119G, N119D, N119C
gRNA1562	ABE	T126A	
gRNA1562	CBE	T126I	T126S, T126N
gRNA1569; gRNA1570	CBE	L122F	
gRNA1582; gRNA1589	ABE	E133G (~20%)	A134V, I137V

The following base editor systems were used in these experiments.

Guides	Base editors
gRNA1572	spCas9-BE4
20 gRNA1577	VRQR spCas9-BE4
gRNA1563	SpCas9-ABE8.8
gRNA1581	spCas9-ABE
gRNA1567	SpCas9-ABE8.8
gRNA1583	spCas9-ABE
25 gRNA1561	SpCas9-ABE8.8
gRNA1579	spCas9-ABE

	gRNA1562	SpCas9-ABE8.8
	gRNA1562	spCas9-BE4
	gRNA1569	spCas9-BE4
	gRNA1570	spCas9-BE4
5	gRNA1582	spCas9-ABE
	gRNA1589	KKH-saCas9-ABE8.8

To determine the impact of amino acid alterations on the function of FcRn, experiments were undertaken using surface plasmon resonance to evaluate binding of IgG1 and Albumin to FcRn polypeptides containing one of the following 10 alterations: D130G, W131R, D130N, D130H, F110L, E116K, E115K, E116Q, M118I, and E133K. The FcRn proteins were produced as follows: EXPI293F cells were transfected in accordance with the Expi293 Expression System User Guide (Catalog# A14525, Publication# MAN0019402). Plasmid ratios for each construct were 1:1 for extracellular domain of FcRn (Wildtype or mutant) and beta2 microglobulin (Wildtype) (a protein with which FcRn associates). Transfected cells were cultured for an additional 4 days post-transfection before harvesting by centrifugation (7068xg for 30mins at room temp) and sterile filtration through 0.22um aPES membrane. Material was stored at 4degC until purification. Purification was performed using an AKTA Pure (Cytiva Life Sciences). Harvest material from the transfection step outlined above was purified over a 1mL HisTrap HP column (Cytiva Life Sciences Cat# 17-5247-01)

The affinities of FcRn variants towards human IgG1-4 and albumin (Sigma A9731) were determined using SPR on a Biacore 3000 instrument equipped with a C1 sensor chip (Cytiva cat. # BR100540) or CM5 sensor chip (Cytiva cat. # BR100399) at 25°C and a flow rate of 30 µL min⁻¹. The running buffer contained 10 mM sodium phosphate pH 5.8, 150 mM NaCl, and 0.05% Tween-20 (HBS-EP) and was adjusted to a final pH of 5.8 by dropwise addition of HCl, then sterile filtered and degassed by sonication under reduced pressure. Wild-type FcRn was purchased from Acro Biosystems (cat. # FCN-H52W7) and compared to wild-type and mutant FcRn produced as described above. Experiments were conducted in two orientations: 1) with FcRn immobilized on the surface, or 2) with soluble FcRn binding to immobilized albumin and IgG. In both cases, one min association and 5 min dissociation phases were used, and regeneration was performed with an injection of PBS, pH 7.4. The response was double-referenced by subtracting the responses from a reference flow cell (without protein)

immobilized), and a blank injection over the same flow cell. Equilibrium and kinetic parameters were determined by fitting the response to a 1:1 Langmuir binding model using Scrubber (BioLogic) software.

5 1) SPR measurements with surface-immobilized FcRn

For immobilization, FcRn was biotinylated with EZ-link Sulfo-NHS-biotinylation kit (ThermoFisher cat. # 31425) according to the manufacturer's instructions, except that the NHS-biotin was included at a 1:1 molar equivalent to sparsely biotinylate the protein. Neutravidin (ThermoFisher cat #31000) was covalently immobilized on the C1 SPR chip surface by
 10 activating the surface carboxylate groups with 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) for 7 minutes followed by a 3 min injection of neutravidin at 10 µg/ml in 10 mM sodium acetate pH 5.0. The neutravidin immobilization level was approximately 1000 RU. The activated surface was quenched with a 7 min injection of 10 wt% 3kDa poly(ethylene glycol) amine (PEG-amine, JenKem Technology
 15 cat. #M-NH₂HCl-3000) pH 8, which was found to reduce non-specific binding to the C1 sensor chip surface. Biotinylated FcRn was then captured on the SPR chip by injecting 1 µg/ml solutions until a capture level of 30 RU was reached. IgG isotypes 1-4 were injected at concentrations ranging up to 3 µM, and albumin was injected at concentrations up to 36 µM.

20 2) SPR measurements with surface-immobilized IgG and albumin

Human IgG and albumin were immobilized onto carboxymethyldextran-containing (CM5) SPR chips by NHS/EDC activation as above. After immobilization of IgG or albumin in 10 mM sodium acetate pH 5.0 to approximately 1000 RU, deactivation was performed with a 7 min injection of 1 M ethanolamine, pH 9.5. Soluble FcRn was screened at 3 µM for binding, and
 25 concentrations up to 15 µM were tested for select mutants.

All ten of the FcRn polypeptides evaluated maintained albumin binding (**FIG. 7A**). Four of the FcRn polypeptides evaluated had reduced IgG binding (**FIG. 7B**), namely W131R, E115K, M118I, and E133K. See also **FIGs. 7C and 7D**. **Tables 23 and 24** provide K_D values
 30 determined for binding of IgG1, IgG2, IgG3, IgG4, or albumin by surface-bound or non-surface-bound FcRn polypeptides containing one of the alterations E133K, M118I, and W131R. FcRn

polypeptides containing the alteration W131R showed non-binding with the IgG polypeptides but was capable of binding albumin. The W131R variant of FcRn showed higher binding affinity for albumin when in solution than wild-type FcRn. FcRn polypeptides containing the alteration M118I showed only intermediate affinity for the IgG polypeptides and were capable of binding albumin.

Having established that it may be beneficial to alter amino acid M118 of an FcRn polypeptide for therapeutic purposes, guide RNA gRNA3265 was designed for targeting a base editor system to effect the corresponding nucleobase alteration in an FcRn polynucleotide. The gRNA3265 was predicted to facilitate the alterations M118(141)T and F117(140)P in an FcRn polypeptide. Also, a base editor system containing gRNA1578 achieved base editing efficiencies of about 21% for the target FcRn polypeptide alteration M118(141)V in HEK293T cells (see FIG. 6A).

Table 23. FcRn-biotin on surface, K_D for binding pairs [nM]

FcRn	IgG1	IgG2	IgG3	IgG4	albumin
wild-type	520	870	610	1500	1400
E133K	non-specific	>> 3300	>> 1000	>> 4400	non-specific
M118I	5200*	6600*	4200*	12600*	1900
W131R	>> 6000	>> 6000	>> 3000	>> 6000	1700

Non-binding interactions are listed as K_D >> {highest concentration tested}

*Intermediate affinity for M118I was estimated based on limited surface saturation

Table 24. FcRn in solution, K_D for binding pairs [nM]

FcRn	IgG1	IgG2	IgG3	IgG4	albumin
wild-type	380	440	298	225	500
E133K	>> 3000	>> 3000	not tested	not tested	960
W131R	>> 15000	>> 15000	>> 3000	>> 3000	440
M118I	not tested	not tested	210	239	480

Non-binding interactions are listed as K_D >> {highest concentration tested}

EXAMPLE 2: Base editing to alter an FcRn polynucleotide in primary human hepatocytes

Experiments were undertaken to evaluate base editing of an FcRn polynucleotide in primary human hepatocytes (PHHs). The design of the experiment is shown in **FIG. 8A**. Primary human hepatocyte (PHH) co-cultures were contacted with a base editor system containing gRNA1583 or gRNA1582 (see **Tables 2A and 2B**) and an adenosine base editor (ABE). Cells were transfected with 150 ng of the guide RNA and 450 ng of mRNA encoding the base editor, which amounted to a sub-saturating dose (600 ng total) of the base editor system. At ten days post-transfection, base editing efficiencies for each observed amino acid alteration were evaluated for each base editor system (**FIG. 8B**). At 13 days post-transfection, FcRn transcript levels were evaluated using RT-qPCR (**FIG. 8C**). The base editor system containing the guide gRNA1583, whose target amino acid alteration in FcRn was W131(154)R, achieved a total base editing efficiency of over 40% in the primary human hepatocytes. Also, mRNA expression level of FcRn in the primary human hepatocytes base-edited using the guide gRNA1583 were about 70% of that measured for unedited cells.

15

EXAMPLE 3: Optimization of guide polynucleotides for use in base editing to alter an FcRn polynucleotide

Experiments were undertaken to evaluate the impact of spacer length on base editing efficiencies achieved in HEK293T cells using base editing systems targeted to introduce a W131(154)R alteration to an FcRn polypeptide. The following guide polynucleotides containing spacers ranging in length from 19 nt to 23 nt were prepared: gRNA3025 (19 nt spacer), gRNA1583 (20 nt spacer), gRNA3026 (21 nt spacer), gRNA3027 (22 nt spacer), and gRNA3028 (23 nt spacer) (see sequences listed in **Tables 2A and 2B**). The experimental design is shown in **FIG. 9A**. HEK293T cells were transfected at day 1 following seeding with 150 ng of the guide polynucleotide and 450 ng of mRNA encoding an adenosine base editor (ABE). At 72 hours post-transfection, cells were harvested and next-generation sequencing was used to evaluate base editing efficiencies for each base editor system measured for each observed FcRn amino acid alteration (**FIG. 9B**). All spacer lengths evaluated were associated with similar base editing efficiencies. The guide with a spacer length of 19 nucleotides (i.e., gRNA3025) had a base editing efficiency that was slightly better than any of the other guides evaluated that contained longer spacers.

25
30

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
5 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.

10 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

What is claimed is:

- 5 1. A method of altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide, the method comprising contacting the FcRn polynucleotide with a base editor system comprising one or more guide polynucleotides and a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain, or one or more polynucleotides encoding the base editor system, wherein
 - 10 (a) the one or more guide polynucleotides comprises a nucleic acid sequence comprising at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in **Table 2B**; or
 - (b) said one or more guide polynucleotides targets said base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from the group consisting of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130,
 - 15 W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:
 FcRn amino acid sequence
 AESHLSELLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWEKE
 TTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQGTWG
 GDWPEALAIISQRWQQQDKAANKELTFLLFSCPHRLREHLERGRGNLEWKEPPSMRLKARPSSPG
 20 FSVLTCSAFSFPPELQLRFLRNGLAAGTGQGDGFGPNSDGSFHASSSLTVKSGDEHHYCCIVQH
 AGLAQPLRVELES PAKSSVLVVGIVIGVLLLTAAAVGGALLWRRMRSGLPAPWISLRGDDTGVL
 LPTPGEAQDADLDKDVNVIPATA (SEQ ID NO: 530), or a corresponding position in another FcRn polypeptide sequence, thereby altering the nucleobase of the FcRn polynucleotide.
- 25 2. The method of claim 1, wherein the alteration of the nucleobase results in one or more of the following amino acid alterations in the FcRn polypeptide encoded by the FcRn polynucleotide relative to the reference sequence: F110L, F110S, F110P, L112P, N113S, N113D, E115G, E115K, E116G, E116K, E116Q, F117P, M118N, M118V, M118I, M118T, N119G, N119D, N119S, N119C, D121G, L122F, L122A, L122P, T126I, T126S, T126N,
- 30 T126A, W127R, G128S, D130G, D130N, D130H, W131R, W131Q, P132L, P132S, P132P, E133G, A134V, L135P, I137V, I137T.

3. The method of claim 1, wherein the one or more guide polynucleotides target the base editor to effect an alteration of a nucleobase in a codon encoding the amino acid M118 or W131 in the reference sequence.
- 5
4. The method of claim 3, wherein the alteration of the nucleobase results in an amino acid alteration in the FcRn polypeptide encoded by the FcRn polynucleotide selected from the group consisting of M118V, M118I, M118T, W131R, and W131Q.
- 10
5. The method of any one of claims 2-4, wherein the one or more amino acid alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to IgG1, IgG2, IgG3, and/or IgG4.
- 15
6. The method of any one of claims 2-5, wherein the one or more amino acid alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to an Fc region of IgG1, IgG2, IgG3, and/or IgG4.
- 20
7. The method of claim 6, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is greater than 3000 nM.
8. The method of any one of claims 2-7, wherein the FcRn polypeptide encoded by the FcRn polynucleotide comprising an altered nucleobase is capable of binding albumin.
- 25
9. The method of claim 8, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 2000 nM.
- 30
10. The method of claim 8, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 1000 nM.

11. The method of claim 8, wherein binding of the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 500 nM.
- 5 12. The method of any one of any one of claims 1-11, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 20%.
13. The method of any one of any one of claims 1-11, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 40%.
- 10 14. The method of any one of claims 1-13, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 50%.
- 15 15. The method of any one of claims 1-14, wherein the deaminase domain is capable of deaminating cytidine or adenine in DNA.
16. The method of any one of claims 1-15, wherein the deaminase domain is an adenosine deaminase domain or a cytidine deaminase domain.
- 20 17. The method of claim 16, wherein the adenosine deaminase converts a target A•T to G•C in the FcRn polynucleotide.
18. The method of claim 16, wherein the cytidine deaminase converts a target C•G to T•A in the FcRn polynucleotide.
- 25 19. The method of claim 16 or claim 18, wherein the cytidine deaminase domain is an APOBEC deaminase domain or a derivative thereof.
20. The method of any one of claims 1-16, wherein the base editor is a BE4 base editor.
- 30

21. The method of claim 16 or claim 17, wherein the adenosine deaminase domain is a TadA deaminase domain.
22. The method of claim 16, 17, or 21, wherein the deaminase domain is an adenosine
5 deaminase domain.
23. The method of claim 22, wherein the adenosine deaminase is a TadA*8 or Tad*9 variant.
24. The method of claim 23, wherein the adenosine deaminase is a TadA*8.1, TadA*8.2,
10 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.
25. The method of any one of claims 16, 17, or 21-24, wherein the deaminase domain is a
15 monomer or heterodimer.
26. The method of any one of claims 1-25, wherein the napDNAbp domain is Cas9 or Cas12.
27. The method of any one of claims 1-26, wherein the napDNAbp is a nuclease inactive or
20 nickase variant.
28. The method of any one of claims 1-27, wherein the napDNAbp domain comprises a
Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h,
Cas12i, or Cas12j/CasΦ polynucleotide or a functional portion thereof.
25
29. The method of any one of claims 1-28, wherein the napDNAbp domain comprises a dead
Cas9 (dCas9) or a Cas9 nickase (nCas9).
30. The method of any one of claims 1-29, wherein the napDNAbp domain is a
30 Staphylococcus aureus Cas9 (SaCas9), Streptococcus thermophilus 1 Cas9 (St1Cas9), a
Streptococcus pyogenes Cas9 (SpCas9), or variants thereof.

31. The method of claim 30, wherein the napDNAbp domain comprises a variant of SpCas9 or SaCas9 having an altered protospacer-adjacent motif (PAM) specificity.
- 5 32. The method of claim 31, wherein the SpCas9 or SaCas9 has specificity for a PAM sequence selected from the group consisting of NGG, NGA, NGC, NNGRRT, and NNNRRT, where N is any nucleotide and R is A or G.
33. The method of any one of claims 1-32, wherein the napDNAbp domain comprises a
10 nuclease active Cas9.
34. The method of any one of claims 1-33, wherein the base editor further comprises one or more uracil glycosylase inhibitors (UGIs), or wherein the method further comprises expressing a UGI in a cell in trans with the base editor.
- 15 35. The method of any one of claims 1-34, wherein the base editor further comprises one or more nuclear localization signals (NLS).
36. The method of claim 35, wherein the NLS is a bipartite NLS.
- 20 37. The method of any one of claims 1-36, wherein the one or more guide polynucleotides comprise a scaffold comprising one of the following nucleotide sequences:
GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGCUUUU (SpCas9 scaffold; SEQ ID NO: 317) or
25 GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAUGCCGUGUUUAUC
UCGUCAACUUGUUGGCGAGAUUUU (SaCas9 scaffold; SEQ ID NO: 436).
38. The method of any one of claims 1-37, wherein the one or more guide polynucleotides comprise one or more modified nucleotides.
- 30

39. The method of claim 38, wherein the one or more modified polynucleotides are at the 5' terminus and/or the 3' terminus of the one or more guide polynucleotides.
40. The method of claim 38 or claim 39, wherein the one or more modified nucleotides are
5 2'-O-methyl-3'-phosphorothioate nucleotides.
41. The method of any one of claims 1-40, wherein the one or more guide polynucleotides comprise a spacer consisting of from 19 to 23 nucleotides.
- 10 42. The method of claim 41, wherein the one or more guide polynucleotides comprise a spacer consisting of 19 or 20 nucleotides.
43. The method of any one of claims 1-42, wherein the base editor comprises a complex comprising the deaminase domain, the napDNAbp domain, and the guide polynucleotide, or the
15 base editor is a fusion protein comprising the napDNAbp domain fused to the deaminase domain.
44. The method of any one of claims 1-43, wherein the FcRn polynucleotide is in a cell.
- 20 45. The method of claim 44, wherein the cell is a hepatocyte, an endothelial cell, a myeloid cell, or an epithelial cell.
46. The method of any one of claim 44 or claim 45, wherein the cell is *in vivo* or *ex vivo*.
- 25 47. The method of any one of claims 44-46, wherein the cell is in a subject.
48. The method of claim 47, wherein the subject is a mammal.
49. The method of claim 48, wherein the mammal is a human.
- 30 50. A cell produced by the method of any one of claims 1-49.

51. A base editor system for altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide, the base editor system comprising: (i) one or more guide polynucleotides, or one or more polynucleotides encoding the one or more guide
 5 polynucleotides, and (ii) a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain, or one or more polynucleotides encoding the base editor, wherein

(a) the one or more guide polynucleotides comprises a nucleic acid sequence comprising at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in **Table 2B**; or

10 (b) said one or more guide polynucleotides targets said base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from the group consisting of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130, W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:

FcRn amino acid sequence

15 AESHLSELLYHLTAVSSPAPGTPAFWVSGWLGPQQYLSYNSLRGEAEPCGAWVWENQVSWYWEKE
 TTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQGTWG
 GDWPEALAI SQRWQQQDKAANKELTFLLEFSCPHRLREHLERGRGNLEWKEPPSMRLKARPSSPG
 FSVLTCSAFSFPPELQLRFLRNGLAAGTGQGDFGPNSDGSFHASSSLTVKSGDEHHYCCIVQH
 AGLAQPLRVELES PAKSSVLVVGIVIGVLLLTAAAVGGALLWRRMRSGLPAPWISLRGDDTGVL
 20 LPTPGEAQDADLKD VNVIPATA (SEQ ID NO: 530), or a corresponding position in another
 FcRn polypeptide sequence.

52. The base editor system of claim 51, wherein the alteration of the nucleobase results in one or more of the following amino acid alterations in the FcRn polypeptide encoded by the
 25 FcRn polynucleotide relative to the reference sequence: F110L, F110S, F110P, L112P, N113S, N113D, E115G, E115K, E116G, E116K, E116Q, F117P, M118N, M118V, M118I, M118T, N119G, N119D, N119S, N119C, D121G, L122F, L122A, L122P, T126I, T126S, T126N, T126A, W127R, G128S, D130G, D130N, D130H, W131R, W131Q, P132L, P132S, P132P, E133G, A134V, L135P, I137V, I137T.

30

53. The base editor system of claim 51, wherein the one or more guide polynucleotides target the base editor to effect an alteration of a nucleobase in a codon encoding the amino acid M118 or W131 in the reference sequence.
- 5 54. The base editor system of claim 53, wherein the alteration of the nucleobase results in an amino acid alteration in the FcRn polypeptide encoded by the FcRn polynucleotide selected from the group consisting of M118V, M118V, M118I, M118T, W131R, and W131Q.
55. The base editor system of any one of claims 52-54, wherein the one or more amino acid
10 alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to IgG1, IgG2, IgG3, and/or IgG4.
56. The base editor system of any one of claims 52-55, wherein the one or more amino acid
15 alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to an Fc region of IgG1, IgG2, IgG3, and/or IgG4.
57. The base editor system of claim 56, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is greater than 3000 nM.
20
58. The base editor system of any one of claims 52-57, wherein the FcRn polypeptide encoded by the FcRn polynucleotide comprising an altered nucleobase is capable of binding albumin.
- 25 59. The base editor system of claim 58, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 2000 nM.
60. The base editor system of claim 58, wherein the FcRn polypeptide comprising the one or
30 more amino acid alterations has a K_D in solution for binding with albumin that is less than 1000 nM.

61. The base editor system of claim 58, wherein binding of the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 500 nM.
- 5
62. The base editor system of any one of claims 51-61, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 20%.
63. The base editor system of any one of claims 51-62, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 40%.
- 10
64. The base editor system of any one of claims 51-63, wherein the nucleobase of the FcRn polynucleotide is altered with a base editing efficiency of at least about 50%.
65. The base editor system of any one of claims 51-64, wherein the deaminase domain is capable of deaminating cytidine or adenine in DNA.
- 15
66. The base editor system of any one of claims 51-8, wherein the deaminase domain is an adenosine deaminase domain or a cytidine deaminase domain.
- 20
67. The base editor system of claim 66, wherein the adenosine deaminase converts a target A•T to G•C in the FcRn polynucleotide.
68. The base editor system of claim 66, wherein the cytidine deaminase converts a target C•G to T•A in the FcRn polynucleotide.
- 25
69. The base editor system of claim 66, wherein the cytidine deaminase domain is an APOBEC deaminase domain or a derivative thereof.
70. The base editor system of any one of claims 51-66, wherein the base editor is a BE4 base editor.
- 30

71. The base editor system of claim 66 or claim 67, wherein the adenosine deaminase domain is a TadA deaminase domain.
- 5 72. The base editor system of any one of claims 51-67 or claim 71, wherein the deaminase domain is an adenosine deaminase domain.
73. The base editor system of claim 72, wherein the adenosine deaminase is a TadA*8 or Tad*9 variant.
- 10 74. The base editor system of claim 72 or claim 73, wherein the adenosine deaminase is a 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, 15 TadA*8.23, or TadA*8.24.
75. The base editor system of any one of claims 72-74, wherein the deaminase domain is a monomer or heterodimer.
- 20 76. The base editor system of any one of claims 51-75, wherein the napDNAbp domain is Cas9 or Cas12.
77. The base editor system of any one of claims 51-76, wherein the napDNAbp domain is a nuclease inactive or nickase variant.
- 25 78. The base editor system of any one of claims 51-77, wherein the napDNAbp domain comprises a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ polynucleotide or a functional portion thereof.
- 30 79. The base editor system of any one of claims 51-78, wherein the napDNAbp domain comprises a dead Cas9 (dCas9) or a Cas9 nickase (nCas9).

80. The base editor system of any one of claims 51-79, wherein the napDNAbp domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

5

81. The base editor system of claim 51-80, wherein the napDNAbp domain comprises a variant of SpCas9 or SaCas9 having an altered protospacer-adjacent motif (PAM) specificity.

82. The base editor system of claim 81, wherein the SpCas9 or SaCas9 has specificity for a PAM sequence selected from the group consisting of NGG, NGA, NGC, NNGRRT, and NNNRRT, where N is any nucleotide and R is A or G.

10

83. The base editor system of any one of claims 51-82, wherein the napDNAbp domain comprises a nuclease active Cas9.

15

84. The base editor system of any one of claims 51-83, wherein the base editor further comprises one or more uracil glycosylase inhibitors (UGIs), or wherein the base editor system further comprises a UGI in trans with the base editor.

20 85. The base editor system of any one of claims 51-84, wherein the base editor further comprises one or more nuclear localization signals (NLS).

86. The base editor system of claim 85, wherein the NLS is a bipartite NLS.

25 87. The base editor system of any one of claims 51-86, wherein the one or more guide polynucleotides comprise a scaffold comprising one of the following nucleotide sequences:
 GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
 CCGAGUCGGUGCUUUU (SpCas9 scaffold; SEQ ID NO: 317) or
 GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAUGCCGUGUUUAUC
 30 UCGUCAACUUGUUGGCGAGAUUUU (SaCas9 scaffold; SEQ ID NO: 436).

88. The base editor system of any one of claims 51-88, wherein the one or more guide polynucleotides comprise one or more modified nucleotides.
89. The base editor system of claim 88, wherein the one or more modified polynucleotides
5 are at the 5' terminus and/or the 3' terminus of the one or more guide polynucleotides.
90. The base editor system of claim 88 or claim 89, wherein the one or more modified nucleotides are 2'-O-methyl-3'-phosphorothioate nucleotides.
- 10 91. The base editor system of any one of claims 88-90, wherein the one or more guide polynucleotides comprise a spacer consisting of from 19 to 23 nucleotides.
92. The base editor system of claim 91, wherein the one or more guide polynucleotides comprise a spacer consisting of 19 or 20 nucleotides.
15
93. The base editor system of any one of claims 51-92, wherein the base editor comprises a complex comprising the deaminase domain, the napDNAbp domain, and the one or more guide polynucleotides, or the base editor is a fusion protein comprising the napDNAbp domain fused to the deaminase domain.
20
94. A polynucleotide encoding the base editor system of any one of claims 51-93.
95. A vector comprising the polynucleotide of claim 94.
- 25 96. The vector of claim 95, wherein the vector comprises a lipid nanoparticle.
97. The vector of claim 96, wherein the lipid nanoparticle comprises a lipid monolayer comprising a lipid selected from the group consisting of lecithin, phosphatidylcholines, phosphatidic acid, phosphatidylethanolamines, phosphatidylglycerols, phosphatidylserines,
30 phosphatidylinositols, cardiolipins, lipid-polyethyleneglycol conjugates, and combinations thereof.

98. The vector of claim 97, wherein the lipid monolayer comprises a PEGylated lipid.
99. The vector of claim 97 or claim 98, wherein the lipid monolayer further comprises a
5 cholesterol.
100. The vector of any one of claims 96-99, wherein the lipid nanoparticle comprises an ionizable cationic lipid selected from the group consisting of: N-methyl-N-(2-(arginoylamino) ethyl)- N, N- Di octadecyl aminium chloride or di stearyl arginyl ammonium chloride]
- 10 (DSAA); N,N-di-myristoyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride (DMGLA); N,N-dimyristoyl-N-methyl-N-2[N2-guanidino-L- lysiny] aminoethyl ammonium chloride; N,N-dimyristoyl-N-methyl-N-2[N'-(N2, N6- di-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-di-stearyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-dioleoyl-N,N-dimethylammonium chloride (DODAC); N-
15 (2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTAP); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTMA); N,N-distearyl- N,N-dimethylammonium bromide (DDAB); 3-(N-(N',N'-dimethylaminoethane)- carbamoyl) cholesterol (DC-Choi); N-(1,2-dimyristyloxyprop-3-yl)-N,N- dimethyl-N-hydroxyethyl ammonium bromide (DMRIE); 1,3-dioleoyl-3- trimethylammonium-propane, N-(1-(2,3-
20 dioleoyloxy)propyl)-N-(2- (sperminecarboxamido)ethyl)-N,N-dimethy- 1 ammonium trifluoroacetate (DOSPA); GAP-DLRIE; DMDHP; 3-p[4N-(H8N-diguanidino spermidine)-carbamoyl] cholesterol (BGSC); 3-P[N,N-diguanidinoethyl-aminoethane)-carbamoyl] cholesterol (BGTC); N,N\N2,N3 Tetra-methyltetrapalmitylspermine (cellfectin); N-t-butyl-N'- tetradecyl-3-tetradecyl-aminopropion-amidine (CLONfectin); dimethyldioctadecyl ammonium bromide (DDAB); 1,3-
25 dioleoyloxy-2-(6-carboxyspermyl)-propyl amide (DOSPER); 4-(2,3-bis-palmitoyloxy-propyl)-1-methyl- 1H-imidazole (DPIM) N,N,N',N'-tetramethyl-N,N'-bis(2-hydroxyethyl)-2,3 dioleoyloxy- 1 ,4- butanediammonium iodide) (Tfx-50); 1,2 dioleoyl-3-(4'-trimethylammonio) butanol-sn- glycerol (DOBT); cholesteryl (4'trimethylammonia) butanoate (ChOTB) where the trimethylammonium group is connected via a butanol spacer arm to either the double chain (for
30 DOTB) or cholesteryl group (for ChOTB); DL-1,2-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORI); DL-1,2-0-dioleoyl-3- dimethylaminopropyl-P-

hydroxyethylammonium (DORIE); 1,2-dioleoyl-3-succinyl-sn-glycerol choline ester (DOSC);
 cholesteryl hemisuccinate ester (ChOSC); dioctadecylamidoglycylspermine (DOGS);
 dipalmitoyl phosphatidylethanolamylspermine (DPPES); cholesteryl-3P- carboxyl-amido-
 ethylenetrimethylammonium iodide; 1-dimethylamino-3- trimethylammonio-DL-2-propyl-
 5 cholesteryl carboxylate iodide; cholesteryl-3-β- carboxyamidoethyleneamine; cholesteryl-3-P-
 oxysuccinamido- ethylenetrimethylammonium iodide; 1-dimethylamino-3-trimethylammonio-
 DL-2- propyl-cholesteryl-3-P-oxysuccinate iodide; 2-(2-trimethylammonio)- ethylmethylamino
 ethyl-cholesteryl-3-P-oxysuccinate iodide; 3-β-N- (polyethyleneimine)-carbamoylesterol,
 DC-cholesterol; N4-cholesteryl-spermine HCl salt (GL67); N1-[2-((1 S)-1-[(3-
 10 aminopropyl)amino]-4-[di(3-amino-propyl)amino]butylcarboxamido)ethyl]-3,4-di[oleoyloxy]-
 benzamide (MVL5); and combinations thereof.

101. The vector of claim 95, wherein the vector comprises a polymer nanoparticle.
- 15 102. The vector of claim 95, wherein the vector is a viral vector.
103. The vector of claim 102, wherein the viral vector is a retroviral vector or an adeno-
 associated virus vector.
- 20 104. A cell comprising the polynucleotide of claim 94 or the vector of any one of claims 95-
 103.
105. The cell of claim 104, wherein the cell is a hepatocyte, an endothelial cell, a myeloid cell,
 or an epithelial cell.
- 25 106. The cell of claim 104 or claim 105, wherein the cell is a mammalian cell.
107. The cell of claim 106, wherein the cell is a human cell.

108. A composition comprising the base editor system of any one of claims 51-93, the polynucleotide of claim 94, the vector of any one of claims 95-103, or the cell of any one of claims 104-107.

5 109. A pharmaceutical composition comprising the composition of claim 108 and a pharmaceutically acceptable excipient.

110. A method of treating an autoimmune disorder mediated by immunoglobulin G in a subject in need thereof, the method comprising altering a nucleobase of an FcRn polynucleotide
10 in the subject by administering to the subject a base editor system comprising one or more guide polynucleotides and a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain and a deaminase domain, or one or more polynucleotides encoding the base editor system, wherein:

(a) the one or more guide polynucleotides comprises a nucleic acid sequence comprising
15 at least 10-23 contiguous nucleotides of a spacer nucleic acid sequence listed in **Table 2B**; or

(b) said one or more guide polynucleotides targets said base editor to effect an alteration of a nucleobase in a codon encoding an amino acid residue selected from the group consisting of F110, L112, N113, E115, E116, F117, M118, N119, D121, L122, T126, W127, G128, D130, W131, P132, E133, A134, L135, and I137 relative to the following reference sequence:

20 FcRn amino acid sequence

AESHLSELLYHLTAVSSPAPGTPAFVWSGWLGPQQYLSYNSLRGEAEPGAWVWENQVSWYWEKE
TTDLRIKEKLFLEAFKALGGKGPYTLQGLLGCELGPDNTSVPTAKFALNGEEFMNFDLKQGTWG
GDWPEALAISSQRWQQQDKAANKELTFLLFSCPHRLREHLERGRGNLEWKEPPSMRLKARPSSPG
FSVLTCSAFSFPPELQLRFLRNGLAAGTGQGDGFGPNSDGSFHASSSLTVKSGDEHHYCCIVQH
25 AGLAQPLRVELESPAKSSVLVVGIVIGVLLLTAAAVGGALLWRRMRSGLPAPWISLRGDDTGVL
LPTPGEAQDADLKDVNVIPATA (SEQ ID NO: 436), or a corresponding position in another FcRn polypeptide sequence, thereby treating the autoimmune disorder.

111. The method of claim 110, wherein the alteration of the nucleobase results in one or more
30 of the following amino acid alterations in the FcRn polypeptide encoded by the FcRn polynucleotide relative to the reference sequence: F110L, F110S, F110P, L112P, N113S,

N113D, .E115G, E115K, E116G, E116K, E116Q, F117P, M118N, M118V, M118I, M118T, N119G, N119D, N119S, N119C, D121G, L122F, L122A, L122P, T126I, T126S, T126N, T126A, W127R, G128S, D130G, D130N, D130H, W131R, W131Q, P132L, P132S, P132P, E133G, A134V, L135P, I137V, I137T.

5

112. The method of claim 110, wherein the one or more guide polynucleotides target the base editor to effect an alteration of a nucleobase in a codon encoding the amino acid M118 or W131 in the reference sequence.

10 113. The method of claim 112, wherein the alteration of the nucleobase results in an amino acid alteration in the FcRn polypeptide encoded by the FcRn polynucleotide selected from the group consisting of M118V, M118V, M118I, M118T, W131R, and W131Q.

114. The method of any one of claims 111-113, wherein the one or more amino acid
15 alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to IgG1, IgG2, IgG3, and/or IgG4.

115. The method of any one of claims 111-113, wherein the one or more amino acid
20 alterations in the FcRn polypeptide reduce or eliminate binding of the FcRn polypeptide to an Fc region of IgG1, IgG2, IgG3, and/or IgG4.

116. The method of claim 115, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with IgG1, IgG2, IgG3, and/or IgG4 that is greater than 3000 nM.

25

117. The method of any one of claims 111-116, wherein the FcRn polypeptide encoded by the FcRn polynucleotide comprising an altered nucleobase is capable of binding albumin.

118. The method of claim 117, wherein the FcRn polypeptide comprising the one or more
30 amino acid alterations has a K_D in solution for binding with albumin that is less than 2000 nM.

119. The method of claim 117, wherein the FcRn polypeptide comprising the one or more amino acid alterations has a K_D in solution for binding with albumin that is less than 1000 nM.

120. The method of claim 117, wherein binding of the FcRn polypeptide comprising the one
5 or more amino acid alterations has a K_D in solution for binding with albumin that is less than 500 nM.

121. The method of any one of claims 110-120, wherein the method comprises decreasing
10 levels of immunoglobulin G polypeptides in the subject by at least about 25%.

122. The method of any one of claims 110-121, wherein the method comprises decreasing
levels of immunoglobulin G polypeptides in the subject by at least about 50%.

123. The method of any one of claims 110-122, wherein the method comprises decreasing
15 levels of immunoglobulin G polypeptides in the subject by at least about 70%.

124. The method of any one of claims 110-123, wherein the nucleobase of the FcRn
polynucleotide is altered with a base editing efficiency of at least about 20%.

20 125. The method of any one of claims 110-124, wherein the nucleobase of the FcRn
polynucleotide is altered with a base editing efficiency of at least about 40%.

126. The method of any one of claims 110-125, wherein the nucleobase of the FcRn
polynucleotide is altered with a base editing efficiency of at least about 50%.

25 127. The method of any one of claims 110-126, wherein the disorder is selected from the
group consisting of myasthenia gravis (gMG), warm autoimmune hemolytic anemia (wAIHA),
idiopathic thrombocytopenia purpura (ITP), Grave's disease, chronic inflammatory
demyelinating polyneuropathy (CIDP), pemphigus vulgaris, and hemolytic diseases of fetus and
30 newborn (HDFN).

128. The method of any one of claims 110-127, wherein the deaminase domain is capable of deaminating cytidine or adenine in DNA.
129. The method of any one of claims 110-128, wherein the deaminase domain is an
5 adenosine deaminase domain or a cytidine deaminase domain.
130. The method of claim 129, wherein the adenosine deaminase converts a target A•T to G•C in the FcRn polynucleotide.
- 10 131. The method of claim 129, wherein the cytidine deaminase converts a target C•G to T•A in the FcRn polynucleotide.
132. The method of claim 129 or claim 131, wherein the cytidine deaminase domain is an APOBEC deaminase domain or a derivative thereof.
- 15 133. The method of any one of claims 110-129, claim 131, or claim 132, wherein the base editor is a BE4 base editor.
134. The method of claim 129 or claim 130, wherein the adenosine deaminase domain is a
20 TadA deaminase domain.
135. The method of any one of claims 129, 130, or 134, wherein the deaminase domain is an adenosine deaminase domain.
- 25 136. The method of claim 135, wherein the adenosine deaminase is a TadA*8 or Tad*9 variant.
137. The method of claim 135 or claim 136, wherein the adenosine deaminase is a 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,
30 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.

138. The method of any one of claims 110-137, wherein the deaminase domain is a monomer
5 or heterodimer.

139. The method of any one of claims 110-138, wherein the napDNAbp domain is Cas9 or Cas12.

10 140. The method of any one of claims 110-139, wherein the napDNAbp domain is a nuclease inactive or nickase variant.

141. The method of any one of claims 110-140, wherein the napDNAbp domain comprises a Cas9, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h,
15 Cas12i, or Cas12j/CasΦ polynucleotide or a functional portion thereof.

142. The method of any one of claims 110-141, wherein the napDNAbp domain comprises a dead Cas9 (dCas9) or a Cas9 nickase (nCas9).

20 143. The method of any one of claims 110-142, wherein the napDNAbp domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

144. The method of claim 143, wherein the napDNAbp domain comprises a variant of SpCas9
25 or SaCas9 having an altered protospacer-adjacent motif (PAM) specificity.

145. The method of claim 144, wherein the SpCas9 or SaCas9 has specificity for a PAM sequence selected from the group consisting of NGG, NGA, NGC, NNGRRT, and NNNRRT, where N is any nucleotide and R is A or G.

30

146. The method of any one of claims 110-145, wherein the napDNAbp domain comprises a nuclease active Cas9.

147. The method of any one of claims 110-146, wherein the base editor further comprises one or more uracil glycosylase inhibitors (UGIs) , or wherein the method further comprises expressing a UGI in a cell in trans with the base editor.

148. The method of any one of claims 110-147, wherein the base editor further comprises one or more nuclear localization signals (NLS).

10

149. The method of claim 148, wherein the NLS is a bipartite NLS.

150. The method of any one of claims 110-149, wherein the one or more guide polynucleotides comprise a scaffold comprising one of the following nucleotide sequences:
15 GUUUUAGAGCUAGAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGCUUUU (SpCas9 scaffold; SEQ ID NO: 317) or
GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAUGCCGUGUUUAUC
UCGUCAACUUGUUGGCGAGAUUUU (SaCas9 scaffold; SEQ ID NO: 436).

20 151. The method of any one of claims 110-150, wherein the one or more guide polynucleotides comprise one or more modified nucleotides.

152. The method of claim 151, wherein the one or more modified polynucleotides are at the 5' terminus and/or the 3' terminus of the one or more guide polynucleotides.

25

153. The method of claim 151 or claim 152, wherein the one or more modified nucleotides are 2'-O-methyl-3'-phosphorothioate nucleotides.

154. The method of any one of claims 110-153, wherein the one or more guide polynucleotides comprise a spacer consisting of from 19 to 23 nucleotides.

30

155. The method of claim 154, wherein the one or more guide polynucleotides comprise a spacer consisting of 19 or 20 nucleotides.
156. The method of any one of claims 110-155, wherein the base editor comprises a complex comprising the deaminase domain, the napDNAbp domain, and the one or more guide polynucleotides, or the base editor is a fusion protein comprising the napDNAbp domain fused to the deaminase domain.
157. The method of any one of claims 110-156, wherein the administration is local administration.
158. The method of any one of claims 110-157, wherein the administration is systemic administration.
159. The method of any one of claims 110-158, wherein the base editor system is administered to the subject using a vector.
160. The method of claim 159, wherein the vector is a lipid nanoparticle.
161. The method of claim 159 or claim 160, wherein the vector targets the liver.
162. The method of any one of claims 110-161, wherein the subject is a mammal.
163. The method of claim 162, wherein the mammal is a human.
164. A kit suitable for use in the method of any one of the above claims and comprising a guide polynucleotide comprising a sequence listed in **Table 2A** or **Table 2B**.
165. A method of altering a nucleobase of a Fc fragment of IgG receptor and transporter (FcRn) polynucleotide, the method comprising contacting the FcRn polynucleotide with a base editor system comprising one or more guide polynucleotides selected from the group consisting

of gRNA1583, gRNA1578, gRNA3265, or one or more polynucleotides encoding the same, and a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain and an adenosine deaminase domain, or one or more polynucleotides encoding the base editor; thereby altering the nucleobase of the FcRn polynucleotide.

5

166. A base editor system comprising one or more guide polynucleotides selected from the group consisting of gRNA1583, gRNA1578, gRNA3265, or one or more polynucleotides encoding the same, and a base editor comprising a nucleic acid programmable DNA binding protein (napDNAbp) domain and an adenosine deaminase domain, or one or more
10 polynucleotides encoding the base editor.

167. A guide polynucleotide comprising a sequence listed in **Table 2A** or **Table 2B**.

168. A method of modifying a neonatal fragment crystallizable receptor (FcRn) protein in a
15 mammalian cell, the method comprising contacting the cell with a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of an *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene.

20

169. The method according to claim 168, wherein the genome editor comprises a base editor or a prime editor.

170. A method of treating an IgG-mediated autoimmune disorder in a subject in need thereof,
25 the method comprising modifying neonatal fragment crystallizable receptor (FcRn) protein in a mammalian cell of the subject.

171. The method according to claim 170, wherein modifying the FcRn protein comprises genome editing an *FCGRT* gene in the mammalian cell of the subject.

30

172. The method according to claim 171, wherein the genome editing comprises contacting the mammalian cell with a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of the *FCGRT* gene and targets the genome editor to effect a modification in the *FCGRT* gene in the cell, wherein the modification
5 alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene.

173. The method according to claim 172, wherein the genome editor comprises a base editor or a prime editor.

10 174. The method according to claim 169 or claim 173, wherein the base editor or prime editor is delivered to the mammalian cell via a nanoparticle, a viral vector, or electroporation.

175. The method according to claim 174, wherein the nanoparticle is a gold nanoparticle, a lipid nanoparticle, or a polymer nanoparticle.

15 176. The method according to claim 174, wherein the viral vector is selected from a retrovirus, an adenovirus, an adeno-associated virus (AAV), a herpesvirus, or a sendai virus.

177. The method according to any of the preceding claims, wherein the modified FcRn protein
20 comprises one or more single nucleotide modifications or changes.

178. The method according to any of the preceding claims, wherein the modified FcRn exhibits reduced ability to bind to an Fc region of an IgG antibody.

25 179. The method according to any of the preceding claims, wherein the mammalian cell is a human cell.

180. The method according to any of the preceding claims, wherein the mammalian cell is *ex vivo*, *in vivo*, or *in vitro*.

30

181. The method according to any of claims 168-180, wherein the contacted cell expresses a variant FcRn protein comprising at least one amino acid alteration relative to a reference FcRn protein.
- 5 182. The method according to any of claims 168-169 or 172-181, wherein the genome editor is a base editor comprising a nucleic acid programmable DNA binding domain and a cytidine deaminase domain that converts a target C-G to T-A or a target G-C to A-T in the *FCGRT* gene.
- 10 183. The method according to any of claims 168-169 or 172-181, wherein the genome editor is a base editor comprising a nucleic acid programmable DNA binding domain and an adenosine deaminase domain that converts a target A-T to G-C or a target T-A to C-G in the *FCGRT* gene.
- 15 184. The method according to claim 182 or claim 183, wherein the nucleic acid programmable DNA binding domain comprises a catalytically inactivated (dead) Cas9 (dCas9) or a Cas9 nickase (nCas9).
- 20 185. The method according to any of claims 168-169 or 172-181, wherein the genome editor is a prime editor comprising a nucleic acid programmable DNA binding domain and a reverse transcriptase and the guide RNA is a prime editing guide RNA (pegRNA), wherein the prime editor replaces one or more nucleotides in the *FCGRT* gene with a different nucleotide.
- 25 186. The method according to claim 185, wherein the nucleic acid programmable DNA binding domain comprises a catalytically inactivated (dead) Cas9 (dCas9) or a Cas9 nickase (nCas9).
- 30 187. The method according to any of the preceding claims, wherein the modified FcRn protein differs from a reference FcRn protein at one or more amino acids selected from the group consisting of: leucine (L) at position 112, glutamic acid (E) at position 115, glutamic acid (E) at position 116, tryptophan (W) at position 131, proline (P) at position 132, and glutamic acid (E) at position 133.

188. The method according to any of the preceding claims, wherein the modified FcRn protein comprises one or more mutations as set forth in Fig. 3.

189. The method according to any of claims 168-169 or 172-188, wherein the guide RNA and
5 the genome editor are conjugated to a targeting moiety that binds to FcRn or albumin.

190. The method according to claim 189, wherein the targeting moiety is selected from the group consisting of an Fc domain of IgG, an antibody that specifically binds FcRn, an antibody that specifically binds albumin, a peptide that binds albumin, albumin, or a fragment or
10 derivative thereof.

191. A composition comprising a guide RNA and a genome editor, wherein the guide RNA comprises a nucleotide sequence that is complementary to a portion of the *FCGRT* gene and targets the base genome editor to effect a modification in the *FCGRT* gene in the cell, wherein
15 the modification alters the amino acid sequence of the FcRn protein encoded by the *FCGRT* gene.

192. The composition according to claim 191, further comprising a delivery vehicle comprising a targeting moiety that binds to FcRn or albumin.
20

193. The method according to claim 192, wherein the targeting moiety is selected from the group consisting of an Fc domain of IgG, an antibody that specifically binds FcRn, an antibody that specifically binds albumin, a peptide that binds albumin, albumin, or a fragment or derivative thereof.
25

194. A lipid nanoparticle (LNP) comprising:
a lipid monolayer membrane comprising at least one fragment crystallizable (Fc) region of an IgG antibody or a functional fragment thereof embedded therein; and
a lipid core matrix enclosed in the lipid monolayer membrane.
30

195. The LNP of claim 194, wherein the lipid core matrix comprises at least one nucleic acid.

196. The LNP of claim 195, wherein the nucleic acid is selected from the group consisting of DNA or RNA.
- 5 197. The LNP of claim 196, wherein the RNA is an siRNA or a guide RNA.
198. The LNP of claim 197, wherein the siRNA or guide RNA modifies or silences an *FCGRT* gene.
- 10 199. The LNP of any of claims 194-198, wherein the lipid monolayer membrane is comprised of a lipid selected from the group consisting of lecithin, phosphatidylcholines, phosphatidic acid, phosphatidylethanolamines, phosphatidylglycerols, phosphatidylserines, phosphatidylinositols, cardiolipins, lipid-polyethyleneglycol conjugates, and combinations thereof.
- 15 200. The LNP of claim 199, wherein at least a portion of the lipids of the lipid monolayer membrane is PEGylated.
201. The LNP of claim 199, wherein the lipid monolayer further comprises cholesterol.
- 20 202. The LNP of any of claims 194-201, wherein the lipid core matrix comprises an ionizable cationic lipid selected from the group consisting of: N-methyl-N-(2-(arginoylamino) ethyl)- N, N- Di octadecyl aminium chloride or di stearoyl arginyl ammonium chloride] (DSAA); N,N-dimyristoyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride (DMGLA); N,N-dimyristoyl-N-methyl-N-2[N2-guanidino-L- lysiny] aminoethyl ammonium chloride; N,N-dimyristoyl-N-methyl-N-2[N'-(N2, N6- di-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-di-stearoyl-N-methyl-N-2[N'-(N6-guanidino-L-lysiny)] aminoethyl ammonium chloride; N,N-dioleyl-N,N-dimethylammonium chloride (DODAC); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTAP); N-(2,3- dioleoyloxy) propyl)-N,N,N-trimethylammonium chloride (DOTMA); N,N-distearyl- N,N- dimethylammonium bromide (DDAB); 3-(N-(N',N'-dimethylaminoethane)- carbamoyl) cholesterol (DC-Choi); N-(1,2-dimyristyloxyprop-3-yl)-N,N- dimethyl-N-hydroxyethyl
- 25
- 30

- ammonium bromide (DMRIE); 1,3-dioleoyl-3- trimethylammonium-propane, N-(1-(2,3-dioleoyloxy)propyl)-N-(2- (sperminecarboxamido)ethyl)-N,N-dimethy- 1 ammonium trifluoroacetate (DOSPA); GAP-DLRIE; DMDHP; 3-p[4N-(H8N-diguanidino spermidine)-carbamoyl] cholesterol (BGSC); 3-P[N,N-diguanidinoethyl-aminoethane)-carbamoyl] cholesterol (BGTC);
- 5 N,N,N₂,N₃ Tetra-methyltetrapalmitylspermine (cellfectin); N-t-butyl-N'- tetradecyl-3-tetradecyl-aminopropion-amidine (CLONfectin); dimethyldioctadecyl ammonium bromide (DDAB); 1,3-dioleoyloxy-2-(6-carboxyspermyl)-propyl amide (DOSPER); 4-(2,3-bis-palmitoyloxy-propyl)-1-methyl- 1H-imidazole (DPIM) N,N,N',N'-tetramethyl-N,N'-bis(2-hydroxyethyl)-2,3-dioleoyloxy- 1, 4- butanediammonium iodide) (Tfx-50); 1,2 dioleoyl-3-(4'-trimethylammonio) butanol-sn- glycerol (DOBT); cholesteryl (4'-trimethylammonia) butanoate (ChOTB) where the trimethylammonium group is connected via a butanol spacer arm to either the double chain (for DOTB) or cholesteryl group (for ChOTB); DL-1,2-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORI); DL-1,2-0-dioleoyl-3- dimethylaminopropyl-P-hydroxyethylammonium (DORIE); 1,2-dioleoyl-3-succinyl-sn-glycerol choline ester (DOSC);
- 15 cholesteryl hemisuccinate ester (ChOSC); dioctadecylamidoglycylspermine (DOGS); dipalmitoyl phosphatidylethanolamylspermine (DPPES); cholesteryl-3P- carboxyl-amido-ethylenetrimethylammonium iodide; 1-dimethylamino-3- trimethylammonio-DL-2-propyl-cholesteryl carboxylate iodide; cholesteryl-3-β- carboxyamidoethyleneamine; cholesteryl-3-P-oxysuccinamido- ethylenetrimethylammonium iodide; 1-dimethylamino-3-trimethylammonio-
- 20 DL-2- propyl-cholesteryl-3-P-oxysuccinate iodide; 2-(2-trimethylammonio)- ethylmethylamino ethyl-cholesteryl-3-P-oxysuccinate iodide; 3-β-N- (polyethyleneimine)-carbamoylcholesterol, DC-cholesterol; N4-cholesteryl-spermine HCl salt (GL67); N1-[2-((1 S)-1-[(3-aminopropyl)amino]-4-[di(3-amino-propyl)amino]butylcarboxamido)ethyl]-3,4-di[oleoyloxy]-benzamide (MVL5); and combinations thereof.
- 25
203. A pharmaceutical composition comprising:
 at least one LNP according to any of claims 194-202; and
 at least one pharmaceutically-acceptable excipient.
- 30 204. A method of treating an IgG-mediated autoimmune disorder in a subject in need thereof, the method comprising administering to the subject the LNP according to any of claims 194-202.

205. A method of silencing expression or modifying a genomic sequence encoding FcRn in a cell, the method comprising contacting the cell with the LNP according to any of claims 194-202.

5

206. A method or composition according to any of the preceding claims, wherein modification of FcRn does not interfere with albumin half-life.