



(51) International Patent Classification:

C12N 7/04 (2006.01) *C12N 9/22* (2006.01)
A61K 31/7125 (2006.01) *C12N 9/78* (2006.01)
A61P 31/20 (2006.01) *C12N 15/113* (2010.01)
C07K 14/005 (2006.01) *C12N 15/90* (2006.01)

(21) International Application Number:

PCT/US2022/077105

(22) International Filing Date:

27 September 2022 (27.09.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/248,938 27 September 2021 (27.09.2021) US
 63/357,623 30 June 2022 (30.06.2022) US
 63/371,634 16 August 2022 (16.08.2022) US

(71) Applicants: **BEAM THERAPEUTICS INC.** [US/US];
 238 Main Street, 9th Floor, Cambridge, Massachusetts

02142 (US). **INSTITUT NATIONAL DE LA SANTÉ
 ET DE LA RECHERCHE MÉDICALE (INSERM)**
 [FR/FR]; 101, rue de Tolbiac, 75654 Paris cedex 13 (FR).

(72) Inventors; and

(71) Applicants: **ZOULIM, Fabien** [FR/FR]; c/o Inserm, 101, rue de Tolbiac, 75654 Paris cedex 13 (FR). **MARTINEZ, Maria, Guadalupe** [FR/FR]; c/o Inserm, 101, rue de Tolbiac, 75654 Paris cedex 13 (FR). **TESTONI, Barbara** [FR/FR]; c/o Inserm, 101, rue de Tolbiac, 75654 Paris cedex 13 (FR).

(72) Inventors: **SMEKALOVA, Elena**; c/o Beam Therapeutics Inc., 238 Main Street, 9th Floor, Cambridge, Massachusetts 02142 (US). **BARRERA, Luis**; c/o Beam Therapeutics Inc., 238 Main Street, 9th Floor, Cambridge, Massachusetts 02142 (US).

(74) Agent: **HUNTER-ENSOR, Melissa**; c/o Greenberg Traurig, LLP, One International Place, Boston, Massachusetts 02110 (US).

(54) Title: COMPOSITIONS AND METHODS FOR TREATING HEPATITIS B VIRUS INFECTION

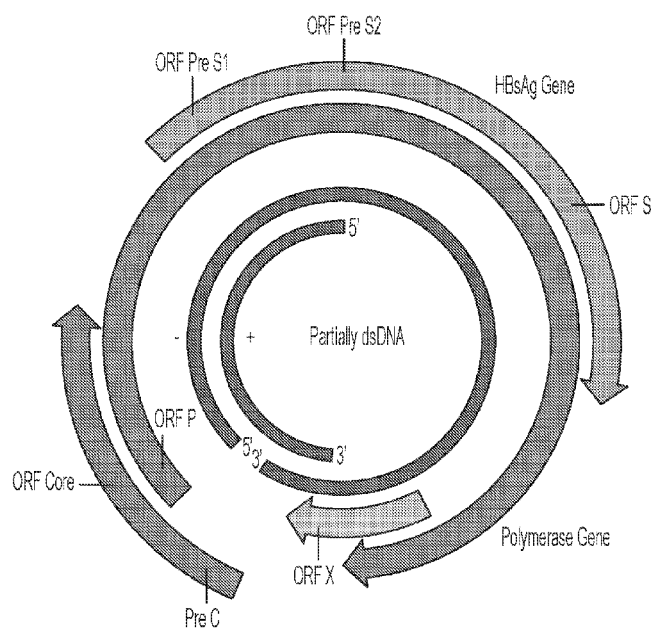


FIG. 1

(57) **Abstract:** Hepatitis B is a serious liver infection caused by the hepatitis B virus (HBV). Current therapeutic approaches to HBV infection have severe limitations. Antiviral medications, e.g., tenofovir, a nucleotide reverse transcriptase inhibitor, can decrease viral replication but do not cure HBV infected patients. Due to the extent of liver damage caused by HBV, a transplant becomes necessary in some cases. In addition to the risks inherent in organ transplants, the cost can be prohibitive. Therefore, improved methods for treating HBV infection are urgently required. The invention features compositions and methods for introducing mutations into the hepatitis B virus (HBV) genome.



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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

COMPOSITIONS AND METHODS FOR TREATING HEPATITIS B VIRUS INFECTION

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims priority to and the benefit of U.S. Provisional Applications No. 63/371,634, filed August 16, 2022, 63/357,623, filed June 30, 2022, and 63/248,938, filed September 27, 2021, the entire contents of which are incorporated herein by reference.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

10 The contents of the electronic sequence listing (180802_053004_PCT_SL.xml; Size: 10,086,238 bytes; and Date of Creation: September 26, 2022) is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

15 Hepatitis B is a serious liver infection caused by the hepatitis B virus (HBV). HBV is a small DNA hepadnavirus that replicates through an RNA intermediate and can persist in infected cells by integrating into a host's genome. Approximately 257 million people worldwide, including between 850,000 and 2.2 million people in the United States, are chronically infected with HBV. Chronic HBV infection manifests as chronic hepatitis, cirrhosis, and/or hepatocellular carcinoma. Between 20% and 30% of adults who have chronic HBV infection
20 develop hepatocellular carcinoma or cirrhosis. HBV infection is responsible for between 600,000 and 1,000,000 deaths per year.

Current therapeutic approaches to HBV infection have severe limitations. Antiviral medications, e.g., tenofovir, a nucleotide reverse transcriptase inhibitor, can decrease viral replication but do not cure HBV infected patients. These antiviral therapies can cost patients as
25 much as \$500 to \$1500 monthly. Due to the extent of liver damage caused by HBV, a transplant becomes necessary in some cases. In addition to the risks inherent in organ transplants, the cost can be prohibitive. Therefore, improved methods for treating HBV infection are urgently required.

30 SUMMARY OF THE INVENTION OF THE DISCLOSURE

As described below, the present invention features compositions and methods for treating hepatitis B virus (HBV) infection by introducing alterations into the HBV genome. In particular embodiments, the invention provides a base editor system (e.g., a fusion protein comprising a programmable DNA binding protein, a nucleobase editor, and gRNA) for modifying the HBV

genome to introduce changes, such as premature stop codons or in the coding sequence of HBV or deamination of nucleobases in HBV covalently closed circular DNA (cccDNA).

In one aspect, the invention of the disclosure features a method of editing a nucleobase of a hepatitis B virus (HBV) genome. The method involves contacting the HBV genome with one or more guide RNAs and a base editor containing a polynucleotide programmable DNA binding domain and an adenosine deaminase. The guide RNA targets the base editor to effect an alteration of the nucleobase of the HBV genome. The one or more guide RNAs contain a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO:406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features a method of editing a nucleobase of a hepatitis B virus (HBV) genome. The method involves (i) contacting a cell containing an HBV genome with one or more guide RNAs and a base editor containing a polynucleotide programmable DNA binding domain and an adenosine deaminase or cytidine deaminase. The guide RNA targets the base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome. The method further involves (ii) contacting the cell with an antiretroviral drug. Contacting the cell with the antiretroviral drug is associated with an increase in base editing efficiency relative to a reference cell not contacted with the antiretroviral drug.

In another aspect, the invention of the disclosure features a method of editing a nucleobase of a hepatitis B virus (HBV) genome. The method involves (i) contacting a cell containing an HBV genome with an mRNA encoding a BE4 and guide RNAs gRNA37 and gRNA40. The guide RNAs targets the base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome. The method further involves (ii) contacting the cell with lamivudine. Contacting the cell with the antiretroviral drug is associated with an increase in base editing efficiency relative to a reference cell not contacted with the antiretroviral drug.

In another aspect, the invention of the disclosure features a method of editing a nucleobase of a hepatitis B virus (HBV) genome. The method involves contacting a cell containing an HBV genome with an mRNA encoding a base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase or cytidine deaminase, and
 5 guide RNAs comprising the sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO: 657) and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662). The guide RNAs target the base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome.

In another aspect, the invention of the disclosure features a method of treating hepatitis B
 10 virus (HBV) infection in a subject. The method involves administering to a subject in need thereof a fusion protein, protein complex, or polynucleotide encoding the fusion protein or protein complex. The fusion protein or protein complex contains a polynucleotide programmable DNA binding domain and a base editor domain that is an adenosine deaminase domain, and one or more guide RNAs that target the base editor domain to effect an A•T to G•C
 15 alteration of a nucleobase of the HBV genome. The one or more guide RNAs contain a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID
 20 NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAUA (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features a method of treating a hepatitis
 25 B virus (HBV) infection in a subject. The method involves administering to a subject in need thereof one or more polynucleotides encoding a polynucleotide programmable DNA binding domain and a base editor domain that is an adenosine deaminase domain, and one or more guide RNAs that target the base editor domain to effect an A•T to G•C, alteration of a nucleobase of
 30 the HBV genome. The one or more guide RNAs contain a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU

(SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411),
 GAGGACAAACGGGCAACAUA (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO:
 413), CCCAAGGUCUUAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID
 NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC
 5 (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features a method of treating a hepatitis
 B virus (HBV) infection in a subject. The method involves administering to a subject in need
 thereof with one or more guide RNAs and a base editor containing a polynucleotide
 programmable DNA binding domain and an adenosine deaminase or cytidine deaminase. The
 10 one or more guide RNAs contain the sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO:
 657) and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662) that target said base editor to effect
 an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV
 genome.

In another aspect, the invention of the disclosure features a composition containing a base
 15 editor(s) bound to a guide RNA(s). The guide RNA(s) contain a nucleotide sequence, from 5' to
 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or
 more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ
 ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA
 (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409),
 20 GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO:
 411), GAGGACAAACGGGCAACAUA (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID
 NO: 413), CCCAAGGUCUUAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU
 (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and
 CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features pharmaceutical composition for
 the treatment of HBV infection. The composition contains (i) a base editor, or a nucleic acid
 encoding the base editor, and one or more guide RNAs (gRNA) containing a nucleic acid
 sequence complementary to an HBV gene in a pharmaceutically acceptable excipient. The one
 or more guide RNAs contain a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide
 30 5' and/or 3' truncation fragment thereof, selected from one or more of
 ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID
 NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA
 (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409),

GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features a method of treating HBV infection. The method involves administering to a subject in need thereof the composition of any one of the above aspects

In another aspect, the invention of the disclosure features a method of treating an HBV infection, the method involves administering to a subject in need thereof the pharmaceutical composition of any one of the above aspects.

In another aspect, the invention of the disclosure features a use of the composition of any one of the above aspects in the treatment of HBV infection in a subject.

In another aspect, the invention of the disclosure features a use of the pharmaceutical composition of any one of the above aspects in the treatment of HBV infection in a subject.

In another aspect, the invention of the disclosure features a guide RNA containing a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In another aspect, the invention of the disclosure features a pharmaceutical composition containing (i) a nucleic acid encoding a base editor; and (ii) the guide RNA of any one of the above aspects.

In any of the above aspects, the antiretroviral drug is selected from one or more of Lamivudine, Entecavir, Tenofovir, Interferon, and Peg-Interferon. In any of the above aspects, the retroviral drug is lamivudine.

In any of the above aspects, step (i) precedes step (ii) or step (ii) precedes step (i). In any of the above aspects, the cell is first contacted with the antiretroviral drug about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 days prior to step (i). In any of the above

aspects, the cell has not been previously contacted with the antiretroviral drug. In any of the above aspects, the cell is in a subject. In any of the above aspects, the contacting is in a eukaryotic cell, a mammalian cell, or a human cell. In any of the above aspects, the cell is *in vivo* or *ex vivo*. In any of the above aspects or embodiments thereof, the subject is a mammal or a human. In any of the above aspects or embodiments thereof, the subject is a mammal. In any of the above aspects or embodiments thereof, the subject is a human.

In any of the above aspects, the HBV genome is contacted with two or more guide RNAs simultaneously.

In any of the above aspects the guide RNA is selected from one or more of
 10 ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUCCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAUA (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413),
 15 CCAAGGUCUUAUAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

In any of the above aspects, the one or more guide RNAs contain a spacer sequence selected from a sequence listed in Table 2A, Table 2B, Table 2C, and/or SEQ ID NOs: 3105-
 20 5485 and 8220-10830.

In any of the above aspects, the one or more guide RNAs contain spacers containing the nucleotide sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO: 657) and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662). In any of the above aspects, the one or more guide RNAs contain the nucleotide sequences

25 mG*mA*mA*AGCCCAGGAUGAUGGGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 424) and

mC*mC*mA*UGCCCCAAAGCCACCCAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422).

In any of the above aspects, base editing efficiency is improved by at least about 5%, 10%, 15%, or 20%.

In any of the above aspects, the adenosine deaminase converts a target A•T to G•C in the hepatitis B virus (HBV) genome.

In any of the above aspects, the nucleobase is in a polynucleotide encoding an HBV protein. In any of the above aspects, the nucleobase is in a polynucleotide sequence encoding an HBV protein selected from HBV core protein (Core), HBV polymerase (Pol), HBV surface protein, or HBV protein X. In embodiments of any of the above aspects, the alteration of the nucleobase in the polynucleotide encoding the HBV protein results in a missense mutation. In
 5 In embodiments of any of the above aspects, the alteration of the nucleobase is associated with a reduction in transcription of a polynucleotide sequence encoding an HBV protein. In embodiments of any of the above aspects, the HBV protein is HBV polymerase (Pol), and/or HBV protein X. In embodiments of any of the above aspects, the missense mutation is in an
 10 HBV pol gene. In embodiments of any of the above aspects, the missense mutation results in an E24G, L25F, P26F, R27C, V48A, V48I, S382F, V378I, V378A, V379I, V379A, L377F, D380G, D380N, F381P, R376G, A422T, F423P, A432V, M433V, P434S, D540G, A688V, D689G, A717T, E718K, P713S, P713L, or L719P in an HBV polymerase protein encoded by the HBV pol gene. In embodiments of any of the above aspects, the missense mutation is in an HBV core
 15 gene. In embodiments of any of the above aspects, the missense mutation results in a T160A, T160A, P161F, S162L, C183R, or *184Q in an HBV core protein encoded by the HBV core gene. In embodiments of any of the above aspects, the missense mutation is in an HBV X gene. In embodiments of any of the above aspects, missense mutation results in a H86R, W120R, E122K, E121K, or L141P in an HBV X protein encoded by the HBV X gene. In embodiments
 20 of any of the above aspects, the missense mutation is in an HBV S gene. In embodiments of any of the above aspects, the missense mutation results in a S38F, L39F, W35R, W36R, T37I, T37A, R78Q, S34L, F80P, or D33G in an HBV S protein encoded by the HBV S gene.

In any of the above aspects, the nucleobase is associated with a transcription site. In
 25 In embodiments of any of the above aspects, the transcription site is selected from one or more of Enhancer II box A, Enhancer I, and the HBX promoter.

In any of the above aspects, the polynucleotide programmable DNA binding domain contains a Cas12 polypeptide. In any of the above aspects, the polynucleotide programmable DNA binding domain contains Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, or Cas12i. In any of the above aspects, the polynucleotide programmable DNA binding domain
 30 contains a sequence with at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In any of the above aspects, the polynucleotide programmable DNA binding domain contains a sequence with at least about 85% amino acid sequence identity to *Bacillus hiasashii* Cas12b (bhCas12b). In any of the above aspects, the polynucleotide

programmable DNA binding domain contains a nuclease inactive or nickase variant. In embodiments of any of the above aspects, the nuclease inactive or nickase variant contains a nuclease inactivated bhCas12b which contains amino acid substitutions D952A, S893R, K846R, and E837G, or corresponding amino acid substitutions thereof. In any of the above aspects, the base editor contains a bhCas12b, or a bhCas12b variant containing amino acid substitutions D952A, S893R, K846R, and E837G.

In any of the above aspects, the base editor contains an adenosine deaminase. In any of the above aspects or embodiments thereof, the adenosine deaminase domain is capable of deaminating adenine in deoxyribonucleic acid (DNA). In any of the above aspects or embodiments thereof, the adenosine deaminase is a TadA deaminase. In embodiments of any of the above aspects, the TadA deaminase is TadA*7.10, 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 embodiments of any of the above aspects, the TadA deaminase is TadA*8.13.

In any of the above aspects, the one or more guide RNAs contains a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), where the crRNA contains a nucleic acid sequence complementary to an HBV nucleic acid sequence. In any of the above aspects, the one or more guide RNAs target three, four, or five nucleobases of the HBV genome.

In any of the above aspects, the base editor is in complex with a single guide RNA (sgRNA) containing a nucleic acid sequence complementary to an HBV nucleic acid sequence.

In any of the above aspects, the method involves editing two or more nucleobases.

In any of the above aspects, the method involves contacting the HBV genome with two or more guide RNAs that target two or more HBV nucleic acid sequences. In any of the above aspects, the method involves delivering the fusion protein, the polynucleotide encoding the fusion protein, or one or more polynucleotides encoding the polynucleotide programmable DNA binding domain and the base editor domain, and the one or more guide RNAs to a cell of the subject. In any of the above aspects or embodiments thereof, the cell is a hepatocyte.

In any of the above aspects, the base editor (i) contains a nuclease inactive bhCas12b; or (ii) contains an amino acid sequence that is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHHPGMNH
RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGAPKKK

RKVGIHGVPAATRSLFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDP
 KNPKKVSKAEIQAEWLDFVLKMQKCNSTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN
 KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEKKKKWEEDKKKDPLAKILGKLAEYGL
 LIPLFIPYTDSEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYKVEK
 5 EYKTLERIKEDIQALKALEQYKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENE
 PSEKYLEVFKDYQRKHPREAGDYSVYEFSLKKNHFIWRNHPEYPYLYATFCEIDKKKKDAKQQ
 ATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKG
 KVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLLGGARVQFDRDHLRRYPHKVES
 GNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEI
 10 GLRVMSIALGQRQAAAASIFEVVDQKPDIEGKLFPIKGTETYAVHRASFNIKLPGETLVKSRE
 VLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENS DVPLVYQDELIQIRE
 LMYKPYKDWVAFKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWS
 LRPTEPGEVRRLEPGQRFQIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPAC
 QIILFEDLSNYPYKERSRFENSRLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTG
 15 SPGIRCRVVTKEKLQDNRRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTH
 ADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIIEFGEQYFILKDGVEE
 WVNAGKLKIKKGSSKQSSSELVDSILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVF
 FGKLERILISKLTNQYSISTIEDSSKQSMKRPAATKKAGQAKKKK (SEQ ID NO: 418).

In any of the above aspects, the composition or pharmaceutical composition further
 20 contains a lipid. In embodiments of any of the above aspects, the lipid is a cationic lipid.

In any of the above aspects, the composition contains a pharmaceutically acceptable
 excipient.

In any of the above aspects, the one or more gRNAs and the base editor are formulated
 together. In any of the above aspects, the one or more gRNAs and the base editor are formulated
 25 separately.

In any of the above aspects, the composition or pharmaceutical composition further
 contains a vector suitable for expression in a mammalian cell, where the vector contains a
 polynucleotide encoding the base editor. In any of the above aspects or embodiments thereof,
 the vector is a viral vector. In any of the above aspects or embodiments thereof, the viral vector
 30 is a retroviral vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated
 viral vector (AAV). In any of the above aspects or embodiments thereof, the composition or
 pharmaceutical composition further contains a ribonucleoparticle suitable for expression in a
 mammalian cell.

In any of the above aspects, the HBV is of genotype A, or of genotype D. In any of the above aspects, the method reduces levels of a marker selected from one or more of Hepatitis B Surface Antigen, Hepatitis B E Antigen, Hepatitis B Virus total DNA, Hepatitis B 3.5 kb RNA, and Hepatitis B covalently closed circular DNA. In any of the above aspects, the method edits the Hepatitis B S antigen site and the PreCore site. In any of the above aspects or embodiments thereof, the editing efficiency is about 30% at the S antigen site and about 60% at the PreCore site.

In any of the above aspects, the nucleic acid encoding the base editor is an mRNA. In any of the above aspects, the one or more guide RNAs comprise the following scaffold sequence:

gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl 3'-phosphorothioate (MS) modification. In any of the above aspects, the one or more guide RNAs

comprise the following nucleotide sequences

gsasasAGCCCAGGAUGAUGGGAgUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucggugcusususu (SEQ ID NO: 424) and cscsasUGCCCCAAGCCACCCAgUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucggugcusususu (SEQ ID NO: 422), where a, c, u, or g represent an A, C, U, or G nucleotide containing a 2'-O-methyl (M) modification, respectively; and as, cs, us, or gs represent an A, C, U, or G nucleotide containing a 2'-O-methyl 3'-phosphorothioate (MS) modification, respectively.

In any of the above aspects, the cell is contacted at a first and a second time point with the one or more guide RNAs and the base editor. In embodiments, at least one week separates the first and second times the cell is contacted. In embodiments, at least two weeks separates the first and second times the cell is contacted.

In any of the above aspects, the base editor is a BE4. In any of the above aspects, the gRNA comprises GAAAGCCCAAGAUGAUGGGA (SEQ ID NO. 10834).

Compositions and articles defined by the invention were isolated or otherwise manufactured in connection with the examples provided below. Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

Definitions

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

By “BE4 polypeptide” is meant a polypeptide having at least about 85% identity to the following amino acid sequence, or a fragment thereof having cytidine deaminase activity:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFITTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNENAHWPYPHLLWVRLYVL
ELYCILGLPPCLNLRKQPKLTFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSET
PGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPKSKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMI
KFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL
FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQN
GRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI
LDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVDVRKMIKSEKQEGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQVNIVKKTEVQTGGFSKESILP

KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
 RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELALPS
 KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKV
 LSAYNKHHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQ
 5 ITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK
 PESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
 NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
 EYKPWALVIQDSNGENKIKMLSGGSPKKKRK (SEQ ID NO: 10832).

By “BE4 polynucleotide” is meant a polynucleotide encoding a BE4 polypeptide.

10 By “HBV polymerase (pol) protein” is meant a polypeptide having at least about 95%
 identity to a wild-type HBV polymerase amino acid sequence or fragment thereof that functions
 in a hepatitis B viral infection. In one embodiment, the HBV polymerase is encoded by an HBV
 A, B, C, D, E, F, G, or H genotype. In one embodiment, the HBV polymerase amino acid
 sequence is provided at UniPro Accession No. Q8B5R0-1, which is reproduced below.

15 MPLSYQHFRLLLLDDEAGPLEEELPRLADEGLNRRVAEDLNLGNLNV SIPWTHKVGNF
 TGLYSSTVPVFNPHWKTPSFPNIHLHQDIKKCEQFVGPLTVNEKRRLQLIMPARFY PKV
 TKYLPLDKGIKPYPPEHLVNHYFQTRHYLHTLWKAGILYKRETT HSASF CGSPYSWEQD
 LQHGAESFHQQS (SEQ ID NO: 370).

Mutations in HBV polymerase include: E24G, L25F, P26F, R27C, V48A, V48I, S382F,
 20 V378I, V378A, V379I, V379A, L377F, D380G, D380N, F381P, R376G, A422T, F423P,
 A432V, M433V, P434S, D540G, A688V, D689G, A717T, E718K, P713S, P713L, or L719P.

Other exemplary HBV DNA polymerases include, for example, NCBI Accession No.
 AAB59972.1, which has the following sequence:

MPLSYQHFRKLLLLDDEAGPLEEELPRLADEGLNRRVAEDLNLGNLNV SIPWTHKVGNF
 25 TGLYSSTVPVFNPHWKTPSFPNIHLHQDIKKCEQFVGPLTVNEKRRLQLIMPARFY PKV
 TKYLPLDKGIKPYPPEHLVNHYFQTRHYLHTLWKAGILYKRETT HSASF CGSPYSWEQD
 LQHGAESFHQQSSGILSRPPVGSSLQSKHSKSRLGLQSQQGH LARRQQGRSWSIRAGFHP
 TARRPFGVEPSGSGHTTNFASKSASCLHQSPDRKAA YPAVSTFEKHSSSGH AVEFHNLS P
 NSARSQSERP VFPCWWLQFRSSKPCSDYCLSLIVNLLEDWGPCA EHGEHHIRIPRTPSRV
 30 TGGVFLVDKNPHNTAESRLVVD FSQFSRGNYRVS WPKF AVPNLQSLTNLLSSNLSWLSL
 DVSAAFYHLPLHPAAMP HLLVGSSGLSRYVARLSSNSRILNHQHGTMPNLHDYCSRNL
 YVSLLLLYQTFGRKLHLYSHPIILGFRKIPMGVGLSPFLLAQFTSAICSVVRRAPPHCLAFS
 YMDDVVLGAKSVQHLES LFTA VTNFLLSLGIHLNPNKTKRWGYS LNFMGYVIGSYGSL
 PQEHIIQKIKECFRKL PINRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQSKQAF

TFSPTYKAFLCKQYLNLYPVARQRPGLCQVFADATPTGWGLVMGHQVRGTFSAPLPI
HTAELLAACFARSRSGANIIGTDNSVLSRKYTSYPWLLGCAANWILRGTSFVYVPSAL
NPADDPSRGRGLSLRPLLRLPFRPTTGRTSLYADSPSVPSHLPDRVHFASPLHVAWRPP
(SEQ ID NO: 371).

5 By “HBV polymerase gene” is meant a polynucleotide encoding an HBV polymerase.

By “Hepatitis B surface antigen (HBsAg) polypeptide” or “HBV surface protein (S)” is meant an antigenic protein or fragment thereof having at least about 85% identity to NCBI Accession No. AAB59969.1, which functions in an HBV viral infection. An exemplary HBsAg amino acid sequence is provided below:

10 MENITSGFLGPLLVLQAGFFLLTRILTIQSLDSWWTSLNFLGGTTVCLGQNSQSPTS NHS
PTSCPPTCPGYRWMCLRRFIIFLLCLIFLLVLLDYQGMLPVCPLIPGSSTTSTGPCRTC
MTTAQGTSMYPSCCCTKPSDGNCTCIPISSWAFGKFLWEWASARFSWLSLLVPFVQWF
VGLSPTVWLSVIWMMWYWGPSLYSILSPFLPLLPIFFCLWVYI (SEQ ID NO: 372).

By “HbsAg polynucleotide” is meant a polynucleotide encoding an HBsAg protein.

15 By “HBV X-protein” or “HBV protein X” is meant a polypeptide or fragment thereof having at least about 85% identity to NCBI Accession No. AAB59970.1, which functions in an HBV viral infection. An exemplary amino acid sequence is provided below:

MAARLCCQLDPARDVLCLRPVGAESCGRPFSGSLGTLSSPSPSAVPTDHGAHLSLRGLP
VCAFSSAGPCALRFTSARRMETTVNAHRMLPKVLHKRTLGLSAMSTTDLEAYFKDCLF
20 KDWEELGEEIRLKVFVLGGCRHKLVCAPAPCNFFTSA (SEQ ID NO: 373).

By “core antigen precursor,” “precore protein,” or “Hepatitis B e antigen (HBeAg)” is meant a polypeptide or fragment thereof having at least about 85% identity to NCBI Accession No. AAB59971.1, reproduced below, which functions in an HBV viral infection.

MQLFHLCLIISCSCPTVQASKLCLGWLWGMDIDPYKEFGATVELLSFLPSDFFPSV
25 RDLLDTASALYREALSPEHCSPHHTALRQAILCWGELMTLATWVGVNLEDPASRDLV
VSYVNTNMGLKFRQLLWFHISCLTFGRETVIEWLVSFGVWIRTPPAYRPPNAPILSTLPET
TVVRRRGRSPRRRTSPRRRRRSQSPRRRRRSQSREPQC (SEQ ID NO: 10831).

By “HBV core protein,” “HBc,” or “core protein” is meant a polypeptide having at least about 95% identity to a wild-type HBV core protein amino acid sequence or fragment thereof.

30 In an embodiment, the HBV core protein functions in a hepatitis B viral infection. In one embodiment, the HBV core protein is encoded by an HBV A, B, C, D, E, F, G, or H genotype. In one embodiment, the HBV core protein amino acid sequence is provided at NCBI GenBank Accession No. AXG50928.1, provided below:

MDIDPYKEFGASVELLSFLPSDFFPSIRDLLDTASALYREALESPEHCSPHHTALR
 QAILCWGELMNLATWVGSNLEDPASRELVVSYVNVNMGLKIRQLLWFHISCLTFGRET
 VLEYLVSFGVWIRTPPAYRPPNAPILSTLPETTIVRRRGRSPRRRTPSPRRRRSQSPRRRR
 SQSRESQC (SEQ ID NO: 374). By "HBV X protein" is meant a polynucleotide encoding an

5 HBV X-protein.

By "HBV X protein (genotype B)" is meant a polypeptide having at least about 95% identity to a wild-type HBV genotype B X protein amino acid sequence or fragment thereof. In an embodiment, the HBV X protein functions in a hepatitis B viral infection. In one embodiment, the HBV genotype B X protein amino acid sequence is provided at NCBI GenBank
 10 Accession No. BAQ95575.1, provided below:

MAARLCCQLDPARDVLCLRPVGAESRGRPLPGPLGALPPASPPVPSDHGAHLRLRGLP
 VCAFSSXGPCALRFTSARRMETTVNAHRNLPKVLHKRTLGLSAMSTTDLEAYFKDCVF
 XEWEELGEEEXRLKVFLGGCRHKLVCSPAPCNFF TSA (SEQ ID NO: 375).

By "HBV X protein (genotype C)" is meant a polypeptide having at least about 95% identity to a wild-type HBV genotype C X protein amino acid sequence or fragment thereof. In an embodiment, the HBV X protein functions in a hepatitis B viral infection. In one embodiment, the HBV genotype C X protein amino acid sequence is provided at NCBI GenBank
 15 Accession No. BAQ95563.1, provided below:

MAARVCCQLDPARDVLCLRPVGAESRGRPVSGPFGPLPSPSSSAVPADYGAHLRLRGLP
 20 VCAFSSAGPCALRFTSARRMETTVNAHQVLPKLLHKRTLGLSAMSTTDLEAYFKDCFLK
 DWEELGEEIRLKVFLGGCRHKLVCSPAPCNFF TSA (SEQ ID NO: 376).

By "HBV S protein" is meant a polypeptide having at least about 95% identity to a wild-type HBV S protein amino acid sequence or fragment thereof. In an embodiment, the HBV S protein functions in a hepatitis B viral infection. In one embodiment, the HBV S protein is encoded by an HBV A, B, C, D, E, F, G, or H genotype. In one embodiment, the HBV S protein amino acid sequence is provided at NCBI GenBank Accession No. ABV02793.1, provided
 25 below:

MENTTSGFLGPLLVLQAGFFLLTRNLTIPQSLDSWWTSLNFLGGAPTCPGQNSQSPTS NH
 SPTSCPPICPGYRWMCLRRFIIFLLIFLLVLLDYQGMLPVCPLLPGTSTTSTGPCKT
 30 CTIPAQGTSMFPSCCCTKPSDGNCTCIPISSWAFARFLWEWASVRFSWLSLLVPFVQWF
 VGLSPTVWLSVIWMMWYWGPSLYNILSPFLPLLPIFFCLWVYI (SEQ ID NO: 377).

The complete genome of Hepatitis B virus subtype ayw, complete genome, which includes polynucleotides encoding HBV polymerase, HBsAg protein, HBV X protein, and the core antigen precursor, is provided at GenBank Accession No. U95551.1, which is reproduced

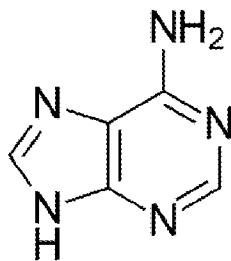
below. The nucleotide locations of regions corresponding to regulatory elements (e.g., Enhancer I, Enhancer II box A, HBX promoter) and polypeptides-encoding sequences within the Hepatitis B virus genome are known the art (see, e.g., Panjaworayan, *et al.* "HBVRegDB: Annotation, comparison, detection and visualization of regulatory elements in hepatitis B virus sequences,"

5 *Virol. J.*, 4:136, DOI: 10.1186/1743-422X-4-136).

AATCCACAACCTTTCACCAAACCTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTTCC
CCTGCTGGTGGCTCCAGTTCAGGAGCAGTAAACCCTGTTCCGACTACTGCCTCTCCC
TTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCGCTGAACATGGAGAACATCACA
TCAGGATTCCTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGA
10 ATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGG
GGAACCTACCGTGTGTCTTGGCCAAAATTGCGAGTCCCCAACCTCCAATCACTCACCA
ACCTCCTGTCCTCCAACCTTGTCCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCA
TCTTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAA
GGTATGTTGCCCCGTTTGTCTCTAATTCCAGGATCCTCAACCACCAGCACGGGACCA
15 TGCCGAACCTGCATGACTACTGCTCAAGGAACCTCTATGTATCCCTCCTGTTGCTGT
ACCAAACCTTCGGACGGAAATTGCACCTGTATTCCCATCCCATCATCCTGGGCTTTC
GGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTG
CCATTTGTTCAAGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGA
TGATGTGGTATTGGGGGCCAAGTCTGTACAGCATCTTGAGTCCCTTTTTACCGCTGTT
20 ACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAAAACAAAGAGATGG
GGTACTCTCTGAATTTTATGGGTTATGTCATTGGAAGTTATGGGTCCTTGCCACAAG
AACACATCATACAAAAAATCAAAGAATGTTTTAGAAAACCTTCCTATTAACAGGCCT
ATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTGGGTTTTGCTGCCCCATTTA
CACAATGTGGTTATCCTGCGTTAATGCCCTTGTATGCATGTATTCAATCTAAGCAGG
25 CTTTCACTTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATACCTGAACCTTTA
CCCCGTTGCCCGGCAACGGCCAGGTCTGTGCCAAGTGTTTGCTGACGCAACCCCCAC
TGGCTGGGGCTTGGTCATGGGCCATCAGCGCGTGCGTGGAACCTTTTCGGCTCCTCT
GCCGATCCATACTGCGGAACCTCCTAGCCGCTTGTTTTGCTCGCAGCAGGTCTGGAGC
AAACATTATCGGGACTGATAACTCTGTTGTCCTCTCCCGCAAATATACATCGTATCC
30 ATGGCTGCTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGT
CCCGTCGGCGCTGAATCCTGCGGACGACCCTTCTCGGGGTCGCTTGGGACTCTCTCG
TCCCCCTTCTCCGTCTGCCGTTCCGACCGACCACGGGGCGCACCTCTCTTTACGCGGA
CTCCCCGTCTGTGCCTTCTCATCTGCCGACCGTGTGCACTTCGCTTCACCTCTGCAC
GTCGCATGGAGACCACCGTGAACGCCACCGAATGTTGCCCAAGGTCTTACATAAG

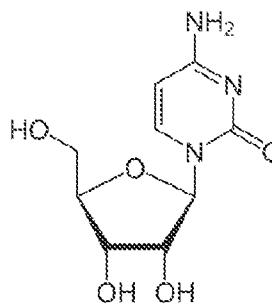
AGGACTCTTGGACTCTCTGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGAC
TGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGATTAAAGGTCTTTGTA
CTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATTGCAACTTTTTTCACC
TCTGCCTAATCATCTCTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGG
5 GTGGCTTTGGGGCATGGACATCGACCCTTATAAAGAATTTGGAGCTACTGTGGAGTT
ACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGATCTTCTAGATAACCGCCT
CAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGTTACCTCACCATACTG
CACTCAGGCAAGCAATTCTTTGCTGGGGGGAATAATGACTCTAGCTACCTGGGTGG
GTGTTAATTTGGAAGATCCAGCATCTAGAGACCTAGTAGTCAGTTATGTCAACACTA
10 ATATGGGCCTAAAGTTCAGGCAACTCTTGTGGTTTCACATTTCTTGTCTCACTTTTGG
AAGAGAAACCGTTATAGAGTATTTGGTGTCTTTCGGAGTGTGGATTGCGACTCCTCC
AGCTTATAGACCACCAAATGCCCCTATCCTATCAACACTTCCGGAAACTACTGTTGT
TAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGGT
CTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAACCTCAATGTTAGTATTCCT
15 TGGACTCATAAGGTGGGGAACTTTACTGGTCTTTATTCTTCTACTGTACCTGTCTTTA
ATCCTCATTGGAAAACACCATCTTTTCCTAATATACATTTACACCAAGACATTATCA
AAAAATGTGAACAGTTTGTAGGCCCACTTACAGTTAATGAGAAAAGAAGATTGCAA
TTGATTATGCCTGCTAGGTTTTATCCAAAGGTTACCAAATATTTACCATTGGATAAG
GGTATTAAACCTTATTATCCAGAACATCTAGTTAATCATTACTTCCAAACTAGACAC
20 TATTTACACACTCTATGGAAGGCGGGTATATTATATAAGAGAGAAACAACACATAG
CGCCTCATTTTGTGGGTCACCATATTCTTGGGAACAAGATCTACAGCATGGGGCAGA
ATCTTTCCACCAGCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTT
CAGAGCAAACACAGCAAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGC
CAGACGCCAACAAGGTAGGAGCTGGAGCATTTCGGGCTGGGTTTACCCCAACCGCAC
25 GGAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTACAACTTTGCCAGC
AAATCCGCCTCCTGCCTCCACCAATCGCCAGACAGGAAGGCAGCCTACCCCGCTGTC
TCCACCTTTGAGAAACACTCATCCTCAGGCCATGCAGTGG (SEQ ID NO: 378).

By “adenine” or “9H-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-[(2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2(1H)-one” is meant an adenine 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 $C_{10}H_{13}N_5O_4$.

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 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). 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.

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.*,
 5 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 Deaminase Base Editor 8 (ABE8) polypeptide” or “ABE8” is meant a
 10 base editor as defined herein comprising an 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 sites listed in Table 15 (*e.g.*, 82 and/or 166) of the following reference sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRAIGLHDPTAHAEIMALR
 15 QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
 RVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1) .

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

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

“Administering” is referred to herein as providing one or more compositions described herein to a patient or a subject.

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

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

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

By “analog” is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polypeptide analog retains the biological activity of a

corresponding naturally-occurring polypeptide, while having certain biochemical 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
5 an unnatural amino acid.

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
10 polynucleotide (*e.g.*, guide RNA (gRNA)). Representative nucleic acid and protein sequences of base editors are provided in the Sequence Listing as SEQ ID NOs: 2-11.

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
15 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
20 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
25 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
30 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 base editor (CBE).

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 deaminase activity, e.g.,
 5 converting A•T to G•C.

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 casn1 nuclease or a CRISPR (clustered regularly interspaced short
 10 palindromic repeat) associated nuclease.

The term “conservative amino acid substitution” or “conservative mutation” refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms
 15 (Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)). According to such analyses, groups of amino acids can be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and Schirmer, R. H., supra). Non-limiting examples of conservative mutations include amino acid substitutions of amino acids, for
 20 example, lysine for arginine and vice versa such that a positive charge can be maintained; glutamic acid for aspartic acid and vice versa such that a negative charge can be maintained; serine for threonine such that a free –OH can be maintained; and glutamine for asparagine such that a free –NH₂ can be maintained.

The term “coding sequence” or “protein coding sequence” as used interchangeably herein
 25 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

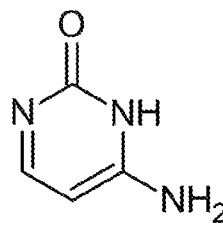
Tryptophan TGG → TGA

TGG → TAG

TGG → TAA

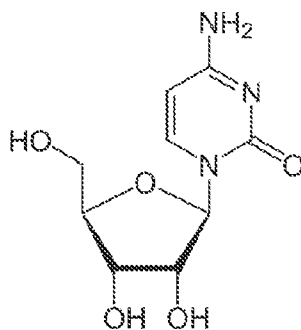
By “complex” is meant a combination of two or more molecules whose interaction relies on inter-molecular forces. Non-limiting examples of inter-molecular forces include covalent and non-covalent interactions. Non-limiting examples of non-covalent interactions include hydrogen bonding, ionic bonding, halogen bonding, hydrophobic bonding, van der 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 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 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 bonds.

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



molecular formula $C_4H_5N_3O$, having the structure 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 deaminase.

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

5 By “cytidine deaminase” or “cytosine deaminase” is meant a polypeptide or fragment thereof capable of deaminating cytidine or cytosine. 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. *Petromyzon marinus* cytosine deaminase 1 (PmCDA1) (SEQ ID NO: 13-14), Activation-induced cytidine
10 deaminase (AICDA) (SEQ ID NOs: 15-21), and APOBEC (SEQ ID NOs: 12-61) are 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**. In embodiments, the cytidine deaminase is a BE4, rrA3F, or ppApo.

By “rrA3F polypeptide” is meant a cytidine deaminase having cytidine deaminase activity
15 and with an amino acid sequence with about or at least about 85%, 90%, 95%, 99%, or 100% sequence identity to the following sequence:

MKPQIRDHRPNPMEAMYPHIFYHFENLEKAYGRNETWLCFTVEIHKQYLPVPWKKGVF
RNQVDPETHCHAEKCFLSWFCNNTLSPKKNYQVTWYTSWSPCPECAGEVAEFLAEHSN
VKLTIYTARLYYFWDTDYQEGRLSLSEEGASVEIMDYEDFQYCWENFVYDDGEPFKRW
20 KGLKYNFQSLTRRLREILQ (SEQ ID NO: 67).

By “ppAPOBEC-1 (ppApo)” is meant a cytidine deaminase having cytidine deaminase activity and with an amino acid sequence with about or at least about 85%, 90%, 95%, 99%, or 100% sequence identity to the following sequence:

MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKRETCCLLYEIKWGMSRKIWRSSGKNTT
25 NHVEVNFIIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLF
WHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMM
LYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWR
(SEQ ID NO: 23).

By “cytosine” is meant a pyrimidine nucleobase with the molecular formula $C_4H_5N_3O$.

30 By “cytosine deaminase activity” is meant catalyzing the deamination 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, 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.

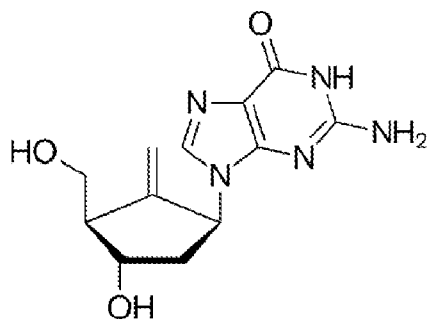
5 “Detect” refers to identifying the presence, absence or amount of the analyte to be detected. In one embodiment, a sequence alteration in a polynucleotide or polypeptide is detected. In another embodiment, the presence of indels is detected.

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

By “disease” is meant any condition or disorder that damages or interferes with the
15 normal function of a cell, tissue, or organ. Examples of diseases include HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection.

By “effective amount” is meant the amount of a required to ameliorate the symptoms of a disease relative to an untreated patient. The effective amount of active compound(s) used to
20 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 an HBV
25 genome in a cell (e.g., a cell in vitro or in vivo). In one embodiment, an effective amount is the amount of a base editor required to achieve a therapeutic effect (e.g., to reduce or control an HBV infection). Such therapeutic effect need not be sufficient to alter an HBV genome in all cells of a subject, tissue or organ, but only to alter an HBV genome in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an
30 effective amount is sufficient to ameliorate one or more symptoms of HBV.

By "Entecavir" is meant a molecule having the structure



, corresponding to CAS no. 142217-69-4, and pharmaceutically acceptable salts thereof. Entecavir has antiretroviral activity.

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

By "guide polynucleotide" is meant a polynucleotide or polynucleotide complex which is specific for a target sequence and can form a complex with a polynucleotide programmable nucleotide binding domain protein (*e.g.*, Cas9 or Cpf1). In an embodiment, the guide polynucleotide is a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule.

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

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

The terms "inhibitor of base repair", "base repair inhibitor", "IBR" or their grammatical equivalents refer to a protein that is capable in inhibiting the activity of a nucleic acid repair enzyme, for example a base excision repair enzyme.

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

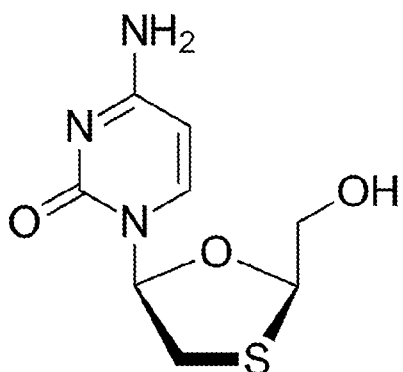
The terms "isolated," "purified," or "biologically pure" refer to material that is free to varying degrees from components which normally accompany it as found in its native state. "Isolate" denotes a degree of separation from original source or surroundings. "Purify" denotes a degree of separation that is higher than isolation. A "purified" or "biologically pure" protein is sufficiently free of other materials such that any impurities do not materially affect the biological properties of the protein or cause other adverse consequences. That is, a nucleic acid or peptide

of this invention is purified if it is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Purity and homogeneity are typically determined using analytical chemistry techniques, for example, polyacrylamide gel electrophoresis or high performance liquid chromatography. The term "purified" can denote that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. For a protein that can be subjected to modifications, for example, phosphorylation or glycosylation, different modifications may give rise to different isolated proteins, which can be separately purified.

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

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

By "lamivudine" or "3TC" is meant a molecule having the structure



, corresponding to CAS no. 134678-17-4, and having the

IUPAC name 2',3'-dideoxy-3'-thiacytidine 4-Amino-1-[(2*R*,5*S*)-2-(hydroxymethyl)-1,3-oxathiolan-5-yl]-1,2-dihydropyrimidin-2-one, and pharmaceutically acceptable salts thereof. Lamivudine has antiretroviral activity. Lamivudine is classified as a nucleoside/nucleotide reverse transcriptase inhibitor (NRTIs).

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

By “marker” is meant any protein or polynucleotide having an alteration in expression level or activity that is associated with a disease or disorder. Examples of diseases include HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection. The marker can be an HBV polynucleotide and/or polypeptide. Non-limiting examples of Hepatitis B virus markers include HBsAg, HBeAg, core protein, 3.5 kb viral RNA, and cccDNA (see, e.g., Bai, *et al.* “Quantification of Pregenomic RNA and Covalently Closed Circular DNA in Hepatitis B Virus-Related Hepatocellular Carcinoma,” *Int J Hepatol*, 2013:849290 (2013), DOI: 10.1155/2013/849290).

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

25 The terms “nucleic acid” and “nucleic acid molecule,” as used herein, refer to a compound comprising a nucleobase and an acidic moiety, e.g., a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, e.g., nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, “nucleic acid” refers to individual nucleic acid residues (e.g. nucleotides and/or nucleosides). In some embodiments, “nucleic acid” refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides (e.g., a string of at least three nucleotides). In some embodiments, “nucleic acid” encompasses RNA as well as single and/or double-

stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms “nucleic acid,” “DNA,” “RNA,” and/or similar terms include nucleic acid analogs, *e.g.*, analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (2'-*e.g.*, fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5'-*N*-phosphoramidite linkages).

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), KRGINDRNFWRGENGKTR (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-hydromethylcytosine. Hypoxanthine and xanthine can be created through mutagen presence, both of them through deamination (replacement of the amine group with a carbonyl group). Hypoxanthine can be modified from adenine. Xanthine can be modified from guanine. Uracil can result from deamination of cytosine. A “nucleoside” consists of a nucleobase and a five carbon sugar (either ribose or deoxyribose). Examples of a nucleoside include adenosine, guanosine, uridine, cytidine, 5-methyluridine (m5U), deoxyadenosine, deoxyguanosine, thymidine, deoxyuridine, and deoxycytidine. Examples of a nucleoside with a modified nucleobase includes inosine (I), xanthosine (X), 7-methylguanosine (m7G), dihydrouridine (D), 5-methylcytidine (m5C), and pseudouridine (Ψ). A “nucleotide” consists of a nucleobase, a five carbon sugar (either ribose or deoxyribose), and at least one phosphate group.

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

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" is meant a mammal, including, but not limited to, a human or non-human mammal, such as a bovine, equine, canine, ovine, rodent, 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, horses, camels, llamas, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other

mammalians that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female. Examples of diseases include HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection.

5 “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 casing mutation”, “disease causing variant”, “deleterious mutation”, or “predisposing mutation” refers to a genetic
10 alteration or mutation that increases an individual’s susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation comprises at least one wild-type amino acid substituted by at least one pathogenic amino acid in a protein encoded by a gene.

 The terms “protein”, “peptide”, “polypeptide”, and their grammatical equivalents are
15 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.

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

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

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

 By “reference” is meant a standard or control condition. In one embodiment, the reference is a wild-type or healthy cell. In other embodiments and without limitation, a reference is an untreated cell that is not subjected to a test condition, or is subjected to placebo or normal
30 saline, medium, buffer, and/or a control vector that does not harbor a polynucleotide of interest. A reference can be a cell or subject free of an HBV infection. A reference can also be a subject prior to receiving a treatment, or prior to an alteration in a treatment.

 A “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a

segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, at least about 20 amino acids, at least about 25 amino acids, about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference
5 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.

10 The term "RNA-programmable nuclease," and "RNA-guided nuclease" are used with (*e.g.*, binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). In some embodiments, the RNA-programmable nuclease is the (CRISPR-
15 associated system) Cas9 endonuclease, for example, Cas9 (CsnI) from *Streptococcus pyogenes*.

The term "single nucleotide polymorphism (SNP)" is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.*, > 1%).

By "specifically binds" is meant a nucleic acid molecule, polypeptide,
20 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.

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
25 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 60%, 80%, 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid level to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example,
30 Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, 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 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.

5 COBALT is used, for example, with the following parameters:

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

10 distance 0.8; Alphabet Regular.

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

- a) Matrix: BLOSUM62;
- b) GAP OPEN: 10;
- c) GAP EXTEND: 0.5;
- 15 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

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

25 include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-

30 stranded molecule between complementary polynucleotide sequences (*e.g.*, a gene described herein), or portions thereof, under various conditions of stringency. (See, *e.g.*, Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

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

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

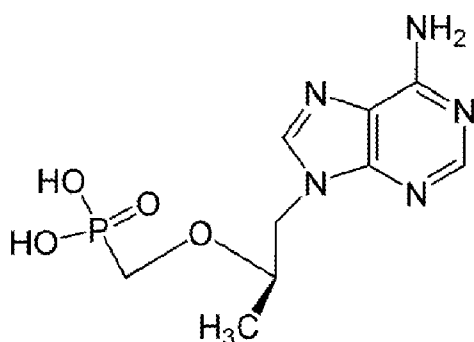
Cloning Techniques, 1987, Academic Press, New York); and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

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

By “Tenofovir” is meant a molecule having the structure



, corresponding to CAS no. 147127-20-6, and

pharmaceutically acceptable salts thereof. Tenofovir has antiretroviral activity. Tenofovir is classified as a nucleoside/nucleotide reverse transcriptase inhibitor (NRTIs).

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

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. Including an inhibitor of

uracil DNA glycosylase (UGI) in the base editor prevents base excision repair which changes the U back to a C. An exemplary UGI comprises an amino acid sequence as follows:

>splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD

5 APEYKPWALVIQDSNGENKIKML (SEQ ID NO: 231).

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, 10 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.

15 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 20 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 25 comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and “contain”) are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. It is contemplated that any embodiment discussed in this specification can be implemented with 30 respect to any method or composition of the present disclosure, and vice versa. Furthermore, compositions of the present disclosure can be used to achieve methods of the present disclosure.

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

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

Reference in the specification to “some embodiments,” “an embodiment,” “one embodiment” or “other embodiments” means that a particular feature, structure, or 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

FIG. 1 is an illustration showing the partially double-stranded and the overlapping open reading frames (ORFs) for the hepatitis B surface antigen (HBsAg) gene, the polymerase gene, the protein X gene, and the core gene. The HBsAg gene comprises ORF PreS1, ORF PreS2, and ORF S, which encode the large, middle, and small surface proteins, respectively. ORF core and Pre C encode capsid proteins.

FIG. 2 is an illustration depicting the HBV life cycle. The term “ER” denotes endoplasmic reticulum. The term “HBsAg” denotes hepatitis B surface antigen. “HBx transcriptional activator” is an HBV-specific transcriptional activator of polymerase II and III promoters.

FIGs. 3A and 3B present a map and a presentation slide. FIG. 3A is a map of the geographic distribution of hepatitis B virus genotypes worldwide. FIG. 3B provides a summary of a base editing strategies for introducing stop codons in viral genes and for generating abasic sites to treat chronic HBV.

FIG. 4 provides a bar graph presenting % A to G editing efficiencies for the indicated guide RNA sequences used in combination with the indicated base editors. The HEK293 lenti HBV system was used and an RNA transfection format. Some of the guide RNA sequences were associated with editing efficiencies of over 50%. The guide RNAs targeted HBV transcriptional elements or introduced missense mutations in HBV genes. Listed on the x-axis are the guide RNA sequences used in combination with the indicated base editors (see **Table 1**).

FIGs. 5A-5C provide bar graphs showing antiviral efficacy of the indicated base editors used in combination with the indicated guide RNAs (listed on x-axis) in the HBV-primary human hepatocytes (HBV-PHH) system. In FIGs. 5A-5C, the term “HBsAG” represents

“Hepatitis B surface antigen,” and the term “HBeAg” represents “Hepatitis B e antigen.” In FIG. 5, the y-axis represents fold-change relative to the no treatment sample.

FIG. 6 provides a schematic diagram outlining an experimental procedure for a functional gRNA screen in HBV-infected HepG2-NTCP cells.

FIG. 7 provides a schematic map of the HBV genome. The genome map is annotated to indicate the location of base edits associated with gRNA37 and gRNA40 used in combination with BE4.

FIG. 8 provides a collection of bar graphs showing the results of a functional gRNA screen in HBV-infected HepG2pNTCP cells. Two lead gRNAs were identified targeting HBsAg and Core genes (i.e., gRNA37 and gRNA40). The guide gRNA37 (Stop-S) was associated with a reduction in HBsAg and gRNA40 (Stop-PreCore) was associated with a reduction in HBeAg. The guide RNAs gRNA37 and gRNA40 both were associated with a lowering of the total HBV DNA and 3.5 kb RNA levels. Listed on the x-axis are the gRNAs used in combination with BE4 (i.e., gRNA12, gRNA37, gRNA19, gRNA190, and gRNA40). The negative control (-) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIG. 9 provides a collection of bar graphs showing that multiplexing gRNAs simultaneously reduced respective HBV viral parameters in HepG2-NTCP. Listed on the x-axis are the gRNAs used in combination with BE4 (i.e., gRNA37 and gRNA40). Both gRNA's targeted the same covalently-closed circular DNA strand (cccDNA) and the targeted sequences were spaced by about 1 kb. Multiplexing gRNA37 and gRNA40 reduced HBsAg and HbBeAg, as well as total HBV DNA and 3.5 kb RNA. The base editor used was BE4. The negative control (-) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIG. 10 provides bar graphs showing that base editor systems containing BE4 and the guide RNA(s) carried out cccDNA editing without reducing the level of cccDNA. The negative control (-) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIG. 11 provides a schematic summarizing an experimental protocol for evaluating the impact of lamivudine pretreatment on editing.

FIG. 12 provides a bar graph showing that pretreatment with lamivudine was associated with increased editing efficiencies. Base editing efficiency was improved by about 20% in HepG2-NTCP cells through pretreatment with lamivudine. High cccDNA editing in the lamivudine pretreated cells suggested that CBE directly targeted cccDNA.

FIG. 13 provides a set of bar graphs showing that base editing using gRNA37 in combination with gRNA40 (i.e., a combinatorial/multiplex gRNA treatment) combined with lamivudine pretreatment was associated with a robust reduction of Hepatitis B virus (HBV)

markers. The negative control (-) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIG. 14 provides a schematic diagram summarizing an experimental protocol for assessing antiviral activity of base editing relative to lamivudine. FIG. 14 lists the HBV parameters that were measured.

FIG. 15 provides a set of bar graphs demonstrating that base editing using gRNA37 in combination with gRNA40 (i.e., a combinatorial/multiplex gRNA treatment) resulted in a reduction in all HBV viral markers by about 70% to about 80% at day 25 following treatment; whereas, treatment for 8 days with lamivudine (LAM) was not. The negative control (-) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIG. 16 provides a plot presenting a time-course of extracellular HBV DNA levels for the indicated treatments. Base editing prevented rebound in primary hepatocyte co-cultures. HBV rebounded after discontinuation of lamivudine-only treatment. The control gRNA targeted the PCSK9 gene. The “HBV” sample was an untreated sample. There was no observed HBV rebound 2 weeks following the 2nd transfection with base editing reagents. HBV replication was assessed by HBV DNA qPCR in s/n.

FIG. 17 provides a bar graph showing the individual base editing efficiencies at the sites targeted by gRNA37 and gRNA40 when the two gRNAs were used together in combination with BE4 to edit HBV cccDNA. About 30% editing efficiency was observed at the S antigen site and about 60% editing efficiency was observed at the PreCore gene. These editing efficiencies were sufficient to enable high antiviral efficacy and prevent rebound in primary human hepatocytes (PHH).

FIG. 18 provides a schematic providing an overview of an experiment evaluating combined treatment with lamivudine and a base editing system comprising BE4 and gRNAs 37 and 40.

FIG. 19 provides a plot presenting a time-course of extracellular HBV DNA levels for the indicated treatments. The “HBV” sample was an untreated sample. Pretreatment with lamivudine improved editing in primary hepatocytes. Combined treatment with the base editor system and lamivudine prevented HBV rebound. HBV replication was assessed by HBV DNA qPCR in s/n.

FIG. 20 provides a bar graph showing base editing efficiencies at the sites targeted by gRNA37 and gRNA40. Cells were contacted with BE4 in combination with gRNA37 and gRNA40 and also treated with lamivudine (LAM). Pretreatment with lamivudine improved editing in primary hepatocytes. Control cells (-) were not treated with lamivudine.

FIG. 21 provides bar graphs showing that base editing in combination with lamivudine treatment in primary hepatocytes resulted in a reduction of all HBV markers by about 70% to about 80%. The designation “HBV” indicates an untreated control.

FIG. 22 provides a bar graph showing base editing at the sites targeted by gRNA37 and gRNA40. Cells were contacted with BE4, ppApo, or rrA3F in combination with gRNA37 and gRNA40. Use of the ppApo and rrA3F base editors, which have reduced off-target activity and lower non-specific effects, in place of BE4 did not negatively alter base editing efficiency relative to BE4. All of the base editors were associated with comparable editing efficiencies.

FIG. 23 provides a bar graph showing that contacting cells with the base editor BE4, ppApo, or rrA3F in combination with gRNA37 and gRNA40 resulted in a reduction in HBsAg levels in the HBV-infected cells. Thus, all three base editor systems had antiviral activity. The negative control (HBV) was untreated cells and the control (contr) was gRNA targeting the PCSK9 gene.

FIGs. 24A-24C provide plots showing *in vivo* levels of hepatitis B virus (HBV) antigens measured in a replication-competent HBVcircle mouse model (n=5) following an injection at day 1 and, in some instances as indicated in the figure, another injection at day 14 of lipid nanoparticles containing a polynucleotide encoding BE4 and the guide RNAs gRNA37 and gRNA40. **FIG. 24A** plots mean % HBsAg as IU/ml over time. **FIG. 24B** plots HBeAg as PEI (Paul Elrich Institut international standard serum) U/ml over time. **FIG. 24C** plots HBV DNA as copies/ml over time. In **FIGs. 24A-24C**, the term “HM” refers to gRNA37-HM and gRNA40-HM containing scaffolds with “heavy modifications” (i.e., heavily modified guide RNAs), the term “SOC” indicates “standard of care,” and “SEM” indicates “standard error of the mean.” In **FIGs. 24A-24C**, “1x” indicates an injection only at day 1 (“LNP-1st inj.”) and “2x” indicates injections at days 1 (“LNP-1st inj.”) and 14 (“LNP 2nd inj.”).

FIGs. 25A-25C provide plots showing *in vivo* levels of hepatitis B virus (HBV) antigens measured in a replication-competent HBVcircle mouse model (n=5) following an injection at day 1 and, in some instances as indicated in the figure, another injection at day 14 of lipid nanoparticles containing a polynucleotide encoding BE4 and the guide RNAs gRNA37 and gRNA40. **FIG. 25A** plots mean % HBsAg as IU/ml over time. **FIG. 25B** plots HBeAg as PEI (Paul Elrich Institut international standard serum) U/ml over time. **FIG. 25C** plots HBV DNA as copies/ml over time. In **FIGs. 25A-25C**, the term “HM” refers to gRNA37-HM and gRNA40-HM containing scaffold with “heavy modifications” (i.e., heavy modification guides RNAs), the term “ETV” indicates “Entecavir,” and “SD” indicates “standard deviation.” In **FIGs. 25A-25C**,

“1x” indicates an injection only at day 1 (“LNP-1st inj.”) and “2x” indicates injections at days 1 (“LNP-1st inj.”) and 14 (“LNP 2nd inj.”).

FIGs. 26A and 26B provide schematics describing a strategy for gRNA selection and screening. This strategy involves targeting conserved HBV regions with a focus on *genotype D*, which is most prevalent worldwide, and for which both cellular and animal models exist. In silico gRNA selection was based on the gRNA’s potential to silence HBV genes. 100 gRNAs introducing stop codons in viral genes (NGG, NGA, NNNRRT PAMs) were designed. 24 conserved gRNAs were predicted to introduce missense mutations into the HBV genome. The final step involved detecting editing in an Hek293-Lenti-HBV cell line.

FIGs. 27A-27D show that base editing prevents HBV rebound in primary human hepatocytes (PHH). **FIG. 27A** provides a schematic showing an experimental schedule used for PHH. **FIG. 27B** is a graph showing HBV replication assessed by HBV DNA qPCR in PHH supernatant at different days post-transfection. Discontinuation of 3TC leads to HBV rebound, while base editing prevents this rebound. **FIG. 27C** is a graph showing that base editing leads to the efficient reduction of the HBsAg, HBeAg, 3.5kbRNA, and HBV DNA, and further improves HBV replication inhibition in PHH. **FIG. 27D** is a graph showing ~30% editing of S antigen and ~60% editing of a PreCore gene was sufficient to enable high antiviral efficacy and prevent rebound in PHH.

FIGs. 28A-28D show that Multiplexing two lead gRNAs reduces HBV parameters in hepatoma cell line (HepG2-NTCP). **FIG. 28A** shows an Experimental schedule. **FIG. 28B** includes a number of graphs showing that base editing leads to the efficient reduction of viral extracellular (HBsAg, HBeAg) as well as intracellular (3.5kbRNA, and HBV DNA) parameters. **FIG. 28C** provides a series of graphs showing that results observed for HBsAg, HBeAg and 3.5kbRNA upon 3TC treatment. **FIG. 28D** shows that an gRNA B + gRNA H combination also inhibited all HBs isoforms as observed in Western blotting.

FIGs. 29A-29C show that Base editing reduced HBs from naturally integrated HBV. **FIG. 29A** provides an experimental protocol used for PLC/PRF5 cells. **FIG. 29B** is a graph showing that extracellular HBs Ag levels were determined by ELISA at 6th day post-transfection of BE4 mRNA and gRNA B*. **FIG. 29C** is a graph showing that ~50% editing of S antigen site was sufficient to enable robust reduction of HBs Ag.

FIGs. 30A and 30B shows that the base editor functions through cccDNA editing without reducing cccDNA level. **FIG. 30A** is a graph. cccDNA levels were assessed by qPCR on the DNA samples pretreated with ExoI/III to eliminate viral replicative intermediates. No cccDNA reduction was observed upon base editing by gRNA B + gRNA H in absence or

presence of 3TC. **FIG. 30B** is graph showing functional editing assessed by NGS on cccDNA enriched samples. Higher editing efficiency was detected in the presence of 3TC.

FIGs. 31A-31E shows that multiplexing two gRNAs with BE4 base editor simultaneously reduced HBV viral parameters in HepG2-NTCP. **FIG. 31A** provides a series of graphs showing that gase editing leads to the efficient reduction of viral extracellular (HBsAg, HBeAg), as well as intracellular (3.5kbRNA, and HBV DNA) parameters relative to a control sample treated with base editing reagents targeting the unrelated PCSK9 gene.

BE4/gRNAs(S1+C2) treatment inhibited all HBs isoforms, as observed in Western blotting.

FIG. 31B provides an experimental schedule in case of pretreatment with lamivudine. **FIG. 31C**

shows that combinatorial treatment with lamivudine lead to the robust reduction of HBV viral markers, similarly to the results shown in panel A. **FIG. 31D** is a graph showing that base editing does not reduce cccDNA level in HepG2-NTCP. **FIG. 31E** shows that pretreatment with Lamivudine increased base editing rates by 20% in HepG2-NTCP. Without intending to be bound by theory high cccDNA editing in Lamivudine pretreated conditions indicates that CBE likely directly targets cccDNA.

FIGs. 32A-32C show that base editing prevents viral rebound in PHH. **FIG. 32A** is a graph showing HBV replication assessed by HBV DNA qPCR in PHH supernatant.

Discontinuation of lamivudine lead to HBV rebound, while base editing prevented viral rebound.

FIG. 32B includes four graphs showing that base editing lead to the efficient reduction of the HBsAg, HBeAg, 3.5kbRNA, and HBV DNA. **FIG. 32C** is a graph showing ~55% editing S antigen and ~80% Editing PreCore gene sufficient to enable high antiviral efficacy and prevent rebound in PHH.

FIG. 33A- 33C show that LNP-mediated delivery of BE4 mRNA and gRNAs S1/C2 lead

to sustained reduction of viral markers in HBV minicircle model. **FIG. 33A** is a graph showing

that there was a > 2 log mean HBsAg reduction. Significantly, 5/9 mice showed HBsAg reduction below the limit of detection. **FIG. 33B** is a graph showing that HBV rebounds in

entecavir treated group (positive control). There was a >3 log sustained reduction in serum HBV DNA in groups treated with base editing. No HBV rebound was observed in the base edited

group. **FIG. 33C** shows a loss of HBeAg expression in all mice below the limit of detection two

weeks after the 1st LNP injection. Data represented as mean +/- SEM, n=4 or 5 per group.

FIGs. 34A and 34B are graphs showing that BE4 in combination with gRNA EMSbeam12 with spacer sequence SEQ ID NO. 578 and MSPbeam37-PLC with spacer sequence SEQ ID NO. 10834 edited the naturally integrated HBV genome present in the Alexander hepatoma cell line, PLC/PRF/5 (**FIG. 34A**) and this reduced HBsAg secretion (**FIG.**

34B). Also included are protospacer sequences: MSPbeam37 protospacer corresponds to SEQ ID NO. 508; MSPbeam37-PLC protospacer (SEQ ID NO. 10833) includes a single nucleobase change relative to SEQ ID NO. 508.

FIG. 35 provides a schematic showing an experimental design for evaluating base editing in HepG2.2.15 cells *in vitro* with and without pretreatment with lamivudine (LAM). In **FIG. 35**, “D0,” “D1,” etc. represent days 0, 1, etc. following commencement of the experiment, “SN” indicates “supernatant,” and “6dpt” indicates “six (6) days post-transduction.”

FIGs. 36A and 36B provide bar graphs showing HBsAg and HBeAg polypeptide levels in HepG2.2.15 cells following base editing using a BE4 base editor in combination with the guides gRNA37 (g37) or gRNA12 (g12) (for base-editing to reduce expression of HBsAg), and/or in combination with the guide gRNA40 (g40) (for base-editing to reduce expression of HBeAg). As a control, cells were contacted with the BE4 base editor in combination with a guide targeting the *PCSK9* gene. Cells were base edited with and without pre-treatment with lamivudine (LAM). **FIG. 36A** shows HBsAg expression levels in HepG2.2.15 cells edited with and without pre-treatment with LAM using the BE4 base editor in combination with the indicated guides. **FIG. 36B** shows HBeAg expression levels in HepG2.2.15 cells base edited with and without pre-treatment with LAM using the BE4 base editor in combination with the indicated guides. Expression levels were compared using ANOVA/non-parametric test / no matching pair/ multiple comparisons to *PCSK9*.

DETAILED DESCRIPTION OF THE INVENTION OF THE DISCLOSURE

The invention of the disclosure features compositions and methods for editing the HBV genome. For example, the compositions contemplated herein can, in some embodiments, include a base editor a guide nucleic acid that targets a particular nucleotide in an HBV gene. In some embodiments, the editing introduces a premature stop codon in the coding sequence of one of the viral proteins. In another embodiment, the editing introduces one or more substitutions (e.g., a missense mutation) in the coding sequence of one or more HBV proteins. In an embodiment, the editing alters a nucleobase in a transcription element (e.g., polyA site, enhancer, or promoter).

Recent studies have shown that a newer generation of ABEs (e.g., ABE8) are able to induce higher editing rates. Further, they have lower gRNA-independent off-target editing relative to the early generation cytidine base editors (CBEs), and thus are of interest as potential

therapeutic approaches. ABEs do not enable generation of stop codons, but they allow targeting AT-rich regions in the HBV genomes.

The use of base editing to treat HBV advantageously prevents HBV rebound by safely introducing permanent mutations in cccDNA and irreversibly silences HBsAg expression from the integrated HBV DNA without DSBs.

As described in more detail below, the present disclosure provides for targeting the established HBV covalently closed circular DNA (cccDNA) pool using cytosine base editors (CBEs, C to T conversions). Infected HepG2-NTCP cells and long-term primary human hepatocyte cultures (PHHs) were co-transfected with selected HBV-targeting gRNAs and mRNA encoding a CBE. Base editing efficiency was evaluated by DNA amplicon sequencing and the consequence of base editing on viral replication was determined by analyzing different viral parameters. Without affecting cccDNA integrity, base editing introduced non-sense mutations in HBs or HBe/HBc open reading frames (ORFs) and efficiently inhibited release of total HBV DNA and HBs/HBe antigens. Further, base editing rates remained high in the context of pretreatment with nucleoside analogs (NAs), which reduced levels of HBV DNA replicative intermediates, indicating that CBEs can directly target cccDNA. Multiplexing two gRNAs with CBE led to a greater reduction of HBsAg, HBeAg, total HBV DNA and 3.5kb viral pregenomic RNA. Importantly, dual gRNA/CBE treatment prevented HBV rebound. Altogether, these results show that CBEs can directly target cccDNA and generate non-sense mutations in HBs and HBe/HBc ORFs that interfere with HBV replication. Moreover, these effects are sustained after CBE degradation, indicating permanent functional changes in HBV cccDNA.

Targeting cccDNA transcriptional activity using base editing

Regions involved in cccDNA transcriptional activity are AT-rich regions and a surrogate to cccDNA degradation for an HBV cure is the permanent disruption of cccDNA transcriptional activity, which is currently unachievable.

One example of a region involved in cccDNA transcriptional activity is the polyadenylation signal (PAS), which is unique for all the HBV viral RNAs and serves also as a regulatory element (TATA box). Targeting this region in cccDNA can destabilize all viral RNA species, including its genomic component pgRNA and lead to cccDNA silencing.

As described in the Examples provided herein, sgRNAs were identified through in-silico analyses that target regulatory regions in cccDNA (e.g., Enhancer I, Enhancer II, Basal Core promoter, Cryptic and canonical polyadenylation signals (PAS), S and X genes promoters).

Over 250 sgRNAs were identified, and these candidates were curated and several candidates with high in silico predicted efficiency were tested in a Hek293-lentiHBV system.

Some aspects of the present disclosure, as described further in the Examples, include selecting highly conserved ABE gRNAs predicted to introduce missense mutations in HBV genes. For example, gRNAs were selected based on the high conservation across HBV genotypes, with a focus on genotypes D and A. High sequence conservation allows for targeting of a wider patient population and also is necessary to avoid appearance of viral mutations that would enable HBV virus to escape base editing treatment. Another example of selection criteria was ability of the gRNAs to generate edits in the HBV genome that would lead to a missense mutation. In silico analysis was performed for amino acids changes generated by the base editing reagents gRNAs were selected that would generate missense amino acid changes with sequences rarely detected in naturally occurring HBV genotypes (<0.05% frequency aa change). This way the likelihood that a gRNAs would introduce a mutation that led to a disruption of the HBV protein/genome was increased.

HBV genome

The HBV genome is about 3.2 kb of partially double-stranded DNA and open read frames (ORFs) encoding seven proteins. Referring to **FIG. 1**, the open reading frame (ORF) P encodes the viral polymerase. The ORF C/PreC encodes capsid proteins. ORF PreS1, ORF PreS2, and ORF S encode large (L), middle (M) and small (S) surface proteins, respectively. ORF X encodes the secretory X protein.

The partially double-stranded HBV genome is converted by host factors to covalently closed circular DNA (cccDNA). The cccDNA is transcribed by a host RNA polymerase to produce viral mRNA including pre-genomic RNA (pgRNA). pgRNA is reversed transcribed by the HBV polymerase into genomic HBV DNA that can be converted into cccDNA, packaged into virions, or integrated into the host cell's genome (**FIG. 2**). cccDNA, a key component of the HBV life cycle, is a stable molecule responsible for chronic HBV infection. Editing of the HBV genome can disrupt the formation of cccDNA, thereby reducing the pathogenicity of the virus.

There are ten different HBV genotypes (A-J) (**FIG. 3A**). A "genotype" is characterized by < 92% sequence identity with any other genome, and a sub-genotype is characterized by < 96 to 92% sequence identity. HBV of genotype D is the most prevalent in the United States (**FIG. 3A**). Research models of HBV genotype D are available including viral stocks (e.g., genotype D, subgenotype ayw (Imquest)) and mouse models (e.g., humanized mouse model (Phoenixbio)).

EDITING OF TARGET GENES

The invention of the disclosure provides methods and compositions that target HBV ORFs for editing. These compositions can comprise a nucleobase editor having a Cas9 or other nucleic acid programmable DNA binding protein domain and an adenosine or cytosine deaminase domain. In some embodiments, the base editor introduces one or more alterations into an HBV ORF. In some embodiments, the alteration results in a mutation in a conserved portion of an HBV protein. In particular embodiments, the alteration introduces one or more stop codons. Throughout the specification, the introduction of a stop codon, resulting in the premature termination of the protein is represented by the amino acid symbol, the amino acid position, and the term STOP (e.g., R87STOP indicates that the codon encoding Arginine at amino acid position 87 is replaced by a Stop codon). Advantageously, the methods of the present invention do not introduce double stranded breaks in the HBV genome. The invention provides strategies for using base editing to treat chronic HBV (**FIG. 3B**). Introducing stop codons into viral genes using the methods and compositions described herein can be accomplished without generating double strand breaks, thereby eliminating or reducing the risk of cutting host genetic material after HBV integrates into the host's genome. Additionally, the compositions can employ a deaminase that is a natural HBV antiviral restriction factor. For example, inducing APOBEC cytidine deaminases with interferon alpha or Lymphotoxin β receptor (LTBR) promotes abasic site formation and cccDNA degradation (**FIG. 3B**). Furthermore, using a base editor without uracil glycosylase inhibitor domains can target cellular uracil glycosylase to cccDNA and promote its degradation.

In embodiments HBV ORFs are edited using a base editor comprising an adenosine deaminase (i.e., an adenosine base editor (ABE)). ABE's have several advantages including robust on-target editing and low gRNA-independent off-target editing relative to first generation rAPOBEC1-based CBEs. ABE's do not induce a uracil N-glycosylase (UNG) response. A>I cccDNA deamination has not been described for HBV. ABE's can be used to edit the polymerase active site and/or cccDNA transcription sites including the RNA polyadenylation site TATAAA common to HBV transcripts. Vital regions involved in cccDNA transcriptional activity are AT-rich regions and a surrogate to cccDNA degradation.

In some aspects, methods and compositions are provided for editing HBV cccDNA with a base editor comprising a cytidine deaminase or adenosine deaminase domain. Exemplary guide RNAs are provided in the following **Table 1**. Further exemplary guide RNAs include those guide RNAs containing spacer sequences provided in any of Tables 1, 2A, 2B,

and/or 2C or listed in the Sequence Listing (e.g., at SEQ ID NOs: 3105-5485 and 8220-10830).

In various embodiments, multiple guide RNAs can be administered to a subject or cell simultaneously, optionally in combination with one or more nucleobase editor polypeptides. In embodiments, the multiple guide RNAs include 2, 3, 4, 5, 6, 7, 8, 9, 10, or more guides,

optionally wherein the guide RNAs are selected from two or more of cas12b-4, cas12b-5, cas12b-11, cas12b-17, cas12b-30, cas12b-42, cas12b-100, cas12b-147, cas12b-154, cas12b-35, cas12b-124, cas12b-127, and cas12b-122b (see **Table 1**). In embodiments, the multiple guide

RNAs administered to a subject or cell contain two guide RNAs (e.g., gRNA37 and gRNA40, see **Table 1**). The guides can be administered simultaneously or sequentially, optionally with

one or more nucleobase editor polypeptide(s) (e.g., a base editor(s) encoded by an mRNA molecule(s)). Non-limiting examples of spacer sequences suitable for use in guide RNA molecules of the present disclosure include any spacer sequence targeting any portion of a Hepatitis B virus genome. Such spacer sequences can be designed and selected using methods

available to the skilled practitioner (e.g., the design of gRNAs targeting a particular sequence is available from Invitrogen as a commercial service). Various strategies can be used for gRNA

selection including, as non-limiting examples, 1) targeting HBV transcription elements, and 2) selecting highly conserved ABE gRNAs predicted to introduce missense mutations in HBV

genes. In embodiments, the guide RNAs can be used to target a base editor for introduction of a missense mutation into an HBV gene (e.g., cas12b-4, cas12b-5, cas12b-11, cas12b-17, cas12b-

30, cas12b-42, cas12b-100, cas12b-147, or cas12b-154). The guide RNAs listed in **Table 1** or any guide RNA comprising a spacer sequence listed in **Table 2A**, **Table 2B**, **Table 2C**, or listed among SEQ ID NOs: 3105-5485 and 8220-10830 or anywhere in the Sequence Listing can also

be used for targeting a base editor for altering a nucleobase at a transcription site (e.g., cas12b-35, cas12b-124, cas12b-127, or cas12b-122b). Transcription sites that can be targeted for base

editing by the guide RNAs of **Table 1** or any guide RNA comprising a spacer sequence listed in **Table 2A**, **Table 2B**, **Table 2C**, or listed anywhere in the sequence listing (e.g., among SEQ ID

NOs: 3105-5485 and 8220-10830) include enhancer II box A (e.g., cas12b-35), enhancer I (e.g., cas12b-124 or cas12b-127), and the HBX promoter (e.g., cas12b-122b). Non-limiting examples

of transcription sites include basal core promoter (BCP), enhancer I, enhancer II, HBV polyA

site, HBX promoter, and S promoter-CCAATC-region II. The guide RNAs of the present

invention can be used in combination with any base editor (e.g., Cas12b-ABE (ABE-bhCas12b); SEQ ID NO: 418). The base editor can target the PAM sequences listed in **Table 2A** or **Table**

2B below (e.g., RTTN sequences).

In embodiments two or more guide RNAs are introduced to a subject, optionally simultaneously. In embodiments, the two or more guide RNAs include cas12b-5 and cas12b-4; cas12b-11 and cas12b-4; cas12b-17 and cas12b-4; cas12b-30 and cas12b-4; cas12b-42 and cas12b-4; cas12b-100 and cas12b-4; cas12b-147 and cas12b-4; cas12b-154 and cas12b-4; cas12b-35 and cas12b-4; cas12b-124 and cas12b-4; cas12b-127 and cas12b-4; cas12b-122b and cas12b-4; cas12b-11 and cas12b-5; cas12b-17 and cas12b-5; cas12b-30 and cas12b-5; cas12b-42 and cas12b-5; cas12b-100 and cas12b-5; cas12b-147 and cas12b-5; cas12b-154 and cas12b-5; cas12b-35 and cas12b-5; cas12b-124 and cas12b-5; cas12b-127 and cas12b-5; cas12b-122b and cas12b-5; cas12b-17 and cas12b-11; cas12b-30 and cas12b-11; cas12b-42 and cas12b-11; cas12b-100 and cas12b-11; cas12b-147 and cas12b-11; cas12b-154 and cas12b-11; cas12b-35 and cas12b-11; cas12b-124 and cas12b-11; cas12b-127 and cas12b-11; cas12b-122b and cas12b-11; cas12b-30 and cas12b-17; cas12b-42 and cas12b-17; cas12b-100 and cas12b-17; cas12b-147 and cas12b-17; cas12b-154 and cas12b-17; cas12b-35 and cas12b-17; cas12b-124 and cas12b-17; cas12b-127 and cas12b-17; cas12b-122b and cas12b-17; cas12b-42 and cas12b-30; cas12b-100 and cas12b-30; cas12b-147 and cas12b-30; cas12b-154 and cas12b-30; cas12b-35 and cas12b-30; cas12b-124 and cas12b-30; cas12b-127 and cas12b-30; cas12b-122b and cas12b-30; cas12b-100 and cas12b-42; cas12b-147 and cas12b-42; cas12b-154 and cas12b-42; cas12b-35 and cas12b-42; cas12b-124 and cas12b-42; cas12b-127 and cas12b-42; cas12b-122b and cas12b-42; cas12b-147 and cas12b-100; cas12b-154 and cas12b-100; cas12b-35 and cas12b-100; cas12b-124 and cas12b-100; cas12b-127 and cas12b-100; cas12b-122b and cas12b-100; cas12b-154 and cas12b-147; cas12b-35 and cas12b-147; cas12b-124 and cas12b-147; cas12b-127 and cas12b-147; cas12b-122b and cas12b-147; cas12b-35 and cas12b-154; cas12b-124 and cas12b-154; cas12b-127 and cas12b-154; cas12b-122b and cas12b-154; cas12b-124 and cas12b-35; cas12b-127 and cas12b-35; cas12b-122b and cas12b-35; cas12b-127 and cas12b-124; cas12b-122b and cas12b-124; or cas12b-122b and cas12b-127 (See **Table 1**).

Table 1 Guide RNAs

Guide ID	SEQ ID NO	gRNA sequence
cas12b-4	379	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC ACAAGAAUCCUCACAAUACC
cas12b-5	380	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC UCGCUUGGAUGUGUCUGCGGC
cas12b-11	381	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAGCAC CACCUGUAUUCUCAUCCCAU

cas12b-17	382	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC GGGGCCAAGUCUGUACAGCA
cas12b-30	383	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC GCUGACGCAACCCCCACUGG
cas12b-42	384	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC GGAGCUACUGUGGAGUUACU
cas12b-100	385	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC UUCUUCUAGGGGACCUGCCU
cas12b-147	386	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC GAGGACAAACGGGCAACAU
cas12b-154	387	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC UUGUCAACAAGAAAAACCCC
cas12b-35	388	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC CCCAAGGUCUUACAUAGAG
cas12b-124	389	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC CCGGGCAACGGGGUAAAGGU
cas12b-127	390	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC ACACAGAAAGGCCUUGUAAG
cas12b-122b	391	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGU GUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUAAGCAC CACGCACGCGCUGAUGGCC
E12 (gRNA12)	419	mG*mA*mC*UUCUCUCAUUUUUCUAGGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
M52 (gRNA52)	420	mU*mC*mA*AUCCCAACAAGGACACCGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
E20 (gRNA20)	421	mU*mC*mC*UCUGCCGAUCCAUAUCUGGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
M40 (gRNA40)	422	mC*mC*mA*UGCCCCAAAGCCACCCAGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
M190 (gRNA190)	423	mG*mC*mU*GCCAACUGGAUCCUGCGGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
M37 (gRNA37)	424	mG*mA*mA*AGCCCAGGAUGAUGGGAGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
E19 (gRNA19)	425	mU*mC*mC*GCAGUAUGGAUCGGCAGGUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUmU*mU*mU*
G-38	5486	mA*mG*mA*AAGGCCUUGUAAGUUGG GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*
G-231	5487	mU*m*G*mACGCAACCCCCACUGGCU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*

G-159	5488	mG*mC*mA*AACACUUGGCACAGACC GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-255	5489	mU*mU*mU*GCUGACGCAACCCCCAC GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-140	5490	mG*mA*mC*GCAACCCCCACUGGCUG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-15	5491	mA*mA*mC*CCCCACUGGCUGGGGU GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-23	5492	mA*mA*mU*GUCAACGACCGACCUUG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-234	5493	mU*mG*mC*CCAAGGUCUUAACAUAAG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-6	5494	mA*mA*mA*GAAUUGGAGCUACUGU GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-48	5495	mA*mG*mA*GAGGUGCGCCCCGUGGU GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-150	5496	mG*mA*mG*GUGAAGCGAAGUGCACA GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-185	5497	mG*mG*mG*GCGCACCCUCUCUUUACG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-200	5498	mG*mU*mG*AAAAAGUUGCAUGGUGC GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-polyA_NGC 1	5499	mG*mA*mC*CCUUAUAAAGAAUUUGG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-polyA_NGC 2	5500	mU*mU*mU*AUAAGGGUCGAUGUCCA GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-772	5501	mU*mG*mG*CGAUUGGUGGAGGCAGG GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-221	5502	mC*mA*mC*AGAAAGGCCUUGUAAGU GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-412	5503	mC*mU*mU*UCUGUGUAAACAAUACC GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-251	5504	mC*mA*mG*GUUUUUUACACAGAA GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-839	5505	mU*mU*mU*ACCCCGUUGCCCGGCAA GUUUUAGAGCUAGAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U

G-553	5506	mG*mG*mG*CAACGGGGUAAAGGUUC GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-638	5507	mG*mU*mU*GCCCCAAGGUCUUACUA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-845	5508	mU*mU*mU*GAAGUAUGCCUCAAGGU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-474	5509	mG*mA*mU*UAAAGGUCUUUGUACUA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-804	5510	mU*mU*mA*AAGGUCUUUGUACUAGG GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
G-144 (EMS3)	5511	mA*mG*mG*AGGUCUAGGCAUAAAA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-8	5512	mA*mA*mA*UUGAGAGAAGUCCACCA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-41	5513	mA*mG*mA*AGAUGAGGCAUAGCAGC GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-85 (MSPbeam3 6)	5514	mC*mA*mC*CACGAGUCUAGACUCUG GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-131 (EMS11)	5515	mG*mA*mA*GAACUCCCUCGCCUCGC GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-133	5516	mG*mA*mA*GAUGAGGCAUAGCAGCA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-142 (MSPbeam6 6)	5517	mG*mA*mG*AAGUCCACCACGAGUCU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-150	5518	mG*mA*mG*GUGAAGCGAAGUGCACA GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
ABE-235 (EMS21)	5519	mU*mG*mG*ACUUCUCUCAUUUUUCU GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U
cas12b-116	5520	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACAGCGCCGACGGGACGUAAAC
cas12b-122a	5521	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACCGUCAGCAAACACUUGGCAC
cas12b-148	5522	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACGAGGACAGGAGGUUGGUGAG
cas12b-152	5523	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACAGAGAAGUCCACCACGAGUC

cas12b-111	5524	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACUUUAUAAGGGUCGAUGUCCA
cas12b-125	5525	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACAGGUAAUUGUUUACACAGAAA
cas12b-126	5526	mG*mU*mU*CUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGC AGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCUCUUACGAGGCAUUG CACUUUACACAGAAAGGCCUUGU
gRNA37- HM	424	gsasasAGCCCAGGAUGAGGGAgUUUAGagcuagaaauagcaaGUUaA aAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucguguc sususu
gRNA40- HM	422	cscsasUGCCCCAAAGCCACCCAgUUUAGagcuagaaauagcaaGUUaA aAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucguguc sususu

For all guides except gRNA37-HM and gRNA40-HM, mA*, mC*, mG*, and mU* represent bases with a 2'-O-methyl 3'-phosphorothioate modification. For gRNA37-HM and gRNA-40HM, a, c, u, or g represents bases with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl 3'-phosphorothioate (MS) modification.

5

To produce the gene edits described above, cells (e.g., a hepatocyte) are contacted with two 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 contacted with at least one nucleic acid, wherein the at least one nucleic acid encodes two 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 nucleotide analogs (e.g., mA*, mC*, mG*, and/or mU*). These nucleotide analogs can inhibit degradation of the gRNA from cellular processes. **Tables 2A-2C** provide exemplary target and spacer sequences to be used for gRNAs. Further exemplary target sequences and spacer sequences are provided in the Sequence Listing as SEQ ID NOs: 724-3104 and 5609-8219, and SEQ ID NOs: 3105-5485 and 8220-10830, respectively. The gRNAs can target all or a portion of all HBV genotypes (e.g., A, B, C, D, E, F, G, and/or H). The gRNAs can target over 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 85%, or all of hepatitis B viruses of a particular genotype or set of genotypes.

In various embodiments, a gRNA (e.g., gRNA37-HM or gRNA40-HM) of the present disclosure contains the following “heavily modified” (HM) scaffold:
gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugGcaccgagucg
gugcusususu (SEQ ID NO: 317), where “a”, “c”, “u”, or “g” represents an A, C, U, or G
nucleotide containing a 2'-O-methyl (M) modification, respectively; and as, cs, us, or gs represent
an A, C, U, or G nucleotide containing a 2'-O-methyl 3'-phosphorothioate (MS) modification,

respectively. In embodiments, a guide RNA of the present disclosure contains the HM scaffold and a spacer sequence listed in any one of **Tables 2A-2C**.

Table 2A: Exemplary Target Sequences

Guide ID	Target Sequence	gRNA Spacer Sequence	PAM Site	HBV gene targeted	Location of Target Site in HBV Genome		
					Start	End	Strand
cas12b-4	ACAAGAATCCTCACAATACC (SEQ ID NO: 392)	ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405)	GTTG	Pol/S	218	242	+
cas12b-5	TCGCTGGATGTGTCTGCGGC (SEQ ID NO: 393)	UCGCUUGGAUGUGUCUGCGGC (SEQ ID NO: 406)	GTTA	Pol/S	367	391	+
cas12b-11	CACCTGTATTCCCATCCCAT (SEQ ID NO: 394)	CACCUGUAUUCCTCAUCCCAU (SEQ ID NO: 407)	ATTG	Pol/S	592	616	+
cas12b-17	GGGGCCAAGTCTGTACAGCA (SEQ ID NO: 395)	GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408)	ATTG	Pol/S	754	778	+
cas12b-30	GCTGACGCAACCCCCACTGG (SEQ ID NO: 396)	GCUGACGCAACCCCCACUGG (SEQ ID NO: 409)	GTTT	Pol	1183	1207	+
cas12b-42	GGAGCTACTGTGGAGTTACT (SEQ ID NO: 397)	GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410)	ATTT	Core	1925	1949	+
cas12b-100	TTCTTCTAGGGGACCTGCCT (SEQ ID NO: 398)	UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411)	GTTT	Pol	2357	2381	-
cas12b-147	GAGGACAAACGGGCAACATA (SEQ ID NO: 399)	GAGGACAAACGGGCAACAUA (SEQ ID NO: 412)	ATTG	Pol/S	461	485	-
cas12b-154	TTGTCAACAAGAAAAACCCC (SEQ ID NO: 400)	UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413)	ATTG	Pol/S	206	230	-
cas12b-35	CCCAAGGTCTTACATAAGAG (SEQ ID NO: 401)	CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414)	GTTG	X	1638	1662	+
cas12b-124	CCGGGCAACGGGGTAAGGT (SEQ ID NO: 402)	CCGGGCAACGGGGTAAGGU (SEQ ID NO: 415)	GTTG	Pol	1140	1164	-
cas12b-127	ACACAGAAAGGCCTTGTAAG (SEQ ID NO: 403)	ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416)	GTTT	Pol	1107	1131	-

cas12 b- 122b	CACGCACGCGCTGAT GGCCC (SEQ ID NO: 404)	CACGCACGCGCUGAU GGCCC (SEQ ID NO: 417)	GTT C	Pol	122 2	12 46	-

“Core” refers to the HBV core protein; “Pol” refers to the HBV polymerase gene; “S” refers to the HBV surface protein; “X” refers to the HBV protein X gene.

Table 2B: Exemplary Target Sequences

Guide ID	Target Sequence	Target SEQ ID NO	gRNA Spacer	Spacer SEQ ID NO	PAM Site	HBV gene target
G-38	AGAAAGGCCCTTGTAAGTTGG	5527	AGAAAGGCCCUUGUAAGUUGG	5568	NGA	Pol
G-231	TGACGCAACCCCCACTGGCT	5528	UGACGCAACCCCCACUGGCU	5569	NGG	Pol
G-159	GCAAAACACTTGGCACAGACC	5529	GCAAAACACUUGGCACAGACC	5570	NGG	Pol
G-255	TTTGCTGACGCAACCCCCAC	5530	UUUGCUGACGCAACCCCCAC	5571	NGG	Pol
G-140	GACGCAACCCCACTGGCTG	5531	GACGCAACCCCAACUGGCU	5572	NGG	Pol
G-15	AACCCCACTGGCTGGGCT	5532	AACCCCAACUGGCU	5573	NGG	Pol
G-23	AATGTCAACGACCGACCTTG	5533	AAUGUCAACGACCGACCUUG	5574	NGG	X
G-234	TGCCCAAGGCTTTACATAAG	5534	UGCCCAAGGUCUUAACUAAAG	5575	NGG	X
G-6	AAAGAATTTGGAGCTACTGT	5535	AAAGAAUUUGGAGCUACUGU	5576	NGA	Core
G-48	AGAGAGGTGGCCCGCTGGT	5536	AGAGAGGUGCGCCCGUGGU	5577	NGG	Pol/X
G-150	GAGGTGAAGCGAAGTGCACA	5537	GAGGUAGAAGCGAAGUGCACA	5578	NGG	Pol/X
G-185	GGGGCGACCTCTCTTTACG	5538	GGGGCGACCCUCUUCUUUACG	5579	NGG	Pol/X
G-200	GTGAAAAAGTTGCATGGTGC	5539	GUGAAAAAGUUGCAUGGUGC	5580	NGG	X/Core
G-polyA NGC1	GACCCCTATAAAGAATTGG	5540	GACCCUUUAAAAGAAUUUGG	5581	NGC	Core
G-polyA NGC2	TTTATAAGGGTCGATGTCCA	5541	UUUAUAAAGGGUCGAUGUCCA	5582	NGC	Core
G-772	TGGCGATTGGTGGAGGCAGG	5542	UGGCGAUUGGUGGAGGCAGG	5583	NGG	Pol/S
G-221	CACAGAAAGGCCCTTGTAAAT	5543	CACAGAAAGGCCCUUGUAAAGU	5584	NGA	Pol
G-412	CTTCTGTGTAAACAATAACC	5544	CUUUCUGUGUAAAACAAUACC	5585	NGG	Pol
G-251	CAGGTATTGTTTACACAGAA	5545	CAGGUUUUUUACACAGAA	5586	NGG	Pol
G-839	TTTACCCCGTTGCCCGGCCAA	5546	UUUACCCCGUUGCCCGGCCAA	5587	NGG	Pol
G-553	GGGCAACGGGGTAAAGGTTT	5547	GGGCAACGGGGUAAAGGUUC	5588		Pol
G-638	GTGCCCCAAGGTCTTACATA	5548	GUUGCCCCAAGGUCUUACAUA	5589	NGA	X
G-845	TTTGAAGTATGCCCTCAAGGT	5549	UUUGAAGUAGCCUCAAGGU	5590	NGG	X
G-474	GATTAAAGGCTTTTGTTACTA	5550	GAUUAAAGGUCUUUUGUACUA	5591	NGA	X

G-804	TTAAAGGTCCTTTGTACTAGG	5551	UUAAAAGGUCUUUUGUACUAGG	5592	NGG	X
G-144 (EMS3)	AGGAGGCTGTAGGCATAAAT	5552	AGGAGGCTGUAGGCAUAAAU	5593	NGG	X
ABE-8	AAATTGAGAGAAAGTCCACCA	5553	AAAUUGAGAGAAAGUCCACCA	5594	NGA	S/Pol
ABE-41	AGAAAGATGAGGCATAGCAGC	5554	AGAAAGUAGAGGCAUAGCAGC	5595	NGG	S/Pol
ABE-85 (MSPbeam36)	CACCACGAGTCTAGACTCTG	5555	CACCACGAGUCUAGACUCUG	5596	NGG	S/pol
ABE-131 (EMS11)	GAAAGAACTCCCTCGCCCTCGC	5556	GAAGAAACUCCUCGCCUCGC	5597	NGA	Core/Pol
ABE-133	GAAGATGAGGCATAGCAGCA	5557	GAAGAUGAGGCAUAGCAGCA	5598	NGA	S/Pol
ABE-142 (MSPbeam66)	GAGAAAGTCCACCACGAGTCT	5558	GAGAAAGUCCACCACGAGUCU	5599	NGA	S/Pol
ABE-150	GAGGTGAAGCGAAGTGCACA	5559	GAGGUGAAGCGAAAGUGCACA	5600	NGG	Pol/X
ABE-235 (EMS21)	TGGACTTCTCTCAATTTTCT	5560	UGGACUUCUCUCAAUUUUCU	5601	NGG	S/Pol
cas12b-116	AGCGCCGACGGACGTAAAC	5561	AGCGCCGACGGACGUAAAC	5602	ATTG	Pol/X
cas12b-122a	CGTCAGCAAAACACTTGGCAC	5562	CGUCAGCAAAACACUUGGCAC	5603	GTTG	Pol
cas12b-148	GAGGACAGGAGGTTGGTGAG	5563	GAGGACAGGAGGUUGGUGAG	5604	GTTG	S/Pol
cas12b-152	AGAGAAGTCCACCACGAGTC	5564	AGAGAAAGUCCACCACGAGUC	5605	ATTG	S/Pol
cas12b-111	TTTATAAGGGTCGATGTCCA	5565	UUUAUAGGGUCGAUGUCCA	5606	ATTG	X/Core
cas12b-125	AGGTATTGTTACACACAGAAA	5566	AGGUAUUGUUUACACAGAAA	5607	GTTG	Pol
cas12b-126	TTTACACAGAAAAGGCCCTTGT	5567	UUUACACAGAAAAGGCCUUGU	5608	ATTG	Pol

“Core” refers to the HBV core protein; “Pol” refers to the HBV polymerase gene; “S” refers to the HBV surface protein; “X” refers to the HBV protein X gene.

Table 2C: Exemplary Target Sequences

Guide ID	Target Sequence	Target SEQ ID NO	gRNA Spacer Sequence	Spacer SEQ ID NO
E1	AAGGCACAGCTTGGAGGCTTG	426	AAGGCACAGCUUGGAGGCUUG	575
E10	GAACTCCCTCGCCTCGCAGAC	427	GAACUCCCUCGCCUCGCAGAC	576
E11	GAAGAACTCCCTCGCCTCGC	428	GAAGAACUCCCUCGCCUCGC	577
E12 (gRNA12)	GACTTCTCTCAATTTTCTAG	429	GACUUCUCUCAUUUUUCUAG	578
E14	GCTGCTATGCCTCATCTTCTT	430	GCUGCUAUGCCUCAUCUUCUU	579
E15	GGACTTCTCTCAATTTTCTA	431	GGACUUCUCUCAUUUUUCUA	580
E16/M2	GTCCACCACGAGTCTAGACTC	432	GUCCACCACGAGUCUAGACUC	581
E18	TAAACGCGCGCAGACACATCC	433	UAAAACGCGCGCAGACACAUC	582
E19 (gRNA19)	TCCGCAGTATGGATCGGCAG	434	UCCGCAGUAUGGAUCGGCAG	583
E2	ACCAATTTATGCCTACAGCCT	435	ACCAAUUUAUGCCUACAGCCU	584
E20 (gRNA20)	TCCTCTGCCGATCCATACTG	436	UCCUCUGCCGAUCCAUAACUG	585
E21	TGGACTTCTCTCAATTTTCT	437	UGGACUUCUCUCAUUUUUCU	586
E22/M9	TTCAAGCCTCCAAGCTGTGCC	438	UUCAAGCCUCCAAGCUGUGCC	587
E23	TTTGCTGACGCAACCCCCAC	439	UUUGCUGACGCAACCCCCAC	588
E24	TACTAACATTGAGGTTCCCG	440	UACUAACAUAUGAGGUUCCCG	589
E3	AGGAGGCTGTAGGCATAAAT	441	AGGAGGCUGUAGGCAUAAAU	590
E4	AGGAGTTCCGCAGTATGGAT	442	AGGAGUUCGCGAGUAUGGAU	591
E6	CAAGGCACAGCTTGGAGGCT	443	CAAGGCACAGCUUGGAGGCU	592
E7	CCGCAGTATGGATCGGCAGA	444	CCGCAGUAUGGAUCGGCAGA	593
E8	CCTCTGCCGATCCATACTGC	445	CCUCUGCCGAUCCAUAACUGC	594
E9	CTAGACTCGTGGTGGACTTCT	446	CUAGACUCGUGGUGGACUUCU	595
E95	CGCCCACCGAATGTTGCCCA	447	CGCCCACCGAAUGUUGCCCA	596
E96	CTCCTCCAGTCTTTAAACA	448	CUCCUCCCAGUCUUUAAACA	597
E96	ACTCCTCCAGTCTTTAAACA	449	ACUCCUCCCAGUCUUUAAACA	598
M1	CAATACCGCAGAGTCTAGACT	450	CAAUACCGCAGAGUCUAGACU	599
M1	AATACCGCAGAGTCTAGACT	451	AAUACCGCAGAGUCUAGACU	600
M10	CCCAGCAAAGAATTGCTTGC	452	CCCAGCAAAGAAUUGCUUGC	601
M10	CCCAGCAAAGAATTGCTTGC	453	CCCAGCAAAGAAUUGCUUGC	602
M11	CCCACCCAGGTAGCTAGAGTC	454	CCCACCCAGGUAGCUAGAGUC	603
M11	CCACCCAGGTAGCTAGAGTC	455	CCACCCAGGUAGCUAGAGUC	604
M12	GAATCCACACTCCGAAAGACA	456	GAAUCCACACUCCGAAAGACA	605
M12	AATCCACACTCCGAAAGACA	457	AAUCCACACUCCGAAAGACA	606
M13	GTCTCAATCGCCGCGTCGCA	458	GUCUCAAU CGCCGCGUCGCA	607
M13	GGTCTCAATCGCCGCGTCGCA	459	GGUCUCAAU CGCCGCGUCGCA	608
M14	GAAGATCTCAATCTCGGGAAC	460	GAAGAUCUCAUUCUGGGAAC	609
M14	AAGATCTCAATCTCGGGAAC	461	AAGAUCUCAUUCUGGGAAC	610
M15	TTTTCCAATGAGGATTAAAG	462	UUUUCCA AUGAGGAUUAAG	611
M15	GTTTTCCAATGAGGATTAAAG	463	GUUUCCA AUGAGGAUUAAG	612
M16	TTTACACCAAGACATTATCA	464	UUUACACCAAGACAUUAUCA	613
M16	ATTTACACCAAGACATTATCA	465	AUUUACACCAAGACAUUAUCA	614
M17	TGAACAGTTTGTAGGCCCACT	466	UGAACAGUUUGUAGGCCCACT	615

Guide ID	Target Sequence	Target SEQ ID NO	gRNA Spacer Sequence	Spacer SEQ ID NO
M17	GAACAGTTTGTAGGCCCACT	467	GAACAGUUUGUAGGCCCACU	616
M18	GATTGCAATTGATTATGCCTG	468	GAUUGCAAUUGAUUAUGCCUG	617
M18	ATTGCAATTGATTATGCCTG	469	AUUGCAAUUGAUUAUGCCUG	618
M189	TGCCAACTGGATCCTGCGCG	470	UGCCAACUGGAUCCUGCGCG	619
M19	CGCCTTCCATAGAGTGTGTA	471	CGCCUCCAUAAGAGUGUGUA	620
M19	CCGCCTTCCATAGAGTGTGTA	472	CCGCCUCCAUAAGAGUGUGUA	621
M190 (gRNA190)	GCTGCCAACTGGATCCTGCG	473	GCUGCCAACUGGAUCCUGCG	622
M191	CTGCCAACTGGATCCTGCGC	474	CUGCCAACUGGAUCCUGCGC	623
M2	TCCACCACGAGTCTAGACTC	475	UCCACCACGAGUCUAGACUC	624
M20	TCCCAAGAATATGGTGACCCA	476	UCCCAAGAAUAUGGUGACCCA	625
M20	CCCAAGAATATGGTGACCCA	477	CCCAAGAAUAUGGUGACCCA	626
M21	ACAAGATCTACAGCATGGGG	478	ACAAGAUCUACAGCAUGGGG	627
M21	AACAAGATCTACAGCATGGGG	479	AACAAGAUCUACAGCAUGGGG	628
M22	CCAGCCTTCAGAGCAAACACA	480	CCAGCCUUCAGAGCAAACACA	629
M22	CAGCCTTCAGAGCAAACACA	481	CAGCCUUCAGAGCAAACACA	630
M23	TTCAGAGCAAACACAGCAAAT	482	UUCAGAGCAAACACAGCAAU	631
M23	TCAGAGCAAACACAGCAAAT	483	UCAGAGCAAACACAGCAAU	632
M24	TCCCCAATCCTCGAGAAGATT	484	UCCCCAAUCCUCGAGAAGAUU	633
M24	CCCCAATCCTCGAGAAGATT	485	CCCCAAUCCUCGAGAAGAUU	634
M25	CCAGGACAAGTTGGAGGACA	486	CCAGGACAAGUUGGAGGACA	635
M25	ACCAGGACAAGTTGGAGGACA	487	ACCAGGACAAGUUGGAGGACA	636
M26	TGCTGTACCAAACCTTCGGAC	488	UGCUGUACCAAACCUUCGGAC	637
M26	GCTGTACCAAACCTTCGGAC	489	GCUGUACCAAACCUUCGGAC	638
M27	ACCCCATCTCTTTGTTTTGT	490	ACCCCAUCUCUUUGUUUUGU	639
M27	AACCCCATCTCTTTGTTTTGT	491	AACCCCAUCUCUUUGUUUUGU	640
M28	GCCACAAGAACACATCATACA	492	GCCACAAGAACACAUCAUACA	641
M28	CCACAAGAACACATCATACA	493	CCACAAGAACACAUCAUACA	642
M29	TACTTTCCAATCAATAGGCCT	494	UACUUUCCAUAUAAUAGGCCU	643
M29	ACTTTCCAATCAATAGGCCT	495	ACUUUCCAUAUAAUAGGCCU	644
M3	CACATCCAGCGATAACCAGG	496	CACAUCCAGCGAUAAACCAGG	645
M3	ACACATCCAGCGATAACCAGG	497	ACACAUCCAGCGAUAAACCAGG	646
M30	TGTCAACGAATTGTGGGTCTT	498	UGUCAACGAAUUGUGGGUCUU	647
M30	GTCAACGAATTGTGGGTCTT	499	GUCAACGAAUUGUGGGUCUU	648
M31	TACACAATGTGGTTATCCTGC	500	UACACAAUGUGGUUAUCCUGC	649
M31	ACACAATGTGGTTATCCTGC	501	ACACAAUGUGGUUAUCCUGC	650
M32	CGGCAACGGCCAGGTCTGTG	502	CGGCAACGGCCAGGUCUGUG	651
M32	CCGGCAACGGCCAGGTCTGTG	503	CCGGCAACGGCCAGGUCUGUG	652
M33	GGTGGTCTCCATGCGACGTGC	504	GGUGGUCUCCAUGCGACGUGC	653
M34	TACCGCAGAGTCTAGACTCG	505	UACCGCAGAGUCUAGACUCG	654
M35	CGCAGAGTCTAGACTCGTGG	506	CGCAGAGUCUAGACUCGUGG	655
M36	CACCACGAGTCTAGACTCTG	507	CACCACGAGUCUAGACUCUG	656

Guide ID	Target Sequence	Target SEQ ID NO	gRNA Spacer Sequence	Spacer SEQ ID NO
M37 (gRNA37)	GAAAGCCCAGGATGATGGGA	508	GAAAGCCCAGGAUGAUGGGA	657
M38	CCACCCAAGGCACAGCTTGG	509	CCACCCAAGGCACAGCUUGG	658
M39	AAGCCACCCAAGGCACAGCT	510	AAGCCACCCAAGGCACAGCU	659
M4	GAAAGCCCAGGATGATGGGAT	511	GAAAGCCCAGGAUGAUGGGAU	660
M4	AAAGCCCAGGATGATGGGAT	512	AAAGCCCAGGAUGAUGGGAU	661
M40 (gRNA40)	CCATGCCCCAAAGCCACCCA	513	CCAUGCCCCAAAGCCACCCA	662
M41	TCAGGCAAGCAATTCTTTGC	514	UCAGGCAAGCAAUUCUUUGC	663
M42	CAGGCAAGCAATTCTTTGCT	515	CAGGCAAGCAAUUCUUUGCU	664
M43	AGGCAAGCAATTCTTTGCTG	516	AGGCAAGCAAUUCUUUGCUG	665
M44	GGCAAGCAATTCTTTGCTGG	517	GGCAAGCAAUUCUUUGCUGG	666
M45	GCAAGCAATTCTTTGCTGGG	518	GCAAGCAAUUCUUUGCUGGG	667
M46	TCCAAGGAATACTAACATTG	519	UCCAAGGAAUACUAACAUUG	668
M47	TTCCAATGAGGATTAAAGAC	520	UUCCAUGAGGAUUAAGAC	669
M48	TGCAATTGATTATGCCTGCT	521	UGCAAUUGAUUAUGCCUGCU	670
M49	TGGGAACAAGATCTACAGCA	522	UGGGAACAAGAUCUACAGCA	671
M5	CGAACCCTGAACAAATGGC	523	CGAACCACUGAACAAAUGGC	672
M5	ACGAACCCTGAACAAATGGC	524	ACGAACCACUGAACAAAUGGC	673
M50	GGGAACAAGATCTACAGCAT	525	GGGAACAAGAUCUACAGCAU	674
M51	GGAACAAGATCTACAGCATG	526	GGAACAAGAUCUACAGCAUG	675
M52 (gRNA52)	TCAATCCCAACAAGGACACC	527	UCAAUCCCAACAAGGACACC	676
M53	GACGCCAACAAGGTAGGAGC	528	GACGCCAACAAGGUAGGAGC	677
M54	TGCTCCAGCTCCTACCTTGT	529	UGCUCAGCUCCUACCUUGU	678
M55	CCACCAATCGCCAGACAGGA	530	CCACCAAUCGCCAGACAGGA	679
M56	AGCCACCAGCAGGGAAATAC	531	AGCCACCAGCAGGGAAAUAC	680
M57	ACCAGGACAAGTTGGAGGAC	532	ACCAGGACAAGUUGGAGGAC	681
M58	CGATAACCAGGACAAGTTGG	533	CGAUAAACCAGGACAAGUUGG	682
M59	CCCATCTCTTTGTTTTGTTA	534	CCCAUCUCUUUGUUUUGUUA	683
M6	CATCCATATAACTGAAAGCCA	535	CAUCCAUAUAACUGAAAGCCA	684
M6	ATCCATATAACTGAAAGCCA	536	AUCCAUAUAACUGAAAGCCA	685
M60	CCCCATCTCTTTGTTTTGTT	537	CCCCAUCUCUUUGUUUUGUU	686
M61	TCAACGAATTGTGGGTCTTT	538	UCAACGAUUUGUGGGUCUUU	687
M62	CAACGAATTGTGGGTCTTT	539	CAACGAUUUGUGGGUCUUU	688
M63	GGTCTCCATGCGACGTGCAG	540	GGUCUCCAUGCGACGUGCAG	689
M64	ACAGGCGGGGTTTTCTTGT	541	ACAGGCGGGGUUUUUUCUUGU	690
M65	GCAGAGTCTAGACTCGTGGT	542	GCAGAGUCUAGACUCGUGGU	691
M66/E13	GAGAAGTCCACCACGAGTCT	543	GAGAAGUCCACCACGAGUCU	692
M67	GACACATCCAGCGATAACCA	544	GACACAUCAGCGAUAAACCA	693
M68	AAGCCCAGGATGATGGGATG	545	AAGCCCAGGAUGAUGGGAUG	694
M69	CCACTCCCATAGGAATTTTC	546	CCACUCCCAUAGGAUUUUC	695
M7	TATACCCAAAGACAAAAGAAA	547	UAUACCCAAAGACAAAAGAAA	696

Guide ID	Target Sequence	Target SEQ ID NO	gRNA Spacer Sequence	Spacer SEQ ID NO
M7	ATACCCAAAGACAAAAGAAA	548	AUACCCAAAGACAAAAGAAA	697
M70	CTGAGCCAGGAGAAACGGGC	549	CUGAGCCAGGAGAAACGGGC	698
M71	TACCACATCATCCATATAAC	550	UACCACAUCAUCCAUAUAAC	699
M72	AGCCACCCAAGGCACAGCTT	551	AGCCACCCAAGGCACAGCUU	700
M73	CCAGCAAAGAATTGCTTGCC	552	CCAGCAAAGAAUUGCUUGCC	701
M74	GACGACGAGGCAGGTCCCCCT	553	GACGACGAGGCAGGUCCCCU	702
M75	GACGAGGCAGGTCCCCTAGA	554	GACGAGGCAGGUCCCCUAGA	703
M76	GGTCTCAATCGCCGCGTCGC	555	GGUCUCAUCGCCGCGUCGC	704
M77	CTCAATCGCCGCGTCGCAGA	556	CUCAAUCGCCGCGUCGCAGA	705
M78	AGTCCAAGGAATACTAACAT	557	AGUCCAAGGAAUACUAACAU	706
M79	TGTTTTCCAATGAGGATTAA	558	UGUUUUCCAUGAGGAUUA	707
M8	CATGCAACTTTTTACCTCTG	559	CAUGCAACUUUUUACCUCUG	708
M8	ATGCAACTTTTTACCTCTG	560	AUGCAACUUUUUACCUCUG	709
M80	TTTCCACCAGCAATCCTCTG	561	UUUCCACCAGCAAUCCUCUG	710
M81	CCCAACAAGGACACCTGGCC	562	CCCAACAAGGACACCUGGCC	711
M82	ACGCCAACAAGGTAGGAGCT	563	ACGCCAACAAGGUAGGAGCU	712
M83	CCTCCACCAATCGCCAGACA	564	CCUCCACCAAUCCGAGACA	713
M84	CCAATCCTCGAGAAGATTGA	565	CCAUAUCCUCGAGAAGAUUGA	714
M85	TCCCCAATCCTCGAGAAGAT	566	UCCCCAAUCCUCGAGAAGAU	715
M86	AGGGTCCCCAATCCTCGAGA	567	AGGGUCCCCAAUCCUCGAGA	716
M87	CTTCTCTCAATTTTCTAGGG	568	CUUCUCUCAUUUUUCUAGGG	717
M89	GATAACCAGGACAAGTTGGA	569	GAUAACCAGGACAAGUUGGA	718
M9	TCAAGCCTCCAAGCTGTGCC	570	UCAAGCCUCCAAGCUGUGCC	719
M90	CCACCAGCACGGGACCATGC	571	CCACCAGCACGGGACCAUGC	720
M92	CTCCATGCGACGTGCAGAGG	572	CUCCAUGCGACGUGCAGAGG	721
M93	GTGGTCTCCATGCGACGTGC	573	GUGGUCUCCAUGCGACGUGC	722
M94	TATGGATGATGTGGTATTGG	574	UAUGGAUGAUGUGGUUAUUGG	723
MSP37 (B*)	GAAAGCCCAAGATGATGGGA	10833	GAAAGCCCAAGAUGAUGGGA	10834

The gRNAs can target the HBV core protein, the HBV polymerase gene, the HBV surface protein, or the HBV X protein. In embodiments, the PAM sequence associated with a sequence targeted by a gRNA is not limiting. Other targets can include the polyadenylation signal (PAS) involved in cccDNA transcriptional activity and unique for all the viral RNAs. PAS serves as a regulatory element (TATA box) as well. Targeting the PAS in cccDNA could destabilize all viral RNA species, including pgRNA and lead to ccc DNA silencing. The sgRNAs can target main regulatory regions in cccDNA (e.g., Enhancer I, Enhancer II, Basal Core promoter, Cryptic and canonical polyadenylation signals (PAS), and S and X gene promoters). The gRNAs can be designed and evaluated *in silico* prior to being tested in cells (e.g., an HEK293-lentiHBV system).

In embodiments any one of guides E1 to E24 are used in combination with a nucleobase editor polypeptide comprising SaCas9 and/or target a sequence associated with an NNRRT, NGG, or NGA PAM sequence. In embodiments, a guide RNA comprising any of the spacer sequences provided herein (e.g., those listed in any of **Tables 2A-2C** or in the Sequence Listing), or variants thereof, are used in combination with any of the base editors and/or nucleic acid programmable DNA binding protein provided herein (e.g., BE4 or Cas12b).

In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity of the fusion proteins. For example, any of the fusion proteins provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the fusion proteins provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). 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-methylated) strand opposite the targeted nucleobase. Mutation of the catalytic residue (e.g., D10 to A10) prevents cleavage of the edited strand containing the targeted A residue. 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.

In some embodiments, a precise modification of an HBV gene decreases the virulence of the virus and/or decreases the ability of the virus to propagate *in vitro*. The modification can be a premature stop codon or other mutation that impairs or otherwise reduces the activity of an HBV protein. In some embodiments, the HBV gene that is targeted for modification is the pol, X, S, pre-S1, pre-S2, the core or pre-core genes, a transcription site, or a combination thereof.

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 or 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.

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

5 embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain can comprise 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.

10 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

15 guide nucleic acid is capable of binding to a nucleic acid sequence during CRISPR (i.e., Clustered Regularly Interspaced Short Palindromic Repeats)-mediated modification of a nucleic acid. Such a protein is referred to herein as a “CRISPR protein.” Accordingly, disclosed herein is a base editor comprising a polynucleotide programmable nucleotide binding domain comprising all or a portion of a CRISPR protein (i.e. a base editor comprising as a domain all or

20 a portion of a CRISPR protein, also referred to as a “CRISPR protein-derived domain” of the base editor). A CRISPR protein-derived domain incorporated into a base editor can be modified compared to a wild-type or natural version of the CRISPR protein. For example, as described below a CRISPR protein-derived domain can comprise one or more mutations, insertions, deletions, rearrangements and/or recombinations relative to a wild-type or natural version of the

25 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, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5,

30 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, CARE, 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.

5 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%,
10 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
15 portion of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syphilidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

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

In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a 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 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 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.

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 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 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.

Nickases

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

position 840. In such 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 can comprise an H840A mutation, while the amino acid residue at position 10 remains a D. In some embodiments, a nickase can be derived from a fully catalytically active (e.g.,
 5 natural) form of a polynucleotide programmable nucleotide binding domain by removing all or a portion of a nuclease domain that is not required for the nickase activity. For example, where a polynucleotide programmable nucleotide binding domain comprises a nickase domain derived from Cas9, the Cas9-derived nickase domain can comprise a deletion of all or a portion of the RuvC domain or the HNH domain.

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

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLVDSITDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
 15 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
 DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
 GTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
 PYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKS
 20 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMILHDDSLTF
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILOTVKVVDLVKVMGRHKPENIVIAMARENQ
 TTQKGQKNSRERMKRIEIEGIELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINR
 25 LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIAK
 SEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKG
 30 KSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFSKR
 ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD (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

5 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).

10 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

15 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

20 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.

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

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDS TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTY
30 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NF
DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFR I
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKS

LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
 5 TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
 LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIAK
 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 10 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKG
 KSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEQKQLFVEQHKHYLDEIIIEQISEFSKR
 ILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQSI TGLYETRIDLSQLGGD (SEQ ID NO: 201).

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

The “efficiency” of non-homologous end joining (NHEJ) and/or homology directed repair (HDR) can be calculated by any convenient method. For example, in some embodiments, efficiency can be expressed in terms of percentage of successful HDR. For example, a surveyor
 25 nuclease assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage. For example, a surveyor nuclease enzyme can be used that directly cleaves DNA containing a newly integrated restriction sequence as the result of successful HDR. More cleaved substrate indicates a greater percent HDR (a greater efficiency of HDR). As an illustrative example, a fraction (percentage) of HDR can be calculated using the
 30 following equation [(cleavage products)/(substrate plus cleavage products)] (e.g., (b+c)/(a+b+c), where “a” is the band intensity of DNA substrate and “b” and “c” are the cleavage products).

In some embodiments, efficiency can be expressed in terms of percentage of successful NHEJ. For example, a T7 endonuclease I assay can be used to generate cleavage products and the ratio of products to substrate can be used to calculate the percentage NHEJ. T7 endonuclease

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

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

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

In order to utilize HDR for gene editing, a DNA repair template containing the desired sequence can be delivered into the cell type of interest with the gRNA(s) and Cas9 or Cas9 nickase. The repair template can contain the desired edit as well as additional homologous sequence immediately upstream and downstream of the target (termed left & right homology arms). The length of each homology arm can be dependent on the size of the change being introduced, with larger insertions requiring longer homology arms. The repair template can be a single-stranded oligonucleotide, double-stranded oligonucleotide, or a double-stranded DNA plasmid. The efficiency of HDR is generally low (<10% of modified alleles) even in cells that express Cas9, gRNA and an exogenous repair template. The efficiency of HDR can be enhanced by synchronizing the cells, since HDR takes place during the S and G2 phases of the cell cycle. Chemically or genetically inhibiting genes involved in NHEJ can also increase HDR frequency.

In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence can have additional sites throughout the genome where partial homology exists. These sites are called off-targets and need to be considered when designing a gRNA. In addition to optimizing gRNA design, CRISPR specificity can also be increased through modifications to Cas9. Cas9 generates double-strand breaks (DSBs) through the combined activity of two nuclease domains,

RuvC and HNH. Cas9 nickase, a D10A mutant of SpCas9, retains one nuclease domain and generates a DNA nick rather than a DSB. The nickase system can also be combined with HDR-mediated gene editing for specific gene edits.

5 *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
 10 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
 15 inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain can comprise one or more deletions of all or a portion of a catalytic domain (*e.g.*, RuvC1 and/or HNH domains). In further embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises a point mutation (*e.g.*, D10A or H840A) as well as a
 20 deletion of all or a 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
 25 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;
 30 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
5 comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

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

In some 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
20 (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB
25 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
30 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 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 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 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 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, 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, 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 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 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.

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.

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 embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation, or one or more corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some embodiments, one of the Cas9 domains present in the fusion protein may be replaced with a guide nucleotide sequence-programmable DNA-binding protein domain that has no requirements for a PAM sequence. In some embodiments, the Cas9 is an SaCas9. Residue A579 of SaCas9 can be mutated from N579 to yield a SaCas9 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, 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, 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, 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

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

PAM	SpCas9 amino acid position												
	111 4	113 5	121 8	121 9	122 1	124 9	132 0	132 1	132 3	133 2	133 3	133 5	133 7
	R	D	G	E	Q	P	A	P	A	D	R	R	T
CAC		N		V								Q	N
CAC				V						N		Q	N

Table 3B

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

Table 3C

PAM	SpCas9 amino acid position															
	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320
R	Y	D	E	K	D	K	K	G	E	Q	A	P	E	N	A	A
SacB.TAT		N				N			V	H						V
SacB.TAT		N						S	V	H		S				

PAM	SpCas9 amino acid position																
	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320	1332
AAT	R	Y	D	E	K	D	K	G	E	Q	A	P	E	N	A	A	D
TAT			N					S	V	H	V	S		K	T		G
TAT	G		N			G		S	V	H		S	K			S	G
TAT	G		N			G		S	V	H		S				S	G
TAT	G	C	N			G		S	V	H		S				S	G
TAT	G	C	N			G		S	V	H		S				S	G
TAT	G	C	N			G		S	V	H		S				S	G
TAT	G	C	N		E	G		S	V	H		S				S	G
TAT	G	C	N	V		G		S	V	H		S				S	G
TAT		C	N			G		S	V	H		S				S	G
TAT	G	C	N			G		S	V	H		S				S	G

Table 3D

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

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 *Alicyclobacillus 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 provided herein may be a Cas12b/C2c1, or a Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a Cas12b/C2c1 protein. In

some embodiments, the napDNAbp is a Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12b/C2c1 or

5 Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any one of the napDNAbp sequences provided herein. It should be
10 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
15 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
20 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
25 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 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
30 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 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 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 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 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 provided herein may be a Cas12j/CasΦ protein. Cas12j/CasΦ is described in Pausch *et al.*, “CRISPR-CasΦ from huge phages is a hypercompact genome editor,” *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337, which is incorporated herein by reference in its entirety. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, 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 with Internal Insertion

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

napDNAbp at a C-terminal end of the napDNAbp, an N-terminal end of the napDNAbp, or inserted at an internal location of the napDNAbp. In some embodiments, the heterologous polypeptide is 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
 5 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.

10 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;

15 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;

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

wherein each instance of “-[-“ is an optional linker.

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
 25 numbered in the TadA reference sequence.

The fusion protein can comprise more than one deaminase. The fusion protein can comprise, for example, 1, 2, 3, 4, 5 or more deaminases. In some embodiments, the fusion protein comprises one or two deaminase. The two or more deaminases in a fusion protein can be an adenosine deaminase, a cytidine deaminase, or a combination thereof. The two or
 30 more deaminases can be homodimers or 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 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 fragment thereof. In some embodiments, the Cas9 polypeptide is a nuclease dead Cas9 (dCas9) polypeptide or a fragment thereof. The Cas9 polypeptide in a fusion protein can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein may not be a full length Cas9 polypeptide. The Cas9 polypeptide can be truncated, for example, at a N-terminal or C-terminal end relative to a naturally-occurring Cas9 protein. The Cas9 polypeptide can be a circularly permuted Cas9 protein. The Cas9 polypeptide can be a fragment, a portion, or a domain of a Cas9 polypeptide, that is still capable of binding the target polynucleotide and a guide nucleic acid sequence.

In some embodiments, the Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus* 1 Cas9 (St1Cas9), or fragments or variants 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 deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase fused to the N-terminus.

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

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

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

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

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

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

In various embodiments, the catalytic domain has DNA modifying activity (e.g., deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (e.g., 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

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:

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

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

The heterologous polypeptide (e.g., deaminase) can be inserted in the napDNAbp (e.g., Cas9 or Cas12 (e.g., Cas12b/C2c1)) at a suitable location, for example, such that the napDNAbp retains its ability to bind the target polynucleotide and a guide nucleic acid. A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and

15 cytidine deaminase) can be inserted into a napDNAbp without compromising function of the deaminase (e.g., base editing activity) or the napDNAbp (e.g., ability to bind to target nucleic acid and guide nucleic acid). A deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp at, for example, a disordered region or a region comprising a high temperature factor or B-factor as

20 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

25 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

30 determined by B-factor analysis of the crystal structure of Cas9 polypeptide. In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in regions of the Cas9 polypeptide comprising higher than average B-factors (e.g., higher B factors compared to the total protein or the protein domain comprising the disordered region). B-factor or temperature factor can

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

A heterologous polypeptide (e.g., deaminase) can be inserted in the napDNABp at an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, or 1249-1250 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. It

should be understood that the reference to the above Cas9 reference sequence with respect to insertion positions is for illustrative purposes. The insertions as discussed herein are not limited to the Cas9 polypeptide sequence of the above Cas9 reference sequence, but include insertion at corresponding locations in variant Cas9 polypeptides, for example a Cas9 nickase (nCas9), nuclease dead Cas9 (dCas9), a Cas9 variant lacking a nuclease domain, a truncated Cas9, or a Cas9 domain lacking partial or complete HNH domain.

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

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

In some embodiments, a 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.

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
5 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
10 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
15 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
20 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
25 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
30 embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 1052, or is inserted to replace amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 1067, or is

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

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

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

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

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

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

Exemplary internal fusions base editors are provided in Table 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

BE ID	Modification	Other ID
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

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

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

- 10 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
- 15 nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some embodiments, the HNH domain is deleted and the deaminase is inserted in its place. In some

embodiments, one or more of the RuvC domains is deleted and the deaminase is inserted in its place.

A fusion protein comprising a heterologous polypeptide can be flanked by a N-terminal and a C-terminal fragment of a napDNAbp. In some embodiments, the fusion protein comprises a deaminase flanked by a N-terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The N terminal fragment or the C terminal fragment can bind the target polynucleotide sequence. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of a flexible loop of a Cas9 polypeptide. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of an alpha-helix structure of the Cas9 polypeptide. The N-terminal fragment or the C-terminal fragment can comprise a DNA binding domain. The N-terminal fragment or the C-terminal fragment can comprise a RuvC domain. The N-terminal fragment or the C-terminal fragment can comprise an HNH domain. In some embodiments, neither of the N-terminal fragment and the C-terminal fragment comprises an HNH domain.

In some embodiments, the C-terminus of the N terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. In some embodiments, the N-terminus of the C terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. The insertion location of different deaminases can be different in order to have proximity between the target nucleobase and an amino acid in the C-terminus of the N terminal Cas9 fragment or the N-terminus of the C terminal Cas9 fragment. For example, the insertion position of an 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.

The N-terminal Cas9 fragment of a fusion protein (i.e. the N-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the N-terminus of a Cas9 polypeptide. The N-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The N-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at

least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The C-terminal Cas9 fragment of a fusion protein (i.e. the C-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the C-terminus of a Cas9 polypeptide. The C-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The C-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

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

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

In some embodiments, the deaminase (e.g., adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) of the fusion protein deaminates no more than two nucleobases within the range of an R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than three nucleobases within the range of the R-loop.

5 In some embodiments, the deaminase of the fusion protein deaminates no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleobases within the range of the R-loop. An R-loop is a three-stranded nucleic acid structure including a DNA:RNA hybrid, a DNA:DNA or an RNA: RNA complementary structure and the associated with single-stranded DNA. As used herein, an R-loop may be formed when a target polynucleotide is contacted with a CRISPR complex or
10 a base editing complex, wherein a portion of a guide polynucleotide, e.g. a guide RNA, hybridizes with and displaces with a portion of a target polynucleotide, e.g. a target DNA. In some embodiments, an R-loop comprises a hybridized region of a spacer sequence and a target DNA complementary sequence. An R-loop region may be of about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,
15 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 to a DNA strand that comprises the complementary strand to a
20 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 described herein can effect target deamination in an editing
25 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
30 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 can comprise more than one heterologous polypeptide. For example, the fusion protein can additionally comprise one or more UGI domains and/or one or more nuclear localization signals. The two or more heterologous domains can be inserted in tandem. The two or more heterologous domains can be inserted at locations such that they 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 is a Cas12 polypeptide, e.g., Cas12b/C2c1, or a fragment thereof. The Cas12 polypeptide can be a variant Cas12 polypeptide. In other embodiments, the N- or C-terminal fragments of the Cas12 polypeptide comprise a nucleic acid programmable DNA binding domain or a RuvC domain. In other embodiments, the fusion protein contains a linker between the Cas12 polypeptide and the catalytic domain. In other embodiments, the amino acid sequence of the linker is GSGSGS (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:

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

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

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

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

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

In various embodiments, the catalytic domain has DNA modifying activity (e.g.,
 5 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 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
 10 fused within Cas12 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas12 are provided as follows:

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

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

20 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 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
 25 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
 30 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
 5 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,
 10 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
 15 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
 20 147 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
 25 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
 30 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 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 further contains a tag (e.g., an influenza hemagglutinin tag).

In some embodiments, the fusion protein comprises a napDNAbp domain (e.g., Cas12-derived domain) with an internally fused nucleobase editing domain (e.g., all or a portion of a deaminase domain, e.g., an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b. In some embodiments, the base editor comprises a BhCas12b domain with an internally fused Tada*8 domain inserted at the loci provided in Table 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 deaminase 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 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
 5 to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine residue (e.g., inosine), which can improve the activity or efficiency of the base editor.

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
 10 comprising an adenosine deaminase can deaminate a target A of a polynucleotide comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR, e.g., ADAR1 or
 15 ADAR2) or tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from
 20 *Escherichia coli* (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, 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*).
 25 In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Shewanella putrefaciens*, *Haemophilus influenzae*, *Caulobacter crescentus*, or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine
 30 deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (e.g., mutations in ecTadA). The corresponding residue in any homologous protein can be identified by e.g., sequence alignment and determination of homologous residues. The mutations in any naturally-occurring adenosine deaminase (e.g., having homology to ecTadA) that correspond to any of

the mutations described herein (e.g., any of the mutations identified in ecTadA) can be 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 the TadA reference sequence) can be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (e.g., bacterial adenosine deaminases). It would be apparent to the skilled artisan that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein. Thus, any of the mutations identified in the TadA reference sequence can be made in other adenosine deaminases (e.g., ecTadA) that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein can be made individually or in any combination in the TadA reference sequence or another adenosine deaminase.

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

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

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

In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E155D, E155G, or E155V mutation. In some embodiments, the adenosine deaminase comprises a D147Y.

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, an adenosine deaminase may contain a D108N, a A106V, a E155V, and/or a D147Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") in TadA reference sequence, or corresponding mutations in another adenosine deaminase: 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
 5 be made in an adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X mutation in TadA reference sequence, or one or more
 10 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,
 15 or D108A, or D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a H8X, D108X, and/or N127X mutation in TadA reference sequence, or one or more corresponding
 20 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 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 TadA reference sequence, or one or more
 25 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
 30 Q154R, E155G or E155V or E155D, K161Q, Q163H, and/or T166P mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X,

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

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, 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, 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, N127X, D147X, and E155X in 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 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 H8Y, D108N, N127S, D147Y, R152C,

and Q154H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R, E155G and Q163H in TadA

5 reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the

10 adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R26W, L68Q,

15 D108N, N127S, D147Y, and E155V in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, A109T, N127S, and E155G in TadA reference sequence, or a corresponding mutation or mutations in another adenosine

20 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 TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some

25 embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and

30 Q154H mutation in 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 TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or

corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in 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 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 TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of S2A, H8Y, I49F, L84F, H123Y, N127S, I156F, and/or K160S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an L84X mutation 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 L84F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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

In some embodiments, the adenosine deaminase comprises an I156X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an I156F mutation in 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 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

5 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 TadA reference sequence, or a
10 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,
15 D147Y, E155V, and I156F in 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 TadA reference sequence.

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

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in TadA reference sequence, or one or more
25 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,
30 A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein corresponding to TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an E25X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M,
5 E25D, E25A, E25R, E25V, E25S, or E25Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R26X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type
10 adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R107X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase,
15 where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R107P, R107K, R107A, R107N, R107W, R107H, or R107S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

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

In some embodiments, the adenosine deaminase comprises an A143X mutation in
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 A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA
30 reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, and/or K161X mutation in TadA reference sequence, or one or more corresponding mutations in

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

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

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

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

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

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

or S146C mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

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

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

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

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

In some embodiments, the adenosine deaminase comprises an R152X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a R152P or R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

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:

- (A106V_D108N),
- 5 (R107C_D108N),
(H8Y_D108N_N127S_D147Y_Q154H),
(H8Y_D108N_N127S_D147Y_E155V),
(D108N_D147Y_E155V),
(H8Y_D108N_N127S),
- 10 (H8Y_D108N_N127S_D147Y_Q154H),
(A106V_D108N_D147Y_E155V),
(D108Q_D147Y_E155V),
(D108M_D147Y_E155V),
(D108L_D147Y_E155V),
- 15 (D108K_D147Y_E155V),
(D108I_D147Y_E155V),
(D108F_D147Y_E155V),
(A106V_D108N_D147Y),
(A106V_D108M_D147Y_E155V),
- 20 (E59A_A106V_D108N_D147Y_E155V),
(E59A cat dead_A106V_D108N_D147Y_E155V),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D103A_D104N),
- 25 (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),
- 30 (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
- 35 (E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

- (A106V_D108N_A142N_D147Y_E155V),
 (R26G_A106V_D108N_A142N_D147Y_E155V),
 (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
 (R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
 5 (E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
 (A106V_R107K_D108N_A142N_D147Y_E155V),
 (A106V_D108N_A142N_A143G_D147Y_E155V),
 (A106V_D108N_A142N_A143L_D147Y_E155V),
 (H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
 10 (N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
 (N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
 (H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
 (N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
 (H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
 15 (H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N)
 (H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
 (L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
 (N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
 20 (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),
 25 (N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
 (P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
 (D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
 (L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
 30 (H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),
 (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
 (L84F_A106V_D108N_D147Y_E155V_I156F),
 (R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
 35 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
 (L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),
 (R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

- (R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
5 (L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
10 (H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N),
15 (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),
20 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
25 _K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F
_K157N),
30 (W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P
35 _E155V_I156F_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
_K161T),

(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V_I156F_K157N).

In some embodiments, the TadA deaminase is TadA variant. In some embodiments, the TadA variant is TadA*7.10. In particular embodiments, the fusion proteins 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 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

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAE IMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1).

In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 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 + 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, an adenosine deaminase variant (*e.g.*, TadA*8) comprises a deletion. In some embodiments, an adenosine deaminase variant comprises a deletion of the 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 TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

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

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

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

In other embodiments, the adenosine deaminase variant is a heterodimer 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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, 16, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In some embodiments the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12,

TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

In other embodiments, a base editor of the disclosure 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, 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 TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in Table 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	TadA amino acid number									
		26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
PACE	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

In one embodiment, a fusion protein of the 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 comprise a single TadA*8 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein 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 GAVLVNLRV IGEGWNRAIG 50
 LHDPTAHAEI MALRQGGLVM QNYRLIDATL **Y**VTFEPCVMC AGAMIHSRIG 100
 RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR 150
 MPRQVFNAQK KAQSSTD (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 TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In particular embodiments, a combination of alterations is selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R + Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

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

C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

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

In particular embodiments, the fusion proteins 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
 10 embodiments, the fusion proteins 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 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
 15 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
 20 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
 25 MPRQVFENAQK 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; 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.

5 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
10 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/2020/049975, which is incorporated herein by reference for its entirety.

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

Details of A to G nucleobase editing proteins are described in International PCT
20 Application No. PCT/2017/045381 (WO2018/027078) and Gaudelli, N.M., *et al.*, “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” *Nature*, 551, 464-471 (2017), the entire contents of which are hereby incorporated by reference.

25 ***C to T Editing***

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

The deamination of a target C in a polynucleotide to give rise to a U is a non-limiting example of a type of base editing that can be executed by a base editor described herein. In another example, a base editor comprising a cytidine deaminase domain can mediate conversion of a cytosine (C) base to a guanine (G) base. For example, a U of a

5 polynucleotide produced by deamination of a cytidine by a cytidine deaminase domain of a base editor can be excised from the polynucleotide by a base excision repair mechanism (*e.g.*, by a uracil DNA glycosylase (UDG) domain), producing an abasic site. The nucleobase opposite the abasic site can then be substituted (*e.g.*, by base repair machinery) with another base, such as a C, by for example a translesion polymerase. Although it is typical for a
10 nucleobase opposite an abasic site to be replaced with a C, other substitutions (*e.g.*, A, G or T) can also occur.

Accordingly, in some embodiments a base editor described herein comprises a deamination domain (*e.g.*, cytidine deaminase domain) capable of deaminating a target C to a U in a polynucleotide. Further, as described below, the base editor can comprise additional
15 domains which facilitate conversion of the U resulting from deamination to, in some embodiments, a T or a G. For example, a base editor comprising a cytidine deaminase domain can further comprise a uracil glycosylase inhibitor (UGI) domain to mediate substitution of a U by a T, completing a C-to-T base editing event. In another example, a base editor can incorporate a translesion polymerase to improve the efficiency of C-to-G base
20 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
25 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-
30 stranded polynucleotide, but the target C can be positioned in a portion of the polynucleotide which at the time of the deamination reaction is in a single-stranded state. For example, in embodiments where the NAGPB domain comprises a Cas9 domain, several nucleotides can be left unpaired during formation of the Cas9-gRNA-target DNA complex, resulting in formation of a Cas9 “R-loop complex”. These unpaired nucleotides can form a bubble of

single-stranded DNA that can serve as a substrate for a single-strand specific nucleotide deaminase enzyme (*e.g.*, cytidine deaminase).

In some embodiments, a cytidine deaminase of a base editor can comprise all or a portion of an apolipoprotein B mRNA editing complex (APOBEC) family deaminase.

- 5 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,
- 10 APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D (“APOBEC3E” now refers to this), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, and Activation-induced (cytidine) deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC1 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC2 deaminase. In some
- 15 embodiments, a deaminase incorporated into a base editor comprises all or a portion of is an APOBEC3 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC3A deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3B deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of
- 20 APOBEC3C deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3D deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3E deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3F deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of
- 25 APOBEC3G deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC3H deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of APOBEC4 deaminase. In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of activation-induced deaminase (AID). In some embodiments a
- 30 deaminase incorporated into a base editor comprises all or a portion of cytidine deaminase 1 (CDA1). It should be appreciated that a base editor can comprise a deaminase from any suitable organism (*e.g.*, a human or a rat). In some embodiments, a deaminase domain of a base editor is from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase domain of the base editor is derived from rat (*e.g.*, rat

APOBEC1). In some embodiments, the deaminase domain of the base editor is human APOBEC1. In some embodiments, the deaminase domain of the base editor is pmCDA1.

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

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

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

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

In some embodiments, an APOBEC deaminase incorporated into a base editor can comprise a H121R and a H122R mutation of rAPOBEC1, or one or more corresponding

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

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

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

A number of modified cytidine deaminases are commercially available, including, but not limited to, SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3, which are available from Addgene (plasmids 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177). In some embodiments, a deaminase incorporated into a base editor comprises all or a portion of an APOBEC1 deaminase.

In some embodiments, the fusion proteins 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 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, 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 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.

A fusion protein of the invention second protein comprises two or more nucleic acid 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 by reference.

5 **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 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
 10 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
 15 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
 20 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
 25 (“sgRNA”, or simply “gRNA”) can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-
 30 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, 5 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 10 “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 “gNRA”) can be engineered so as 15 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.

In some embodiments, the guide polynucleotide is at least one single guide RNA (“sgRNA” or “gNRA”). In some embodiments, a guide polynucleotide comprises two or 20 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 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 25 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.

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 30 guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20 nucleotides in length.

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.

5 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 the
10 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 can comprise both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide
15 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
20 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
25 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
30 polynucleotide comprises multiple nucleic acid molecules, the protein-binding segment of can include all or a portion of multiple separate molecules that are for instance hybridized along a region of complementarity. In some embodiments, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75

of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of “segment,” unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and can include regions of RNA molecules that are of any total length and can include regions with complementarity to other molecules.

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.

A gRNA molecule can be transcribed *in vitro*.

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 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 or a guide polynucleotide can comprise three regions: a first region at the 5' end that can be complementary to a target site in a chromosomal sequence, a second internal region that can form a stem loop structure, and a third 3' region that can be single-stranded. A first region of each gRNA can also be different such that each gRNA guides a fusion protein to a specific target site. Further, 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 can comprise from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or
5 from about 10 nucleotides to about 25 nucleotides; or from 10 nucleotides to about 25 nucleotides; or from about 10 nucleotides to 25 nucleotides) or more. For example, a region of base pairing between a first region of a 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
10 21 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
15 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.

A gRNA or a guide polynucleotide can also comprise a third region at the 3' end that
20 can be essentially single-stranded. For example, a third region is sometimes not complementary to any chromosomal sequence in a cell of interest and is sometimes not complementary 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 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
25 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 of a gene can be targeted.

A gRNA or a guide polynucleotide can target a nucleic acid sequence of about 20
30 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 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 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 acid sequence, *e.g.*, to minimize total off-target activity across the genome. For example, for each possible targeting domain choice using *S. pyogenes* Cas9, all off-target sequences (preceding selected PAMs, *e.g.*, NAG or NGG) may be identified across the genome that contain up to certain number (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. First regions of gRNAs complementary to a target site can be identified, and all first regions (*e.g.*, crRNAs) can be ranked according to its total predicted off-target score; the top-ranked targeting domains represent those that are likely to have the greatest on-target and the least off-target activity. Candidate targeting gRNAs can be functionally evaluated by using methods known in the art and/or as set forth herein.

As a non-limiting example, target DNA hybridizing sequences in crRNAs of a 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) can comprise 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. In some embodiments, such gRNAs can be designed such that the mutated start codon will not be base-paired with the gRNA. The guide polynucleotides can comprise standard ribonucleotides, modified ribonucleotides (*e.g.*, pseudouridine), ribonucleotide isomers, and/or ribonucleotide analogs. In some embodiments, the guide polynucleotide can comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

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

A guide polynucleotide can comprise one or more modifications to provide a nucleic acid with a new or enhanced feature. A guide polynucleotide can comprise a nucleic acid affinity tag. A guide polynucleotide can comprise synthetic nucleotide, synthetic nucleotide analog, nucleotide derivatives, and/or modified nucleotides.

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

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

A gRNA or a guide polynucleotide can also be modified by 5' adenylate, 5' guanosine-triphosphate cap, 5' N7-Methylguanosine-triphosphate cap, 5' triphosphate cap, 3' phosphate, 3' thiophosphate, 5' phosphate, 5' thiophosphate, Cis-Syn thymidine dimer,

trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9, 3'-3' modifications, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3' DABCYL, black hole quencher 1, 5 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, 10 phosphothioate DNA, phosphorothioate RNA, UNA, pseudouridine-5'-triphosphate, 5'-methylcytidine-5'-triphosphate, or any combination thereof.

In some cases, a modification is permanent. In other cases, a modification is transient. In some cases, multiple modifications are made to a gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide modification can alter physiochemical 15 properties of a nucleotide, such as their conformation, polarity, hydrophobicity, chemical 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, 20 such as using virus-mediated gene delivery. A gRNA or a guide polynucleotide can be 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.

25 A modification can also be a phosphorothioate substitute. In some cases, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases and; a modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a gRNA or a guide polynucleotide. A modification can also enhance biological activity. In 30 some cases, a phosphorothioate enhanced RNA gRNA can inhibit RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA gRNAs to be used in applications where exposure to nucleases is of high probability *in vivo* or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'- or 3'-end of a gRNA which can inhibit exonuclease degradation.

In some cases, phosphorothioate bonds can be added throughout an entire gRNA to reduce attack by endonucleases.

In some embodiments, the guide RNA is designed to disrupt a splice site (*i.e.*, a splice acceptor (SA) or a splice donor (SD). In some embodiments, the guide RNA is designed
5 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
10 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
15 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
20 capable of binding a nucleotide sequence that contains a canonical or non-canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion of CRISPR proteins that have different PAM specificities.

For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9),
25 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'
30 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
 5 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
Cpf1	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 variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed “MQKFRAER”).

15 In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the NGT PAM variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or 1219. In some embodiments, the NGT PAM variant is created through targeted mutations at

one or more residues 1219, 1335, 1337, 1218. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335. In some embodiments, the NGT PAM variant is selected from the set of targeted mutations provided in **Tables 8A** and **8B** below.

5

Table 8A: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

Table 8B: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

Variant	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		

36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

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

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

Table 9: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

10

In some embodiments, the NGT PAM is selected from the variants provided in **Table 10** below.

Table 10. NGT PAM variants

	NGTN variant	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
Variant 1	LRKIQK	L	R	K	I	–	Q	K
Variant 2	LRSVQK	L	R	S	V	–	Q	K
Variant 3	LRSVQL	L	R	S	V	–	Q	L
Variant 4	LRKIRQK	L	R	K	I	R	Q	K
Variant 5	LRSVRQK	L	R	S	V	R	Q	K
Variant 6	LRSVRQL	L	R	S	V	R	Q	L

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

In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some
10 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 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
15 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
20 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
25 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
 5 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
 10 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
 15 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
 20 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
 25 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 Jakimo *et al.*, (Chatterjee, *et al.*, "A Cas9 with PAM recognition for adenine dinucleotides", *Nature Communications*, vol. 11, article no. 2474 (2020)), and is in the
 30 Sequence Listing as SEQ ID NO: 237.

In some embodiments, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA or RNA. Such a Cas9 protein has a reduced ability to cleave a

target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). 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, 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). Also, mutations other than alanine substitutions are suitable.

In some embodiments, a CRISPR protein-derived domain of a base editor can comprise all or a portion of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM 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 Comprising a NapDNAbp and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain or other nucleic acid programmable DNA binding protein (*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;

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

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:

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

In some embodiments, any of the Cas12 domains or Cas12 proteins provided herein
 10 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;
 NH2-[Cas12 domain]-[adenosine deaminase]-COOH;
 15 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;
 20 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.

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

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

In some embodiments, the adenosine deaminase of the fusion protein comprises a
 30 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;

NH₂-[TadA*8]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or

5 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:

NH₂-[TadA*9]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

10 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.

In some embodiments, the fusion protein can comprise a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some

15 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 comprising a cytidine deaminase or
20 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 presence of an optional linker. In some embodiments, cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For
25 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 of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein may comprise inhibitors, cytoplasmic localization sequences, export sequences, such
30 as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags,

glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, 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 comprises one or more His tags.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/2017/044935, PCT/US2019/044935, and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

Fusion Proteins Comprising a Nuclear Localization Sequence (NLS)

In some embodiments, the fusion proteins provided herein further comprise one or more (*e.g.*, 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In one embodiment, a bipartite NLS is used. In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.*, by nuclear transport). In some embodiments, 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), KRGINDRNFWRGENGRKTR (SEQ ID NO: 193),
RKSGKIAAIVVKRPRKPKKKRKV (SEQ ID NO: 329), or
MDSLMLNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 196).

In some embodiments, the fusion proteins 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 presence of an optional linker. In some embodiments, the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the cytidine deaminase and adenosine deaminase and the napDNAbp are fused via any of the linkers provided 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;

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;

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;

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 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)

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 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.

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 can comprise a nuclease, a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain.

In some embodiments, a base editor can comprise 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 comprising a UGI domain.

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

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

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

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[APOBEC1]-[nucleobase editing domain]-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 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;

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.

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 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 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.

In some embodiments, a base editing system as provided herein provides a new approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a 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 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/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage” *Nature* 533, 420-424 (2016); Gaudelli, N.M., *et al.*, “Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage” *Nature* 551, 464-471 (2017); and Komor, A.C., *et al.*, “Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity” *Science Advances* 3:eaa04774 (2017), the entire contents of which are hereby incorporated by reference.

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

A base editor system may further comprise a guide polynucleotide component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and

interactions thereof. In some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a guide polynucleotide. For example, in some embodiments, the nucleobase editing component of the base editor system, e.g., the deaminase component, can comprise an additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (e.g., a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain (e.g., polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the deaminase domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

In some embodiments, a base editor system can further comprise an inhibitor of base excision repair (BER) component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. The inhibitor of BER component may comprise a base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be a uracil DNA glycosylase inhibitor (UGI). In some embodiments, the inhibitor of base excision repair can be an inosine base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the polynucleotide programmable nucleotide binding domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain and an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable

nucleotide binding domain can target an inhibitor of base excision repair to a target nucleotide sequence by non-covalently interacting with or associating with the inhibitor of base excision repair. For example, in some embodiments, the inhibitor of base excision repair component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the guide polynucleotide. For example, in some embodiments, the inhibitor of base excision repair can comprise an additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the guide polynucleotide (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the inhibitor of base excision repair. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition 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. In some embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edit of base pair is upstream of a PAM site. In some embodiments, the intended edit of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edit of base-pair is downstream of a

PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site.

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
5 embodiments, the linker or spacer is 1-25 amino acids in length. In some embodiments, the linker or spacer is 5-20 amino acids in length. In some embodiments, the linker or spacer is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined
10 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 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
15 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
20 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
25 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
30 embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

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

a polynucleotide programmable nucleotide binding domain. In some embodiments, an NLS of the base editor is localized C-terminal to a polynucleotide programmable nucleotide binding domain.

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

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

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
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

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

5 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

10 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

15 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

20 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).

25 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). 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 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 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-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, 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

5 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,

10 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

15 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

20 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

25 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

30 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
5 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
10 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
15 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
20 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
25 (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24

[1] 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,
30 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.

[2]

Table 12: Adenosine Deaminase 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	Adenosine Deaminase	Adenosine Deaminase Description
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 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 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
 5 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
 10 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
 15 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).

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
 20 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.*,
 25 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein).

30

Table 13: Additional Adenosine Deaminase 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)

5

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).

10

15

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

Table 14. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K

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

5 one of the ABEs as shown in Table 15 below.

Table 15. Residue Identity in Evolved TadA

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	R	L	A	L	I	V	F	V	N	Y	C	Y	P	Q	V	F	N	T
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												
ABE8.6-d																		R
ABE8.7-d														R				
ABE8.8-d										H		R		R				
ABE8.9-d					Y							R		R				
ABE8.10-d												R		R				R
ABE8.11-d												T		R				
ABE8.12-d												T		S				
ABE8.13-d					Y					H		R		R				
ABE8.14-d					Y	S												

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE8.15-d						S						R						
ABE8.16-d						S				H		R						
ABE8.17-d						S								R				
ABE8.18-d						S				H				R				
ABE8.19-d						S				H		R		R				
ABE8.20-d					Y	S				H		R		R				
ABE8.21-d												R		S				
ABE8.22-d						S								S				
ABE8.23-d						S				H								
ABE8.24-d						S				H		T						

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

5 ABE8.1_Y147T_CP5_NGC PAM_monomer
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAE IMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD **SGGSSGGSSGSETPGTSES**
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
10 GRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
VAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLI IKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQH
KHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENI IHLFTLTNLGAPRAF
KYFDTTIARKEYRSTKEVLDATLIHQSI TGLYE TRIDLSQLGGDGGSSGGSSGGSSGGSSGGSS
15 GGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIY LALAHMIKFRGHFLIEGDLNPDNSDVKL
FIQLVQTYNQLFEE NPINASGVDAKAIL SARLSKSRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKD TYDDDL DNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
20 NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPF
LKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIER
MTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
25 IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL

DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVK
 VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIIEGIELGSQILKEHPVENT
 QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTTRSDKNRGK
 SDNVPSEEVVKMKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQIT
 5 KHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRE INNYHHAHDAYLN
 AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEGADKRTADGSEFESPKKKRKV
 (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
 10 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).

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
 15 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
 No. PCT/2020/049975, which is incorporated herein by reference for its entirety.

20 **Table 16. Adenosine Deaminase 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 Description	Alterations
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 Description	Alterations
ABE9.58_heterodimer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R

In some embodiments, the base editor comprises a domain comprising all or a portion of a uracil glycosylase inhibitor (UGI). In some embodiments, the base editor comprises a domain comprising all or a portion of a nucleic acid polymerase. In some embodiments, a base editor can comprise as a domain all or a portion of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion of a eukaryotic NAP. In some embodiments, a NAP or portion thereof incorporated into a base editor is a DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor has translesion polymerase activity. In some 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 can comprise 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.

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

Linkers

In certain embodiments, linkers may be used to link any of the peptides or peptide domains of the invention. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length. In certain embodiments, the linker is a polypeptide or based on amino acids. In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In certain embodiments, the linker is a carbon-nitrogen bond

of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

Typically, a linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, a linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, a linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, a linker is 2-100 amino acids in length, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. In some embodiments, the linker is about 3 to about 104 (*e.g.*, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100) amino acids in length. Longer or shorter linkers are also contemplated.

In some embodiments, any of the fusion proteins provided herein, comprise a cytidine or 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), (GGGGGS)_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., 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 (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 comprises the amino acid sequence of:

SGGSSGSETPGTSESATPESSGGS (SEQ ID NO: 356),
 SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 357), or
 GGS GGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS
 EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS (SEQ ID NO: 358).

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: SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (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:

SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS (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:
 PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG
 TSTEPSEGSAPGTSESATPESGPGSEPATSGGSSGGS (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
 5 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
 10 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
 15 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
 20 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 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 base editing at the site with reduced adjacent target editing, for example, as described
 25 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/s13059-017-1306-z, the contents of each of which are incorporated by reference herein in their
 30

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

5 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 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

10 approaches, for example, 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 complexes comprising any of the fusion

20 proteins 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. 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

25 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

30 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 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 disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for 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. 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 to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

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 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 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 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 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 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-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;

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;

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;

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;

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;

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

5 NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

10 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.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (e.g., a “deamination window”). In some embodiments, a target can be within a 4-
15 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
20 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.

A defined target region can be a deamination window. A deamination window can be
25 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
30 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 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 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 domains having one or more of the activities described herein.

5 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)
10 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
15 binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

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

20 Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA described herein.

In some embodiments, a fusion protein of the invention is used for editing a target
25 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
30 reduced or 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 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 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 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. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

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 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)

5 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.*, HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection. In some embodiments, the intended

10 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.* HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation within the coding region

15 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.* HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection. In some

20 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.

25 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 some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.*

30 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. 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 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, 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 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 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 promoter 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 methods of using the fusion proteins 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, 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, 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, any of base editor systems provided herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less 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 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 base editor systems comprising one of the ABE8 base editor variants described herein result in less than 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in at most 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one

of the ABE8 base editor variants described herein result in less than 0.3% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising one of

5 ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising an ABE7.10.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has reduction in indel frequency compared to a base editor

10 system comprising one of the ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at

15 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%,

20 at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising an ABE7.10.

The invention provides adenosine deaminase variants (e.g., ABE8 variants) that have increased efficiency and specificity. In particular, the adenosine deaminase variants

25 described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (e.g., “bystanders”).

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations. In some embodiments, an unintended editing or mutation is a bystander mutation or bystander editing,

30 for example, base editing of a target base (e.g., A or C) in an unintended or non-target position in a target window of a target nucleotide sequence. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system

comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced bystander editing or mutations by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing. In some embodiments, an unintended editing or mutation is a spurious mutation or spurious editing, for example, non-specific editing or guide independent editing of a target base (e.g., A or C) in an unintended or non-target region of the genome. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, e.g., ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE8 base editor variants described herein has reduced spurious editing by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 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 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 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 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%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% on-target base editing efficiency. In some embodiments, any of the ABE8 base editor variants described herein have on-target base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited target nucleobases in a population of cells.

In some embodiments, any of the ABE8 base editor variants described herein has higher on-target base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE8 base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher on-target base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9

fold, or at least 5.0 fold higher on-target base editing efficiency compared to an ABE7 base editor, e.g., ABE7.10.

The ABE8 base editor variants described herein may be delivered to a host cell via a plasmid, a vector, a LNP complex, or an mRNA. In some embodiments, any of the ABE8 base editor variants described herein is delivered to a host cell as an mRNA. In some embodiments, an ABE8 base editor delivered via a nucleic acid based delivery system, e.g., an mRNA, has on-target editing efficiency of at least at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases. In some embodiments, an ABE8 base editor delivered by an mRNA system has higher base editing efficiency compared to an ABE8 base editor delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at

least 4.9 fold, or at least 5.0 fold higher on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of base editor systems comprising one of the ABE8 base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% off-target editing in the target polynucleotide sequence.

In some embodiments, any of the ABE8 base editor variants described herein has lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least about 2.2 fold decrease in guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the ABE8 base editor variants described herein has lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at

least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE8 base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 5.0 fold, at least 10.0 fold, at least 20.0 fold, at least 50.0 fold, at least 70.0 fold, at least 100.0 fold, at least 120.0 fold, at least 130.0 fold, or at least 150.0 fold lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein has 134.0 fold decrease in guide-independent off-target editing efficiency (e.g., spurious RNA deamination) when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE8 base editor variants described herein does not increase guide-independent mutation rates across the genome.

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 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 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 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 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
5 embodiments, the base 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
10 editing method described herein results in at least 75% of a cell population that have been successfully edited. In some embodiments, the 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
15 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 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.

20 In some embodiments, the live cell recovery following a base editing intervention is greater than at least 60%, 70%, 80%, 90% of the starting cell population at the time of the base editing event. In some embodiments, the live cell recovery as described above is about 70%. In some embodiments, the live cell recovery as described above is about 75%. In some
25 embodiments, the live cell recovery as described above is about 80%. In some embodiments, the live cell recovery as described above is about 85%. In some embodiments, the live cell recovery as described above is about 90%, or about 91%, 92%, 93%, 94% 95%, 96%, 97%, 98%, or 99%, or 100% of the cells in the population at the time of the base editing event.

In some embodiments the 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-
30 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 100-fold.

The number of intended mutations and indels can be determined using any suitable method, for example, as described in International PCT Application Nos. PCT/2017/045381 (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:eaao4774 (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 methods of using the fusion proteins provided herein.

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

Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference. 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 ***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. In embodiments, multiplex editing involves administering to a cell and/or subject one or more base editors in combination with two or more distinct gRNAs (e.g., gRNA37 and gRNA40). Any two or more of the guide RNAs listed in Table 1, or gRNAs comprising a spacer sequence(s) listed in **Tables 2A-2C** or in the Sequence Listing (e.g., SEQ ID NOs: 3105-5485 and 8220-10830) can be administered simultaneously to a subject or cell for multiplexed editing. 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 can comprise one or more guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more base editor systems. In some embodiments, the multiplex editing can comprise one or more base editor systems with a single guide polynucleotide or a plurality of guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides with a single base editor system. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that does or does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing can comprise a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any 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.

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
5 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 coding region and at least one protein non-coding region.

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

25 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
30 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.

DELIVERY SYSTEM

The suitability of nucleobase editors to target one or more nucleotides in a gene (e.g., the HBV core protein (Core), the HBV polymerase gene (Pol), the HBV surface protein (S), or the HBV protein X gene (X)) is evaluated as described herein. In one embodiment, a single cell of interest is transfected, transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a 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 hepatocytes and HEK293 cells. 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

5 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 is assessed as described herein, i.e., by sequencing the genome of the cells to detect alterations in a target sequence. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed,
10 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 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
15 throughput sequencing (for example on an Illumina MiSeq). The fusion proteins 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 interest. In one embodiment, a nucleobase editor of the invention is delivered to cells
20 (e.g., hepatocytes or HEK293 cells) 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., the HBV core protein (Core), the HBV polymerase gene (Pol), the HBV surface protein (S), or the HBV protein X gene (X)). In some
25 embodiments, a base editor is targeted by one or more guide RNAs to introduce one or more edits to the sequence of one or more genes of interest (e.g., the HBV core protein (Core), the HBV polymerase gene (Pol), the HBV surface protein (S), or the HBV protein X gene (X)). In some embodiments, the one or more edits to the sequence of one or more genes of interest decrease or eliminate expression of the protein encoded by the gene in the host cell (e.g., a hepatocyte or HEK293 cell). In some embodiments, expression of one or more proteins
30 encoded by one or more genes of interest (e.g., the HBV core protein (Core), the HBV polymerase gene (Pol), the HBV surface protein (S), or the HBV protein X gene (X)) is completely knocked out or eliminated in the host cell (e.g., hepatocyte or HEK293 cell).

In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell.

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 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 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).

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-Dioleyloxy)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-Dioleyloxy-N-[2(sperminocarboxamido-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
Dimyristoxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbonyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyl-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
Diioctadecylamidoglycylspermidin	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
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolium chloride	DOTIM	Cationic
N1 -Cholesterylloxycarbonyl-3,7-diazaanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoyleoxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoylel-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoylel-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 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	YES	Transient	NO	Nucleic Acids and Proteins

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Viral	Phosphate transfection				
	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., a hepatocyte) 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 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.

5 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, 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.
10 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
15 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
20 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 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
25 suitable capsid protein of any viral vector. Thus, in some aspects, the disclosure relates to the viral delivery of a fusion protein. 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
30 be used to genetically modify the cell.

Viral Vectors

A base editor described herein can therefore be delivered with viral vectors. 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 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, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, 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 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 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. Adeno-associated virus ("AAV") vectors can also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (See, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). 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).

AAV is a small, single-stranded DNA dependent virus belonging to the parvovirus family. The 4.7 kb wild-type (wt) AAV genome is made up of two genes that encode four replication proteins and three capsid proteins, respectively, and is flanked on either side by 145-bp inverted terminal repeats (ITRs). The virion is composed of three capsid proteins, Vp1, Vp2, and Vp3, produced in a 1:1:10 ratio from the same open reading frame but from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. A phospholipase domain, which functions in viral infectivity, has been identified in the unique N terminus of Vp1.

Similar to wt AAV, recombinant AAV (rAAV) utilizes the cis-acting 145-bp ITRs to flank vector transgene cassettes, providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. Although there are numerous examples of rAAV success using this system, in

vitro and in vivo, the limited packaging capacity has limited the use of AAV-mediated gene delivery when the length of the coding sequence of the gene is equal or greater in size than the wt AAV genome.

Viral vectors can be selected based on the application. For example, for in vivo gene 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 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. In some embodiments, the disclosed base editors are 4.5 kb or less in length.

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

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

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

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

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.

Inteins

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.

- 5 While the endogenous substrates of protein splicing are proteins found in intein-containing organisms, inteins can also be used to chemically manipulate virtually any polypeptide backbone.

In protein splicing, the intein excises itself out of a precursor polypeptide by cleaving two peptide bonds, thereby ligating the flanking extein (external protein) sequences via the
10 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.

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,
15 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 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
20 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 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
25 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 excises the intein from the reactive complex
30 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, a portion or fragment of a nuclease (e.g., Cas9) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an

AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, an N-terminal fragment of a base editor (e.g., ABE, CBE) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. 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 deaminase 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 kfkvlgntr hsikknliga llfdsgetae
61 atrlkrtrarr rytrkrnric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
301 llSdilrvnT eiTkaplsas mikrydehhq dltllkalvr qqlpekykei ffdqSkngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
421 ailrrqedfy pflkdnreki ekiltfripy yvgplArgnS rfAwmTrkSe eTiTpwnfee
481 vvdkgasaqs fiermtndfk nlpnekvlpk hsllyeyftv yneltkvkvv tegmrkpafll
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kieCfdSvei sgvedrfrnAS lgtyhdllki
601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfdckvmkq lkrrrytgwg
661 rlsrklngi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediql aqvsgggdsl
721 hehianlags paikkgilqt kvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
841 ivpqsflkdd sidknvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrnm tkydendkli revkvitlks

```

961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvkr
 1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
 1081 atvrkvlsmq qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
 1141 ysvlvvakve kgkskklksv kelligitime rssfeknpid fleakgykev kkdliiklpk
 5 1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqqlfve
 1261 qhkhyldiei eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
 1321 paafkyfddt idrkrytstk evldatlihq sitglyetri dlsqllggd (SEQ ID NO: 197)

Pharmaceutical Compositions

10 In some aspects, the present invention provides a pharmaceutical composition comprising any of the genetically modified cells, base editors, fusion proteins, or the fusion protein-guide polynucleotide complexes described herein.

In embodiments, a pharmaceutical composition of the present disclosure comprises an antiretroviral drug (e.g., lamivudine) for treatment of a retroviral infection (e.g., an HBV
 15 infection). Non-limiting examples of antiretroviral drugs include adefovir, tenofovir, telbivudine, interferons (e.g., Interferon alfa-2b (Intron A), Peg-Interferon), Lymphotoxin β receptor (LTBR), nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) (e.g., abacavir (Ziagen), Didanosine, emtricitabine (Emtriva), lamivudine (Epivir), Stavudine, tenofovir disoproxil fumarate (Viread), Tenofovir alafenamide, and zidovudine (Retrovir)),
 20 non-nucleoside reverse transcriptase inhibitors (NNRTIs) (e.g., Delavirdine, doravirine (Pifeltro), efavirenz (Sustiva), etravirine (Intelence), nevirapine (Viramune), rilpivirine (Eduvant), and Delavirdine), post-attachment inhibitors or monoclonal antibodies (e.g., Ibalizumab-uiyk), CCR5 Antagonists (e.g., Maraviroc), Atazanavir, Darunavir, Cobicistat, Lopinavir, Maraviroc, Nevirapine, Rilpivirine, Elvitegravir + TDF + FTC + cobicistat,
 25 Elvitegravir + TAF + FTC + cobicistat, Integrase inhibitors (e.g., Bictegravir, Cabotegravir, Cabotegravir and rilpivirine, Dolutegravir, Elvitegravir, fosamprenavir (Lexiva, Telzir), and Raltegravir), Fusion inhibitors (e.g., Enfuvirtide), entry inhibitors (e.g., Maraviroc), protease inhibitors (PIs) (e.g., Atazanavir, Darunavir, Lopinavir, indinavir (Crixivan), lopinavir/ritonavir (Kaletra), ritonavir (Norvir), saquinavir (Invirase), and tipranavir
 30 (Aptivus)), attachment and post-attachment inhibitors (e.g., Fostemsavir, Ibalizumab), and Entecavir. In an embodiment, the antiretroviral drug is selected from one or more of Lamivudine; Entecavir; Tenofovir; Interferon; and Peg-Interferon.

The pharmaceutical compositions of the present invention can be prepared in accordance with known techniques. See, e.g., Remington, The Science And Practice of
 35 Pharmacy (21st ed. 2005). In general, the cell, population thereof, base editors, fusion protein, or any other component of the compositions of the disclosure 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 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
5 any amount suitable to maintain the pH of the formulation at a predetermined level.

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

The pharmaceutical compositions of the present invention can include at least one
20 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 nucleic acid sequence encoding a cytokine peptide. In some embodiments, the pharmaceutical compositions comprising the cell or population thereof can
25 be administered separately from an additional therapeutic agent.

One consideration concerning the therapeutic use of compositions of the invention is the quantity of an agent necessary to achieve an optimal or satisfactory effect. For example, the quantity of modified cells to be administered may vary for the subject being treated. In one embodiment, between 10^4 to 10^{10} , between 10^5 to 10^9 , or between 10^6 and 10^8 genetically
30 modified cells of the invention are administered to a human subject. In some embodiments, at least about 1×10^8 , 2×10^8 , 3×10^8 , 4×10^8 , and 5×10^8 genetically modified cells of the invention are administered to a human subject. Determining the precise effective dose may be based on factors for each individual subject, including their size, age, sex, weight, and

condition. Dosages can be readily ascertained by those skilled in the art from this disclosure and the knowledge in the art.

The skilled artisan can readily determine the amount of an agent and amount of optional additives, vehicles, and/or carriers in compositions and to be administered in methods of the invention. Typically, additives are present in an amount of 0.001 to 50 % (weight) solution in phosphate buffered saline, and the active ingredient is present in the order of micrograms to milligrams, such as about 0.0001 to about 5 wt%, preferably about 0.0001 to about 1 wt%, still more preferably about 0.0001 to about 0.05 wt% or about 0.001 to about 20 wt%, preferably about 0.01 to about 10 wt%, and still more preferably about 0.05 to about 5 wt %. Of course, for any composition to be administered to an animal or human, and for any particular method of administration, it is preferred to determine therefore: 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 cited herein. And, the time for sequential administrations can be ascertained without undue experimentation.

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 diseased site (e.g., the liver). 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.

- 5 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
10 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
15 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
20 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
25 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
30 stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. et al., Gene Ther. 1999, 6: 1438-47). Positively charged lipids such as N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, e.g., U.S. Patent Nos.

4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

The pharmaceutical composition described herein can be administered or packaged as a unit dose, for example. The term “unit dose” when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material 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, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided

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

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 are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

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

Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with

an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or
5 other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in
10 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.

15 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.

The compositions, as described above, can be administered in effective amounts. The
20 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.

25 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

Some aspects of the present invention provide methods of treating a subject in need,
30 the method comprising administering to a subject in need an effective therapeutic amount of a pharmaceutical composition as described herein. In other embodiments, the methods of the invention comprise expressing or introducing into a cell a base editor polypeptide and one or

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

In embodiments, the methods of treatment involve a pretreatment, co-treatment, or simultaneous treatment with an antiretroviral drug(s) (e.g., lamivudine) for treatment of a retroviral infection (e.g., an HBV infection). In embodiments, the treatment (e.g., pretreatment, co-treatment, or simultaneous treatment) has a duration of about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 days. The treatment has a duration of no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 days. In embodiments, administration of the antiretroviral drug stops about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days prior to (i.e., “pretreatment”) administration of a base editing system to a subject or cell administered the pretreatment. In embodiments, administration of the antiretroviral drug stops no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days prior to administration of a base editing system to a subject or cell administered the pretreatment. In embodiments, administration of the antiretroviral drug begins about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days prior to administration of a base editing system to a subject or cell administered the pretreatment. In embodiments, administration of the antiretroviral drug ends no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days prior to administration of a base editing system to a subject or cell administered the pretreatment. In embodiments, administration of the antiretroviral drug ends no more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days following administration of a base editing system and treatment. In embodiments, the antiretroviral drug is first administered to a subject or cell about or at least about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days prior to administration of a base editing system to the subject or cell and continues for about, at least about, or no more than about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 days following an administration of a base editing system, where in some instances the administration of the base editing system is a first administration of the base editing system.

Non-limiting examples of antiretroviral drugs include adefovir, tenofovir, telbivudine, interferons (e.g., Interferon alfa-2b (Intron A) or Interferon alpha-2a; pegylated interferons (e.g., peginterferon alpha-2a or peginterferon alpha-2b)), Lymphotoxin β receptor (LTBR), nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) (e.g., abacavir (Ziagen), Didanosine, emtricitabine (Emtriva), lamivudine (Epivir), Stavudine, tenofovir disoproxil

fumarate (Viread), Tenofovir alafenamide, and zidovudine (Retrovir)), non-nucleoside reverse transcriptase inhibitors (NNRTIs) (e.g., Delavirdine, doravirine (Pifeltro), efavirenz (Sustiva), etravirine (Intelence), nevirapine (Viramune), rilpivirine (Edurant), and Delavirdine), post-attachment inhibitors or monoclonal antibodies (e.g., Ibalizumab-uiyk),
 5 CCR5 Antagonists (e.g., Maraviroc), Atazanavir, Darunavir, Cobicistat, Lopinavir, Maraviroc, Nevirapine, Rilpivirine, Elvitegravir + TDF + FTC + cobicistat, Elvitegravir + TAF + FTC + cobicistat, Integrase inhibitors (e.g., Bictegravir, Cabotegravir, Cabotegravir and rilpivirine, Dolutegravir, Elvitegravir, fosamprenavir (Lexiva, Telzir), and Raltegravir), Fusion inhibitors (e.g., Enfuvirtide), entry inhibitors (e.g., Maraviroc), protease inhibitors
 10 (PIs) (e.g., Atazanavir, Darunavir, Lopinavir, indinavir (Crixivan), lopinavir/ritonavir (Kaletra), ritonavir (Norvir), saquinavir (Invirase), and tipranavir (Aptivus)), attachment and post-attachment inhibitors (e.g., Fostemsavir, Ibalizumab), and Entecavir. In an embodiment, the antiretroviral drug is selected from one or more of Lamivudine; Entecavir; Tenofovir; Interferon; and Peg-Interferon.

15 In embodiments, pretreatment, co-treatment, or simultaneous treatment of a subject and/or cell with one or more of the antiretroviral drug(s) (e.g., in combination with a base editor system of the present disclosure) is associated with an increase in base editing efficiency of a target sequence(s) in a polynucleotide (e.g., in cccDNA) of about or at least about 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%,
 20 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300%, 400%, 500%, or more relative to a reference subject and/or cell not administered the antiretroviral drug.

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,
 25 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 edited cell or a base editor system or polynucleotide encoding such system. In any of such methods, the methods may comprise administering one or more doses
 30 of an effective amount of a base editor system, optionally wherein the doses are separated by a period of about or at least about 1 hour, 12 hours, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, 10 days, 11 days, 12 days, 13 days, 14 days, or more. In any of such methods, the methods may comprise administering two or more doses of an effective amount of a base editor system.

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, 5 intratumorally, intradermally, intraperitoneally, transtracheally, subcutaneously, subcuticularly, intraarticularly, subcapsularly, subarachnoidly and intrasternally.

In some embodiments, a composition and/or agent described herein (e.g., edited cell, base editor system, and/or antiretroviral drug) 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 10 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 15 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 20 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 25 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 30 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 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 administered intravenously once a day, seven times in a seven day period.

In some embodiments, the composition is administered over a period of about, of at least about, or less than about, substantially instantaneously (e.g., over a period of less than 1 second, as may be the case when a composition is administered as an injection), 1 second, 10 seconds, 30 seconds, 1 minute, 0.1 h, 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 administered over a period of 2 h.

Methods of Treating HBV Infection

The present invention provides methods of treating an HBV infection or symptoms thereof that comprise administering to a subject (e.g., a mammal, such as a human) a therapeutically effective amount of a pharmaceutical composition that comprises a polynucleotide encoding a base editor system (e.g., base editor and gRNA) described herein. In some embodiments, the base editor is a fusion protein that comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain or a cytidine deaminase domain. A cell of the subject is transduced with the base editor and one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration (if the cell is transduced with an adenosine deaminase domain) or a C•G to U•A alteration (if the cell is transduced with a cytidine deaminase domain) of a nucleic acid sequence encoding an HBV polypeptide.

In some embodiments, treatment of chronic Hepatitis B includes a combination of approaches. For example, a subject infected with HBV can be administered a therapeutically effective amount of a pharmaceutical combination described above that targets and modifies HBV cccDNA. For example, a BE4 base editor without a UGI domain can be used with a gRNA targeting a region of the HBV genome. Without being bound by theory, omitting the inhibitor makes C->U deamination susceptible to uracil glycosylase which damages HBV cccDNA, thus destabilizing it. This treatment can be combined with a treatment that reduce or inhibit HBsAg expression (including from integrated HBV DNA), such as targeting the S or pol gene of the HBV genome using the base editors and guide RNAs described herein. Treatment can also comprise stimulating the immune system. In some embodiments, each of

these three distinct therapeutic goals can be accomplished using the base editing reagents and techniques described herein.

The methods herein include administering to the subject (including a subject identified as in need of such treatment) an effective amount of a composition described herein. Identifying a subject in need of such treatment can be in the judgment of a subject or a health care professional and can be subjective (e.g. opinion) or objective (e.g. measurable by a test or diagnostic method).

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

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

Kits

The invention provides kits for the treatment of an HBV infection, as well as related diseases and disorders, including cirrhosis, hepatocellular carcinoma (HCC), and any other disease associated with or resulting from HBV infection, in a subject. In some embodiments, the kit further includes a base editor system or a polynucleotide encoding a base editor

system, wherein the base editor polypeptide system 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 base editor is a mRNA sequence. In some embodiments, the deaminase is a cytidine deaminase or an adenosine deaminase. In some embodiments, the kit comprises an edited cell and instructions regarding the use of such cell.

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

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 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 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 inventors regard as their invention.

EXAMPLES

5 **Example 1: Targeted base editing of the HBV genome resulted in missense mutations or alterations in transcription start sites in the HBV genome.**

Guide RNAs (see **Tables 1 and 2**) were designed *in silico* to target a Cas12b-ABE base editor to introduce missense mutations or transcription site nucleobase alterations at conserved regions of HBV genes (see **Table 20A**). HBV Genotype D, subgenotype ayw, was
10 analyzed using a CRISPR design tool available on Benchling Software to identify potential guide RNAs, and the level of sequence conservation across all HBV genotypes was determined for each guide RNA. 13 guide RNAs were selected for testing (Tables 1 and 2). These guide RNAs were adjacent or in proximity to RTTN PAM sequences. For purposes of comparison, additional base editor systems indicated in **FIG. 4** were evaluated. The
15 additional base editor systems contained NGG-ABE8.8-m, Cas12b-ABE, VRQR-ABE, or NGC-ABE base editors and guide RNAs listed in **Table 1**. Conservation across HBV genomes of the sequences targeted by these additional base editor systems as well as the functional changes resulting from base edits effected by the additional base editors are provided below in **Tables 20B and 20C**.

20 The gRNAs were selected based on the high conservation across HBV genotypes, with a focus on genotypes D and A. Such a design strategy targeting highly conserved regions allowed for the design of gRNAs suitable for targeting a wider patient population and that avoid the appearance of viral mutations that would enable HBV virus to escape base editing treatment.

25 A selection criterion used in the design of the gRNAs for introducing missense mutations was the ability to generate edits in the HBV genome that would lead to a missense mutation. Further, an *in silico* analysis of the amino acid changes generated by the base editing reagents, and gRNAs were selected that would generate missense amino acid changes with sequences rarely detected in naturally occurring HBV genomes (<0.05% frequency of
30 amino acid change). This strategy increased the likelihood that a gRNA would introduce a mutation leading to a disruption of the HBV protein/genome.

Table 20A: Conservation across HBV genomes of sequences targeted by guide RNAs.

Guide ID	HBV gene targeted	Conservation of target site across HBV genotypes A-H
cas12b-4	Pol/S	>90% A,C,D,G
cas12b-5	Pol/S	>85% A, B, D
cas12b-11	Pol/S	>90% A, D, G, H
cas12b-17	Pol/S	>60% A, D
cas12b-30	Pol	>90% all genotypes
cas12b-42	Core	>65% A,D,E
cas12b-100	Pol	>80% A,B,C,D,E
cas12b-147	Pol/S	>75% A, B, D, E
cas12b-154	Pol/S	>80% A-E, G

Table 20B: Conservation across HBV genomes of sequences targeted by guide RNAs, editing efficiencies associated with the guide RNAs, and functional changes associated with genome alterations effected by base editing systems containing the guide RNAs

Guide ID	Base Editor	PAM	HBV gene	% Editing average	Conservation across HBV genomes	Functional Change
cas12b-152	Cas12b-ABE	ATTG	S / Pol	54%	>90% in all genotypes	missense mutations in both S and Pol
ABE-41*	Abe7.10	NGG	S / Pol	58%	>90% in all genotypes	missense mutations in both S and Pol
ABE-85*	Abe7.10	NGG	S / Pol	60%	>90% in all genotypes	missense mutations in both S and Pol
ABE-133*	VRQR-ABE	NGA	S	65%	>90% in all genotypes	missense mutations in S
MSPbeam94*	Abe7.10	NGG	S / Pol	73%	>90% in all genotypes	active site of the pol, missense in S
G-polyA-NGC1	NGC-ABE	NGC	polyA	75%	~50% in genotypes A,D	
G-polyA-NGC2	NGC-ABE	NGC	polyA	67%	~9% in genotype D	
G-251	Abe7.10	NGG	Enhancer I	77%	>50% in genotypes D, E	targeting cccDNA transcription Enhancer I
G-159	Abe7.10	NGG	Enhancer I	51%	>90% in genotypes A, D, E, G	targeting cccDNA transcription Enhancer I
G-553	ABE8.8-m	NGG	Enhancer I	64%	~70% in genotypes D, E	targeting cccDNA transcription Enhancer I
G-845	Abe7.10	NGG	Enhancer IIb	70%	>90% in genotypes B, C, D, E	targeting cccDNA transcription Enhancer IIb
G-200	ABE8.8-m	NGG	HBX promoter	59%	>90% in genotypes B-F	targeting HBX promoter
G-772	Abe7.10	NGG	S promoter	66%	~15% in genotype D	targeting S promoter
G-474	VRQR-ABE	NGA	Basal Core	74%	~20% in genotype D, >60% genotype E	targeting basal core promoter

Table 20C: Conservation across HBV genomes of sequences targeted by guide RNAs.

Guide ID	Base Editor	PAM	HBV gene	Conservation across HBV genotypes A-H
ABE-8*	VRQR-ABE	NGA	S	97%
ABE-41*	ABE7.10/ABE8	NGG	Pol/S	99%
ABE-85 (MSPbeam36)*	ABE7.10/ABE8	NGG	Pol/S	92%
ABE-131 (EMS11)*	VRQR-ABE	NGA	Core	98%
ABE-133*	VRQR-ABE	NGA	S	99%
ABE-142 (MSPbeam66)*	VRQR-ABE	NGA	Pol/S	99%
ABE-150*	ABE7.10/ABE8	NGG	Pol/X	98%
ABE-235 (EMS21)*	ABE7.10/ABE8	NGG	Pol/S	96%
cas12b-116	Cas12b-ABE	ATTG	Pol/X	>75% A, B, D, F
cas12b-122	Cas12b-ABE	GTTG	Pol	>85% A, B, D, F, G
cas12b-148	Cas12b-ABE	GTTG	Pol/S	>90% A, D
cas12b-152	Cas12b-ABE	ATTG	Pol/S	>85% all genotypes

Gene editing efficiency was evaluated as follows. For plasmid transfections, a vector encoding guide RNA and a vector encoding a base editor were transiently transfected into HEK293T cells previously transduced with a lentiviral vector comprising the HBV genome (i.e., a Hek293-lentiHBV system). The base editor was Cas12b-ABE (bhCas12b-ABE8.13 D952A, S893R, K846R, E837G).

Transfection was carried out using a high efficiency, low toxicity DNA transfection reagent (Mirus TransIT293 or Lipofectamine 2000) optimized for Hek293 cells in a 3:1 ratio using 250 ng of gRNA plasmid and 750 ng of base editor plasmid. After four days for plasmid transfections, genomic DNA was extracted with a simple lysis buffer of 0.05% SDS, 25 µg/ml proteinase K, and 10 mM Tris pH 8.0, followed by heat inactivation at 85°C. Genomic sites were PCR amplified and sequenced on a MiSeq. Editing efficiencies determined for each gRNA are presented in **Table 21** below and in **FIG. 4**.

Table 21. Base editing efficiencies.

Guide ID	Editing Efficiency 1 (%)	Editing Efficiency 2 (%)	Editing Efficiency 3 (%)	Average Editing Efficiency (%)
cas12b-4	14.70	26.37	18.93	20.00
cas12b-5	0.12	0.36	1.14	0.54
cas12b-11	44.34	60.28	69.47	58.03
cas12b-17	26.42	34.50	44.29	35.07
cas12b-30	0.02	0.02	0.01	0.02
cas12b-42	29.48	23.94	33.54	28.99
cas12b-100	7.97	4.21	3.35	5.18
cas12b-147	13.02	4.44	15.03	10.83
cas12b-154	16.12	30.06	37.02	27.74
cas12b-35	0.31	0.04	0.70	0.35
cas12b-124	29.50	11.78	11.15	17.47
cas12b-127	50.16	37.40	39.21	42.25
cas12b-122b	8.22	3.91	4.39	5.51

The guides cas12b-4, cas12b-5, cas12b-11, cas12b-17, cas12b-30, cas12b-42, cas12b-100, cas12b-147, and cas12b-154 targeted the base editor to introduce missense mutations into their respective target sequences (see **Table 2A**). The guide cas12b-35 targeted the base editor to introduced a nucleobase alteration in enhancer II box A. The guides cas12b-124 and cas12b-127 each targeted the base editor to introduce a nucleobase alteration in enhancer I. The guide cas12b-122b targeted the base editor to introduce a nucleobase alteration in the HBX promoter.

Example 2: Antiviral and Editing Efficacy of Base Editor Systems

Base editor systems comparing guide RNAs evaluated in Example 1 were subsequently tested for anti-viral and editing efficacy using the Hepatitis B virus-primary human hepatocyte (HBV-PHH) system. Cells were transfected with the guide RNAs and mRNAs encoding the base editor as described in Example 1. It was determined (**FIGs. 5A-5C** and **Table 22**) that base editor systems comprising ABE-133, ABE-41, MSPbeam94, ABE-85, PolyA-NGC1, PolyA-NGC2, Cas12b-11, and/or Cas12b-152 guide RNA effectively edited covalently-closed circular HBV DNA (cccDNA) and/or reduced viral replication in the HBV-PHH system, as indicated by a robust reduction in all HBV markers relative to the no treatment control. Guide RNAs ABE-133, MSPbeam94, and ABE-85 were associated with a 70-80% reduction of HBV DNA as compared to controls. The guide RNAs

PolyA-NGC1 and PolyA-NGC2 were associated with high editing followed by 90% efficacy in reduction of HBeAg.

Table 22: Base editing efficiencies associated with the indicated guide RNAs

gRNA	cccDNA Editing in PHH A>G%
ABE-133	76.0%
ABE-41	56.0%
MSPbeam94	41.0%
ABE-85	29.0%
PolyA-NGC1	60.4%
PolyA-NGC2	62.9%

5

The HBV-PHH system for studying the Hepatitis B virus is described, for example, by Winer, et al., “Long-term hepatitis B infection in a scalable hepatic co-culture system,” *Nat. Commun.* 8:125 (2017) and by Zhou, et al. “Long-term maintenance of human fetal hepatocytes and prolonged susceptibility to HBV infection by co-culture with non-parenchymal cells,” *J. Virol Methods*, 195:185-93 (2014); and is provided commercially by ImQuest BioSciences (imquestbio.com/programs/virology/hbv/).

10

Example 3: Multiplexing gRNAs reduced HBV viral parameters in HepG2-NTCP Cells

Experiments were completed using guide RNA's listed in Table 1 to identify combinations of guide RNA's effective in treating HBV-infected cells. HepG2-NTCP cells were transfected with mRNA encoding BE4 and a guide RNA(s) selected from gRNA12, gRNA37, gRNA19, gRNA190, and gRNA40. HepG2-NTCP cells are described in Sun, *et al.* “NTCP-Reconstituted In Vitro HBV Infection System,” *Methods Mol. Biol.* 1540:1-14 (2017), DOI: 10.1007/978-1-4939-6700-1_1, the disclosure of which is incorporated herein by reference in its entirety for all purposes. Cells were infected at day 0 with HBV and then transfected with the mRNA and gRNA(s) at day 4, see **FIG. 6**. Measurements of HBsAg, HBeAg, total HBV DNA, and 3.5 kb viral DNA were taken at day 14, see **FIG. 6**. The guide RNA, gRNA37, targeted HBsAg and gRNA40 targeted the HBV core protein, see **FIG. 7**. The use of gRNA37 was associated with a reduction in HBsAg levels and the use of gRNA40 was associated with a reduction in HBeAg levels, see **FIG. 8**. Also, each of gRNA37 and gRNA40 was associated with a lowering of total HBV DNA and 3.5 kb RNA levels. As

15

20

25

shown in **FIG. 8**, guide RNAs gRNA37 and gRNA40 were identified as particularly effective gRNAs.

Having identified gRNA37 and gRNA40 as effective gRNAs, the antiviral efficacy of a base editor system containing BE4, gRNA37 and gRNA40 (i.e., multiplexing gRNAs) was evaluated, **FIGs. 8-10**. It was determined that the base editor system multiplexed with the two gRNAs was associated with a reduction of HBV viral parameters in HepG2-NTCP cells, see **FIG. 9**. As demonstrated in **FIG. 10**, the base editor system functioned through cccDNA editing, without reducing cccDNA level, and the base editor system (which contained gRNA37 and gRNA40) was associated with robust cccDNA editing (about 38% for gRNA37 and about 60% for gRNA40). For the evaluation of editing of cccDNA, the cccDNA was enriched via Hirt extraction and ExoI/ExoIII nuclease treatment to remove rcDNA (relaxed circular DNA). Throughout the Examples provided herein, editing of cccDNA was evaluated using next generation sequencing completed using cccDNA enriched in this way. Thus, it was demonstrated that it is advantageous in terms of observed antiviral effect to administer multiple distinct gRNA molecules (e.g., gRNA37 and gRNA40) in combination with a base editor (e.g., BE4) to modify the HBV genome.

Example 4: Antiviral efficacy of base editing in combination with lamivudine treatment

Experiments were undertaken to determine whether there may be any advantage to combining the base editor systems described herein with an antiviral medication (specifically, an antiretroviral medication), such as lamivudine.

First, experiments were carried out as described in the schematic provided in **FIG. 11**. HepG2-NTCP cells were infected with HBV at day 0, at day 4 an 8-day treatment with lamivudine commenced (which included a 3-day pretreatment period within the 8-day total), at day 7 cells were transfected with mRNA encoding BE4 and gRNAs 37 and 40, and measurements were taken at day 15. As shown in **FIGs. 12 and 13**, it was found that the 3-day pretreatment (8-day total treatment duration) with lamivudine was associated with an improvement in base editing of about 20% (**FIG. 12**). Not being bound by theory, pretreatment with lamivudine removes intermediate HBV DNA species, leaving only cccDNA. This improvement was further associated with high antiviral efficacy in the HepG2-NTCP cells. Not being bound by theory, high covalently-closed circular DNA (cccDNA) editing in lamivudine pretreated conditions is consistent with the base editor

system directly targeting cccDNA. Treatment with the combination of gRNA38 and gRNA40 lead to a robust reduction of HBV viral markers (**FIG. 13**).

Example 5: The base editors are more effective than an antiviral medication

Given that HBV infections can be treated using antiretroviral medications, such as lamivudine, experiments were undertaken to determine the efficacy of the base editor systems relative to treatment with lamivudine.

The design of the experiments is described schematically in **FIG. 14**. Primary human hepatocyte (PHH) cells were infected with HBV at day 0. At days 4 and 11, the cells were transfected with mRNA encoding BE4 and with gRNA37 and gRNA40. Measurements of HBV DNA were taken over time, and measurements of HBsAg, HBeAg, HBV total DNA, 3.5 kb RNA, and cccDNA editing were taken at day 25. The experiments were carried out using a PHH co-culture system available from PhoenixBio. The PHH co-culture system maintained hepatocyte differentiation and metabolic activity for over 30 days. The HBV infection in the cells was persistent and a cccDNA level was maintained.

As shown in **FIG. 16**, HBV rebounded after discontinuation of lamivudine, but HBV did not rebound for at least 2 weeks after the second transduction with base editing reagents. Further, cells transfected with the base editor system showed a reduction in all HBV markers by between about 70% and about 80% at day 25, see **FIG. 15**. Thus, the base editor system containing BE4 and the two guide RNAs (gRNA37 and gRNA40) was more effective in controlling the HBV infection than treatment with an antiretroviral medication.

Further, the base editor system was associated with an editing efficiency of about 30% at the S antigen site and about 60% at the PreCore gene site, see **FIG. 17**. These editing levels are sufficient to enable high antiviral efficacy and prevent rebound of HBV in primary human hepatocyte (PHH) cells.

Cytosine base editing resulted in gRNA-specific reductions of HBV viral markers in relevant *in vitro* systems. Multiplex editing led to a simultaneous reduction of HBsAg, HBeAg, HBV DNA, and 3.5 kb RNA in HepG2-NTCP cells and primary hepatocytes. Not being bound by theory, the reduction in viral markers was likely driven by base editing of cccDNA, but not a reduction in cccDNA levels. Combined treatment of cells with the base editing reagents and standard antiviral lamivudine resulted in higher base editing efficiencies. Base editing was associated with the prevention of HBV rebound in infected primary

hepatocytes. Not being bound by theory, cytosine base editing introduced permanent mutations in cccDNA preventing HBV rebound in clinically relevant *in vitro* models.

Example 6: Influence of lamivudine on base editing efficiency in primary hepatocytes

Experiments were undertaken to further investigate the influence of the antiretroviral drug lamivudine on base editing using base editing systems containing BE4 and gRNAs 37 and 40.

The design of the experiments is described schematically in **FIG. 18**. Primary human hepatocyte (PHH) cells were infected with HBV at day 0, at day 4 an 8-day. Treatment with lamivudine commenced (which included a 3-day pretreatment period within the 8-day total), and cells were transfected with mRNA encoding BE4 and the gRNAs gRNA37 and gRNA40 at days 7 and 14. HBV DNA levels were measured over time, and additional HBV markers were measured at day 21.

Pretreatment with lamivudine improved editing in primary hepatocytes, see **FIGs. 19 and 20**. Combined treatment with lamivudine and the base editor system prevented HBV rebound, see **FIG. 19**. As shown in **FIG. 21**, the combined treatment and administration of the base editor system alone both resulted in a reduction in HBV markers by between about 70% and 80% at day 21.

Example 7: Base editing efficiency and antiviral efficacy of base editors with reduced off-target activity

It may be advantageous to use base editors that have reduced off-target activity and lower non-specific effects (e.g., as compared to BE4) without compromising editing efficiencies. Thus, experiments were undertaken to evaluate the antiviral efficacy of base editor systems containing ppApo or rrA3F, which are described in Yu, *et al.* "Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity," *Nature Communications*, 11:2052 (2020), DOI: 10.1038/s41467-020-15887-5, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

It was found, as shown in **FIGs. 22 and 23**, that base editor systems containing BE4, ppApo, or rrA3F and gRNA37 and gRNA40 had comparable editing efficiencies and were associated with reductions in the HBV marker HBsAg.

Example 8: Cytosine base editing inhibited Hepatitis B Virus replication and reduced HBsAg expression *in vitro* and *in vivo*

A base editing strategy was devised to introduce stop codons in HBV genes (e.g., HBs, and Precore) using two distinct gRNAs. The antiviral efficacy of this approach was assessed *in vitro* in two different HBV-infected cell models: HepG2-NTCP and primary human hepatocytes. HepG2-NTCP and primary human hepatocytes were transfected with an mRNA encoding CBE BE4 and with gRNA37 and gRNA40. This transfection resulted in robust cccDNA editing (e.g., up to 50% editing in a gene encoding the Hepatitis B surface antigen (HBsAg) and up to 80% editing in the Precore gene), leading to a reduction in the encoded viral markers. Transfection with CBE mRNA and both gRNAs led to a sustained reduction in each of the following: HBsAg, HBeAg, which is an HBV viral protein, 3.5kb viral RNA, and total intracellular HBV DNA. This significant reduction in HBsAg, HBeAg, 3.5kb viral RNA, and total intracellular HBV DNA was observed when the CBE and gRNAs were administered as a single treatment and when they were used in combination with nucleoside analogs.

Having established the efficacy of these base editing reagents *in vitro*, experiments were then undertaken to evaluate these same base editing reagents for antiviral efficacy *in vivo* using the HBV minicircle mouse model. This *in vivo* model exhibits persistent HBV replication and antigen expression resulting from hydrodynamic injection with cccDNA-like plasmid. After establishing HBV viral replication, hepatic delivery of the base editing reagents was achieved via systemic administration of lipid nanoparticles (LNPs) formulated with the following base editing reagents: CBE mRNA and HBV-targeting gRNAs (e.g., gRNA37 and gRNA40 (see **Table 1**)). The mice were administered a combination of base editors, where the combination was either gRNA37 and gRNA40 or gRNA37-HM and gRNA40-HM. The guides gRNA37-HM and gRNA40-HM each included heavily modified(HM) scaffolds (see **Table 1**).

Intravenous injection of the lipid nanoparticles (LNPs) containing the base editing reagents resulted in a significant reduction in mouse serum HBV DNA, HBsAg and HBeAg. This reduction was more pronounced when a first injection was given and a second injection was given two weeks later (**FIGs. 24A-24C and 25A-25C**). Two injections of BE4/gRNA(37+40) LNP at 2mg/kg led to an approximately 2 log reduction in HBsAg and HBV DNA in the mouse serum, and undetectable levels of HBeAg. It was also found that it was beneficial to administer guide RNAs with heavily modified scaffolds (see, e.g., **FIG. 25A**). Mice administered BE4/gRNA(37+40)-HM (Heavy mod. gRNA) at days 1 and 14

also showed an approximately 2 log reduction in HBsAg as compared to controls. All single-dose (i.e., only at D1) and double-dose (i.e., at D1 and D14) treatments with the base editing reagents led to reductions in HBeAg to levels below detection after 2 weeks. Mice treated with either Entecavir or the base editing reagents showed a ~2 log reduction in serum levels of HBV DNA at day 21.

Table 23 below provides a summary of the experimental design used to evaluate the base editing reagents *in vivo*. The mice contained an HBVcircle polynucleotide that enabled HBV replication *in vivo*. See, for example Yan et al., J. Hepatol 2017 66:1149-1157 for a description of an exemplary mouse model. The mice were administered lipid nanoparticles (BL4 LNP) containing mRNA encoding BE4 (a cytidine base editor) and either the guide RNAs gRNA37 and gRNA40 or the guide RNAs gRNA37-HM and gRNA-40-HM containing a scaffold with heavy modifications (HM). Efficacy of treatment using the base editing reagents was compared to efficacy of treatment using Entecavir. The mice were administered Entecavir daily on days 1 through day 14, or the mice were administered the base editing reagents either once on only day 1 or twice with the first administration being on day 1 and the second administration being on day 14. Entecavir (ETV) was administered orally. HBsAg, HbeAg, and HBV DNA levels in serum were measured weekly (**FIGs. 24A-24C and 25A-25C**).

Table 23. Experimental design for evaluating base editing reagents *in vivo*.

Group	Number of Mice	Drug / LNP Formulation	Base editor (BE)/ gRNA	Dose (RNA, mg/kg)	Schedule
1	5	-	-	-	
2	5	ETV		0.03mg/kg	D1-D14
3	5	BL4 LNP	BE4 / mus.PCSK9 (control)	2 mg/kg	D1
4	5	BL4 LNP	BE4 / mus.PCSK9 (control)	2 mg/kg x 2	D1, D14
5	5	BL4 LNP	BE4/gRNA(37+40)	2 mg/kg	D1
6	5	BL4 LNP	BE4/gRNA(37+40)	2 mg/kg x 2	D1, D14
7	5	BL4 LNP	BE4/gRNA(37+40)-HM	2 mg/kg	D1
8	5	BL4 LNP	BE4/gRNA(37+40)-HM	2 mg/kg x 2	D1, D14

“LNP” indicates “lipid nanoparticle,” group 1 corresponded to untreated mice, “D1” indicates “day 1,” “D14” indicates “day 14,” “Schedule” indicates days of administration, “ETV” indicates “Entecavir,” “BE” indicates “base editor,” “x2” indicates administration at two different days (i.e., D1 and D14), “gRNA(37+40)” indicates

a combination of gRNA37 and gRNA40, “gRNA(37-40)-HM” indicates a combination of gRNA37-HM and gRNA-40-HM, and BL4 indicates a lipid nanoparticle formulation. As a negative control, a guide targeting mouse PCSK9 (mus.PCSK9) was used.

In sum, the data indicate that base editing can inactivate cccDNA by introducing mutations abrogating HBV replication and silencing viral protein expression.

Example 9: Functional inactivation of Hepatitis B virus genomic reservoir by cytosine base editing

gRNAs targeting conserved regions of the HBV genome were designed, and screened (FIG. 26A and 26B). Interestingly, base editing with BE4 and the following gRNAs:

gRNA S1 (gRNA H) protospacer: GAAAGCCCAGGATGATGGGA

gRNA C2 (gRNA B) protospacer: CCATGCCCCAAAGCCACCCA

gRNA EMSbeam12 protospacer: GACTTCTCTCAATTTTCTAG

prevented HBV rebound in primary human hepatocytes (FIG. 27A-D). Multiplexing two gRNAs introducing stop codons with CBE reduced HBsAg, HBeAg, HBV DNA, and 3.5kb RNA in HepG2-NTCP (FIG. 28A-D) and in primary human hepatocytes. Base editing reduced HBs from naturally integrated HBB (FIG. 29A-C). Without intending to be bound by theory, a reduction in viral markers is likely driven by base editing of cccDNA without affecting cccDNA levels. As shown in FIG. 30A-C, base editing strongly reduced HBs from naturally integrated HBV. Combinatorial treatment of the base editing reagents with standard antiviral 3TC (lamivudine) resulted in higher detection of base editing efficiency.

Example 10: Cytosine base editing inhibited Hepatitis B Virus replication and reduced HBsAg expression in vitro and in vivo

Multiplexing the gRNAs of Example 9 with a BE4 base editor simultaneously reduced HBV viral parameters in HepG2-NTCP (FIG. 31A-E). Base editors function through cccDNA editing, without reducing cccDNA level. Base editing prevented viral rebound in Primary Human Hepatocytes (PHH) (Fig.32A-C).

Lipid nanoparticle-mediated delivery of BE4 mRNA and gRNAs S1/C2 lead to a sustained reduction of viral markers in an HBV minicircle murine model (Fig. 33A-C). An HBV minicircle mouse model supports durable production of HBV-like viral particles and HBV antigen expression in immunocompetent mice. Four weeks after a hydrodynamic

injection with a cccDNA-like minicircle plasmid, mice received one or two doses (2x) of the base editing reagents (mRNA & gRNA formulated into a lipid nanoparticle (LNP), at 2mg/kg). Entecavir (ETV) treated mice received antiviral at 0.03mg/kg orally for two weeks, then the treatment was discontinued.

5 In sum, multiplexing two gRNAs introducing stop codons in HBV genes HBs and Precore with CBE lead to a simultaneous reduction of HBsAg, HBeAg, HBV DNA, and 3.5kb RNA in HepG2-NTCP and human primary hepatocytes. Without intending to be bound by theory, a reduction in viral markers is likely to be driven by base editing of cccDNA, but not a reduction in cccDNA levels.

10 In vivo proof-of-concept in an HBV minicircle mouse model: IV injection(s) with LNP formulated with HBV targeting base editors at 2mg/kg lead to a sustained reduction in HBsAg, with several mice showing loss in HBsAg, as well as reduction of HBeAg and serum HBV DNA. Combined, the data indicated that base editing can inactivate cccDNA by introducing mutations abrogating HBV replication and silencing viral protein expression
15 (FIG. 29).

BE4 in combination with gRNA EMSbeam12 and MSPbeam37-PLC edited the naturally integrated HBV genome present in the Alexander hepatoma cell line, PLC/PRF/5 (FIG. 34A). This editing reduced HBsAg secretion (FIG. 34B).

The following methods were used in the above examples.

20 *Primary human hepatocytes system for studying HBV infections*

Primary hepatocytes (PHH) co-cultures infected with HBV were used to test antiviral efficacy of base editing systems. Primary hepatocyte co-culture (PHH) infected with HBV virus is a clinically relevant system for assessing anti-viral activity of base editing systems.

25 The base editing reagents (mRNA encoding a base editor(s) and synthetic gRNA) were transfected via lipofection twice over the course of two weeks. The first transfection was performed about 3, 4, or 7 days after infection (days post infection (dpi)). Not being bound by theory, this 3-to-7-day period ensured that the cccDNA was completely formed at the time of the transfection. The second transfection was at about 7, 10, 11, or 14 days after infection.
30 Extracellular parameters (e.g., HBsAg, HBeAg, and HBV DNA) were monitored over the course of the experiment (including the end of the experiment) and intracellular parameters (HBV DNA, viral RNA, and editing) were measured at the end of the experiment. Cells were lysed at the termination of the experiment (e.g., at 15, 17, 21, or 25 days).

Transfection details: RNA format, 800ng total RNA per well in a 24 well plate (600ng base editor-encoding mRNA + 200ng gRNA) transfected via lipofection (lipofectamin messengerMax, thermofisher Scientific) according to the vendor protocol.

5 **Example 11: Cytosine base editing of the hepatoma cell line HepG2.2.15 *in vitro* with and without pretreatment with lamivudine (LAM) reduced levels of HBsAg and HBeAg**

Experiments were undertaken to evaluate base editing *in vitro* of the hepatoma cell line HepG2.2.15 with and without pre-treatment with lamivudine (LAM). The HepG2.2.15 cells were base edited and polypeptide expression was subsequently evaluated according to the experimental scheme shown in **FIG. 35**. Cells were seeded at day 0 (D0). At day 1 (D1), LAM pretreatments were initiated for cells pre-treated with LAM. At day 4 (D4) cells were transferred to new wells with or without LAM. At day 5 (D5) cells were transfected with mRNA encoding a BE4 base editor and one or more of the guide polynucleotides gRNA12 (EMSbeam12 targeting the *HBs* gene), gRNA37 (MSPbeam37 targeting the *HBs* gene), and gRNA40 (MSPbeam40 targeting the gene encoding HBeAg) (**FIGs. 36A and 36B**). The growth medium was replaced at day 8 (D8). At day 11 (D11 and 6 days post-transduction) culture supernatants were collected for analysis using an enzyme-linked immunosorbent assay to measure levels of expression for HBsAg and HBeAg. Robust reduction in HBsAg expression was observed for cells edited using the guide gRNA37 or gRNA12 targeting the *HBs* gene, and robust reduction in HBeAg expression was observed for cells edited using the guide gRNA40 targeting the gene encoding HBeAg. Cells edited using both gRNA40 and gRNA37 showed reduced expression of both HBsAg and HBeAg.

Other Embodiments

25 From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

30 The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically

and individually indicated to be incorporated by reference. The present disclosure may be related to PCT/US20/32226, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

CLAIMS

What is claimed is:

- 5 1. A method of editing a nucleobase of a hepatitis B virus (HBV) genome, the method comprising contacting the HBV genome with one or more guide RNAs and a base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase, wherein said guide RNA targets said base editor to effect an alteration of the nucleobase of the HBV genome, and wherein said one or more guide RNAs comprise a

10 nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID

15 NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), GAAAGCCCAAGAUGAUGGGA (SEQ ID NO: 10834), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

20
2. A method of editing a nucleobase of a hepatitis B virus (HBV) genome, the method comprising:

(i) contacting a cell comprising an HBV genome with one or more guide RNAs and a base editor comprising a polynucleotide programmable DNA binding domain and an

25 adenosine deaminase or cytidine deaminase, wherein said guide RNA targets said base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome; and

(ii) contacting the cell with an antiretroviral drug, wherein contacting the cell with the antiretroviral drug is associated with an increase in base editing efficiency relative to a

30 reference cell not contacted with the antiretroviral drug.
3. A method reducing HBsAg produced from integrated HBV DNA, the method comprising contacting a cell comprising an HBV genome with one or more guide RNAs and a base editor comprising a polynucleotide programmable DNA binding domain and an

adenosine deaminase or cytidine deaminase, wherein said guide RNA targets said base editor to effect an alteration of the nucleobase of the HBV genome, thereby reducing HBsAg production from an integrated HBV DNA.

- 5 4. The method of claim 2, wherein the antiretroviral drug is selected from the group consisting of Lamivudine, Entecavir, Tenofovir, Interferon, and Peg-Interferon.
5. The method of any one of claims 2-4, wherein step (i) precedes step (ii), or step (ii) precedes step (i).
- 10 6. The method of any one of claims 2-5, wherein the cell is first contacted with the antiretroviral drug about or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 days prior to step (i).
- 15 7. The method of any one of claims claim 2-6, wherein the cell has not been previously contacted with the antiretroviral drug.
8. The method of any one of claims 2-7, wherein the cell is in a subject.
- 20 9. The method of any one of claims 1-8, wherein the HBV genome is contacted with two or more guide RNAs simultaneously.
10. The method of claim any one of claims 2-9, wherein the guide RNA is selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405),
 25 UCGCUGGAUGUGUCUGCGGC (SEQ ID NO:406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410),
 UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCAAGGUCUUACAUAAGAG
 30 (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415),
 ACACAGAAAGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

11. The method of any one of claims 2-10, wherein the one or more guide RNAs comprise a spacer sequence selected from a sequence listed in Table 2A, Table 2B, Table 2C, and/or SEQ ID NOs: 3105-5485 and 8220-10830.

5 12. The method of any one of claims 2-11, wherein the one or more guide RNAs comprise spacers comprising the nucleotide sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO: 657), 5'-GAAAGCCCAAGATGATGGGA-3', and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662).

10 13. The method of any one of claims 2-12, wherein the one or more guide RNAs comprise the nucleotide sequences
mG*mA*mA*AGCCCAGGAUGAUGGGAGUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU
AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO:
424) and

15 mC*mC*mA*UGCCCCAAAGCCACCCAGUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCU
AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO:
422).

14. The method of any one of claims 1-13, wherein the one or more guide RNAs comprise
20 the following scaffold sequence:
gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases
with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl
3'-phosphorothioate (MS) modification.

25 15. The method of any one of claims 2-12, wherein the one or more guide RNAs
comprise the following nucleotide sequences
gsasasAGCCCAGGAUGAUGGGAgUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGU
ccGUUAucAAcuugaaaaagugGcaccgagucggugcusususu (SEQ ID NO: 424) and
30 cscsasUGCCCCAAAGCCACCCAgUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGU
ccGUUAucAAcuugaaaaagugGcaccgagucggugcusususu (SEQ ID NO: 422), where
a, c, u, or g represent an A, C, U, or G nucleotide containing a 2'-O-methyl (M) modification,

respectively; and as, cs, us, or gs represent an A, C, U, or G nucleotide containing a 2'-O-methyl 3'-phosphorothioate (MS) modification, respectively.

16. The method of any one of claims 2-15, wherein the cell is contacted at a first and a second time point with the one or more guide RNAs and the base editor.

17. The method of claim 16, wherein at least one week separates the first and second times the cell is contacted.

18. The method of claim 16, wherein at least two weeks separates the first and second times the cell is contacted.

19. The method of any one of claims 2-18, wherein base editing efficiency is improved by at least about 5%, 10%, 15%, or 20%.

15

20. A method of editing a nucleobase of a hepatitis B virus (HBV) genome, the method comprising contacting a cell comprising an HBV genome with an mRNA encoding a base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase or cytidine deaminase, and one or more guide RNAs comprising the sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO: 657) and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662), wherein said one or more guide RNAs target said base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome.

21. The method of claim 20, further comprising contacting the cell with lamivudine, wherein contacting the cell with lamivudine is associated with an increase in base editing efficiency relative to a reference cell not contacted with the antiretroviral drug.

22. The method of claim 20 or claim 21, wherein the guide RNAs comprise the following scaffold sequence:

gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases

with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl 3'-phosphorothioate (MS) modification.

23. The method of claim 20 or claim 21, wherein the guide RNAs comprise the following scaffold sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422), where mA*, mC*, mG*, and mU*
represent bases with a 2'-O-methyl 3'-phosphorothioate modification.

24. The method of any one of claims 20-23, wherein the cell is contacted at a first and a second time point with the one or more guide RNAs and the base editor.

25. The method of claim 24, wherein at least one week separates the first and second times the cell is contacted.

26. The method of any one of claims 20-25, wherein the base editor is a BE4.

27. The method of any one of claims 1-26, wherein the nucleobase is in a polynucleotide encoding an HBV protein.

28. The method of any one of claims 1-26, wherein the contacting is in a eukaryotic cell, a mammalian cell, or a human cell.

29. The method of any one of claims 2 to 28, wherein the cell is *in vivo* or *ex vivo*.

30. The method of any one of claims 1 to 29, wherein the adenosine deaminase converts a target A•T to G•C in the hepatitis B virus (HBV) genome.

31. The method of any one of claims 1-29, wherein the nucleobase is in a polynucleotide sequence encoding an HBV protein selected from HBV core protein (Core), HBV polymerase (Pol), HBV surface protein, or HBV protein X.

32. The method of any one of claims 1-31, wherein alteration of the nucleobase in the polynucleotide encoding the HBV protein results in a missense mutation.

33. The method of any one of claims 1-32, wherein the nucleobase is associated with a transcription site.

5 34. The method of claim 33, wherein the transcription site is selected from one or more of Enhancer II box A, Enhancer I, and the HBX promoter.

35. The method of any one of claims 1-34, wherein alteration of the nucleobase is associated with a reduction in transcription of a polynucleotide sequence encoding an HBV protein.
10

36. The method of claim 35, wherein the HBV protein is HBV polymerase (Pol), and/or HBV protein X.

15 37. The method of claim 32, wherein the missense mutation is in an HBV pol gene.

38. The method of claim 37, wherein the missense mutation results in an E24G, L25F, P26F, R27C, V48A, V48I, S382F, V378I, V378A, V379I, V379A, L377F, D380G, D380N, F381P, R376G, A422T, F423P, A432V, M433V, P434S, D540G, A688V, D689G, A717T, E718K, P713S, P713L, or L719P in an HBV polymerase protein encoded by the HBV pol gene.
20

39. The method of claim 32, wherein the missense mutation is in an HBV core gene.

25 40. The method of claim 39, wherein the missense mutation results in a T160A, T160A, P161F, S162L, C183R, or *184Q in an HBV core protein encoded by the HBV core gene.

41. The method of claim 32, wherein the missense mutation is in an HBV X gene.

30 42. The method of claim 41, wherein the missense mutation results in a H86R, W120R, E122K, E121K, or L141P in an HBV X protein encoded by the HBV X gene.

43. The method of claim 32, wherein the missense mutation is in an HBV S gene.

44. The method of claim 43, wherein the missense mutation results in a S38F, L39F, W35R, W36R, T37I, T37A, R78Q, S34L, F80P, or D33G in an HBV S protein encoded by the HBV S gene.

5 45. The method of any one of claims 1-44, wherein the polynucleotide programmable DNA binding domain comprises a Cas12 polypeptide.

46. The method of any one of claims 1-45, wherein the polynucleotide programmable DNA binding domain comprises Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h,
10 or Cas12i.

47. The method of any one of claims 1-46 wherein the polynucleotide programmable DNA binding domain comprises a sequence with at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13
15 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b.

48. The method of claim 47, wherein the polynucleotide programmable DNA binding domain comprises a sequence with at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b (bhCas12b).
20

49. The method of any one of claims 1-48, wherein the polynucleotide programmable DNA binding domain comprises a nuclease inactive or nickase variant.

50. The method of claim 49, wherein the nuclease inactive or nickase variant comprises a
25 nuclease inactivated bhCas12b which comprises amino acid substitutions D952A, S893R, K846R, and E837G, or corresponding amino acid substitutions thereof.

51. The method of any one of claims 1-50, wherein the adenosine deaminase domain is capable of deaminating adenine in deoxyribonucleic acid (DNA).
30

52. The method of any one of claims 1-51, wherein the adenosine deaminase is a TadA deaminase.

53. The method of claim 52, wherein the TadA deaminase is TadA*7.10, 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
5 TadA*8.24.

54. The method of claim 53, wherein the TadA deaminase is TadA*8.13.

55. The method of any one of claims 1-54, wherein the one or more guide RNAs
10 comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), wherein the crRNA comprises a nucleic acid sequence complementary to an HBV nucleic acid sequence.

56. The method of any one of claims 1-55, wherein the base editor is in complex with a
15 single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an HBV nucleic acid sequence.

57. The method of any one of claims 1-56, comprising editing two or more nucleobases.

20 58. The method of any one of claims 1-57, comprising contacting the HBV genome with two or more guide RNAs that target two or more HBV nucleic acid sequences.

59. A method of treating hepatitis B virus (HBV) infection in a subject comprising administering to a subject in need thereof a fusion protein, protein complex, or
25 polynucleotide encoding said fusion protein or protein complex, said fusion protein or protein complex comprising a polynucleotide programmable DNA binding domain and a base editor domain that is an adenosine deaminase domain, and one or more guide RNAs that target the base editor domain to effect an A•T to G•C alteration of a nucleobase of the HBV genome, wherein said one or more guide RNAs comprise a nucleotide sequence, from 5' to 3', or a 1,
30 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO:406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409),

GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

60. A method of treating a hepatitis B virus (HBV) infection in a subject, comprising administering to a subject in need thereof one or more polynucleotides encoding a polynucleotide programmable DNA binding domain and a base editor domain that is an adenosine deaminase domain, and one or more guide RNAs that target the base editor domain to effect an A•T to G•C, alteration of a nucleobase of the HBV genome, wherein said one or more guide RNAs comprise a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

61. A method of treating a hepatitis B virus (HBV) infection in a subject, comprising administering to a subject in need thereof with one or more guide RNAs and a base editor comprising a polynucleotide programmable DNA binding domain and an adenosine deaminase or cytidine deaminase, wherein said one or more guide RNAs comprise the sequences GAAAGCCCAGGAUGAUGGGA (SEQ ID NO: 657) and CCAUGCCCCAAAGCCACCCA (SEQ ID NO: 662) that target said base editor to effect an alteration of the nucleobase of the HBV genome, thereby editing a nucleobase in the HBV genome.

62. The method of any one of claims 59-61, wherein the subject is a mammal or a human.

63. The method of any one of claims 59-62, comprising delivering the fusion protein, the polynucleotide encoding said fusion protein, or one or more polynucleotides encoding the polynucleotide programmable DNA binding domain and the base editor domain, and said one or more guide RNAs to a cell of the subject.

5

64. The method of claim 63, wherein the cell is a hepatocyte.

65. The method of any one of claims 59-64, wherein the nucleobase is in a polynucleotide encoding an HBV protein.

10

66. The method of claim 65, wherein the HBV protein is HBV core protein (Core), HBV polymerase (Pol), HBV surface protein, or HBV protein X.

15

67. The method of claim 65 or claim 66, wherein alteration of the nucleobase in the polynucleotide encoding the HBV protein results in a missense mutation.

68. The method of any one of claims 59-67, wherein the nucleobase is associated with a transcription site.

20

69. The method of claim 68, wherein the transcription site is selected from one or more of Enhancer II box A, Enhancer I, and the HBX promoter.

25

70. The method of any one of claims 59-69, wherein alteration of the nucleobase is associated with a reduction in transcription of a polynucleotide sequence encoding an HBV protein.

71. The method of claim 70, wherein the HBV protein is HBV polymerase (Pol), and/or HBV protein X.

30

72. The method of claim 67, wherein the missense mutation is in an HBV pol gene.

73. The method of claim 72, wherein the missense mutation results in an E24G, L25F, P26F, R27C, V48A, V48I, S382F, V378I, V378A, V379I, V379A, L377F, D380G, D380N, F381P, R376G, A422T, F423P, A432V, M433V, P434S, D540G, A688V, D689G, A717T,

E718K, P713S, P713L, or L719P in an HBV polymerase protein encoded by the HBV pol gene.

74. The method of claim 67, wherein the missense mutation is in an HBV core gene.

75. The method of claim 74, wherein the missense mutation results in a T160A, T160A, P161F, S162L, C183R, or *184Q in an HBV core protein encoded by the HBV core gene.

76. The method of claim 67, wherein the missense mutation is in an HBV X gene.

77. The method of claim 76, wherein the missense mutation results in a H86R, W120R, E122K, E121K, or L141P in an HBV X protein encoded by the HBV X gene.

78. The method of claim 67, wherein the missense mutation is in an HBV S gene.

79. The method of claim 78, wherein the missense mutation results in a S38F, L39F, W35R, W36R, T37I, T37A, R78Q, S34L, F80P, or D33G in an HBV S protein encoded by the HBV S gene.

80. The method of any one of claims 59-79, wherein the polynucleotide programmable DNA binding domain comprises a Cas12 polypeptide.

81. The method of any one of claims 59-80, wherein the polynucleotide programmable DNA binding domain comprises Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, or Cas12i.

82. The method of any one of claims 59-81, wherein the polynucleotide programmable DNA binding domain comprises a sequence with at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b.

83. The method of claim 82, wherein the polynucleotide programmable DNA binding domain comprises a sequence with at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b (bhCas12b).

84. The method of any one of claims 59-83, wherein the polynucleotide programmable DNA binding domain is a nuclease inactive or nickase variant.

5 85. The method of claim 84, wherein the nuclease inactive or nickase variant is a nuclease inactivated bhCas12b which comprises amino acid substitutions D952A, S893R, K846R, and E837G, or corresponding amino acid substitutions thereof.

10 86. The method of any one of claims 59-85, wherein the adenosine deaminase domain is capable of deaminating adenine in deoxyribonucleic acid (DNA).

87. The method of any one of claims 59-86, wherein the adenosine deaminase is a TadA deaminase.

15 88. The method of claim 87, wherein the TadA deaminase is TadA*7.10, 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.

20 89. The method of claim 88, wherein the TadA deaminase is TadA*8.13.

90. The method of any one of claims 59-89, wherein the one or more guide RNAs comprises a CRISPR RNA (crRNA) and a trans-encoded small RNA (tracrRNA), wherein
25 the crRNA comprises a nucleic acid sequence complementary to an HBV nucleic acid sequence.

91. The method of any one of claims 59-90, wherein the base editor is in complex with a single guide RNA (sgRNA) comprising a nucleic acid sequence complementary to an HBV
30 nucleic acid sequence.

92. The method of any one of claims 59-91, comprising editing two or more nucleobases.

93. The method of any one of claims 59-92, comprising contacting the HBV genome with two or more guide RNAs that target two or more HBV nucleic acid sequences.

94. The method of claim 92 or claim 93, comprising two or more guide RNAs that target three, four, or five nucleobases of the HBV genome.

95. The method of any one of claims 59-93, wherein the guide RNAs comprise the following scaffold sequence:

gUUUUAGagcuagaaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl 3'-phosphorothioate (MS) modification.

96. The method of any one of claims 59-93, wherein the guide RNAs comprise the following scaffold sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422), where mA*, mC*, mG*, and mU* represent bases with a 2'-O-methyl 3'-phosphorothioate modification.

97. The method of claim 63 or claim 64, wherein the cell is contacted at a first and a second time point with the one or more guide RNAs and the base editor.

98. The method of claim 97, wherein at least one week separates the first and second times the cell is contacted.

99. The method of claim 97, wherein at least two weeks separates the first and second times the cell is contacted.

100. A composition comprising a base editor(s) bound to a guide RNA(s), wherein said guide RNA(s) comprises a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of
ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCTCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA

(SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409),
 GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID
 NO: 411), GAGGACAAACGGGCAACUA (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC
 (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414),
 5 CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID
 NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

101. The composition of claim 100, wherein the guide RNAs comprise the following
 scaffold sequence:

10 gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
 caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases
 with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl
 3'-phosphorothioate (MS) modification.

15 102. The composition of claim 100, wherein the guide RNAs comprise the following scaffold
 sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
 CACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422), where mA*, mC*, mG*, and mU*
 represent bases with a 2'-O-methyl 3'-phosphorothioate modification.

20

103. The composition of any one of claims 100-102, wherein the base editor comprises an
 adenosine deaminase.

104. The composition of claim 103, wherein the adenosine deaminase is capable of
 25 deaminating adenine in deoxyribonucleic acid (DNA).

105. The composition of claim 104, wherein the adenosine deaminase is a TadA
 deaminase.

30 106. The composition of claim 105, wherein the TadA deaminase is TadA*7.10,
 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.

107. The composition of claim 106, wherein the TadA deaminase is TadA*8.13.

5

108. The composition of any one of claims 100-106, wherein the base editor

(i) comprises a nuclease inactive bhCas12b; or

(ii) comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to

10 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAH
AEIMALRQGGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD
VLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESA
TPESSGAPKKKRKVGIIHGVPAAATRSEFILKIEPNNEEVKKGLWKTHEVLNHGIAYYMNILKLI
RQEAIYEHHEQDPKNPKKVSKEIQAEIQLWDFVLKMQKCNSTHEVDKDEVFNILRELYEELV
15 PSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEEKKWEED
KKKDPLAKILGKLAEYGLIPLFIPTDSNEPIVKEIKWMEKSRNQSVRRDKDMFIQALERF
LSWESWNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLS
KRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNH
PEYPYLYATFCEIDKKKKDAKQOATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEK
20 LKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKF
PLKGTLLGARVQFDRDHLRRYPHKVESGNVGRIYFNMVTNIEPTESPVSKSLKIHRDDFPKV
VNFKPKELTEWIKDSKGKLLKSGIESLEIGLRVMSIALGQRQAAAASIFEVVDQKPDIEGKL
FFPIKGTELYAVHRASFNIKLPGETLVKSREVLRLKAREDNLKLMNQKLNFLRNVLHFQQFED
ITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFQKQLHKLRLVEIGKEVK
25 HWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFADQLNHLN
ALKEDRLKKMANTIIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYKERSRFENSRL
MKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCRVVTKEKLQDNRFKLN
QREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFY
KVYCKAYQVDGQTVYIPESKDQKQKIIIEEFGEYGFILKDGVEWVNAGKLIKKGSSKQSSS
30 ELVDS DILKDSFDLASELKGEKLMLYRDP SGNVFPSDKWMAAGVFFGKLERILISKLTNQYS
ISTIEDDSSKQSMKRPAATKKAGQAKKKK (SEQ ID NO: 418).

109. The composition of any one of claims 100-108, further comprising a lipid.

110. The composition of claim 109, wherein the lipid is a cationic lipid.

111. The composition of any one of claims 100-110, further comprising a pharmaceutically acceptable excipient.

5

112. A pharmaceutical composition for the treatment of HBV infection comprising:

- (i) a base editor, or a nucleic acid encoding the base editor, and one or more guide RNAs (gRNA) comprising a nucleic acid sequence complementary to an HBV gene in a pharmaceutically acceptable excipient, wherein said one or more guide RNAs comprise a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO:406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408),
- 15 GCUGACGCAACCCCCACUGG (SEQ ID NO: 409), GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCCAAGGUCUUACAUAAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and
- 20 CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

113. The pharmaceutical composition of claim 112, wherein the base editor

- (i) comprises a nuclease inactive bhCas12b; or
- (ii) comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 96%,
- 25 97%, 98%, 99%, or 100% identical to
- MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESS
GAPKKRKRVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI
30 YEHHEQDPKNPKKVSKAEIQALWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEEKKKWEEDKKKDP
LAKILGKLAEYGLIPLFIPYTDSEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
WNLKVKEEYKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR

GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
 LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKL
 TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
 LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
 5 KELTEWIKDSKGKKLKSGIESLEIGLRVMSIALGQRQAAAASIFEVVDQKPDIEGKLEFFPIK
 GTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITERE
 KRVTKWISRQENS DVPLVYQDELIQIRELMYKPYKDWVAFLEKQLHKLRLVEIGKEVKHWRKS
 LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFADQLNHLNALKED
 RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYPYKERSRFENSRLMKWSR
 10 REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCVVTKEKLQDNRFKLNQREGR
 LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLOKRFWTRTHGFYKVYCK
 AYQVDGQTVYIPESKDQKQKIIIEFGEGYFILKDGVEWVNAGKLKIKKGSSKQSSSELVDS
 DILKDSFDLASELKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQYSISTIE
 DDSSKQSMKRPAATKKAGQAKKKK (SEQ ID NO: 418).

15

114. The pharmaceutical composition of claim 112 or claim 113, wherein the base editor comprises a bhCas12b, or a bhCas12b variant comprising amino acid substitutions D952A, S893R, K846R, and E837G.

20

115. The pharmaceutical composition of any one of claims 112-114, wherein the one or more gRNAs and the base editor are formulated together.

116. The pharmaceutical composition of any one of claims 112-115, wherein the one or more gRNAs and the base editor are formulated separately.

25

117. The pharmaceutical composition of any one of claims 112-116, further comprising a vector suitable for expression in a mammalian cell, wherein the vector comprises a polynucleotide encoding the base editor.

30

118. The pharmaceutical composition of claim 117, wherein the vector is a viral vector.

119. The pharmaceutical composition of claim 118, wherein the viral vector is a retroviral vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated viral vector (AAV).

120. The pharmaceutical composition of any one of claims 112-119, further comprising an ribonucleoparticle suitable for expression in a mammalian cell.

121. The composition of any one of claims 112-120, wherein the guide RNAs comprise the following scaffold sequence:

gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases
with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl
3'-phosphorothioate (MS) modification.

122. The composition of any one of claims 112-120, wherein the guide RNAs comprise the following scaffold sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422), where mA*, mC*, mG*, and mU*
represent bases with a 2'-O-methyl 3'-phosphorothioate modification.

123. A method of treating HBV infection, the method comprising administering to a subject in need thereof the composition of any one of claims 100-111.

124. A method of treating an HBV infection, the method comprising administering to a subject in need thereof the pharmaceutical composition of any one of claims 112-123.

125. The method of any of the above claims, wherein the HBV is of genotype A, or of genotype D.

126. The method of any of the above claims, wherein the method reduces levels of a marker selected from the group consisting of Hepatitis B Surface Antigen, Hepatitis B E Antigen, Hepatitis B Virus total DNA, Hepatitis B 3.5 kb RNA, and Hepatitis B covalently closed circular DNA.

127. The method of any of the above claims, wherein the method edits the Hepatitis B S antigen site and the PreCore site.

128. The method of any of the above claims, wherein the editing efficiency is about 30% at the S antigen site and about 60% at the PreCore site.

5 129. Use of the composition of any one of claims 100-111 in the treatment of HBV infection in a subject.

130. Use of the pharmaceutical composition of any one of claims 112-123 in the treatment of HBV infection in a subject.

10

131. The use of claim 129 or 130, wherein the subject is a mammal.

132. The use of any one of claims 129-131, wherein the subject is a human.

15 133. A guide RNA comprise a nucleotide sequence, from 5' to 3', or a 1, 2, 3, 4, or 5 nucleotide 5' and/or 3' truncation fragment thereof, selected from one or more of ACAAGAAUCCUCACAAUACC (SEQ ID NO: 405), UCGCUGGAUGUGUCUGCGGC (SEQ ID NO: 406), CACCUGUAUUCCCAUCCCAU (SEQ ID NO: 407), GGGGCCAAGUCUGUACAGCA (SEQ ID NO: 408), GCUGACGCAACCCCCACUGG (SEQ ID NO: 409),
 20 GGAGCUACUGUGGAGUUACU (SEQ ID NO: 410), UUCUUCUAGGGGACCUGCCU (SEQ ID NO: 411), GAGGACAAACGGGCAACAU (SEQ ID NO: 412), UUGUCAACAAGAAAAACCCC (SEQ ID NO: 413), CCAAGGUCUUACAUAGAG (SEQ ID NO: 414), CCGGGCAACGGGGUAAAGGU (SEQ ID NO: 415), ACACAGAAAGGCCUUGUAAG (SEQ ID NO: 416), and CACGCACGCGCUGAUGGCCC (SEQ ID NO: 417).

25

134. The guide RNA of claim 133, wherein the guide RNA comprises the following scaffold sequence:

gUUUUAGagcuagaaauagcaaGUUaAaAuAaggcuaGUccGUUAucAAcuugaaaaagugG
 caccgagucggugcusususu (SEQ ID NO: 317), wherein a, c, u, or g represents bases

30 with a 2'-O-methyl (M) modification, and as, cs, us, or gs represents bases with a 2'-O-methyl 3'-phosphorothioate (MS) modification.

ELECTRONIC DEPOSIT DATE: September 27, 2022

135. The guide RNA of claim 133, wherein the guide RNA comprises the following scaffold sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUmU*mU*mU* (SEQ ID NO: 422), where mA*, mC*, mG*, and mU*

5 represent bases with a 2'-O-methyl 3'-phosphorothioate modification.

136. A pharmaceutical composition comprising (i) a nucleic acid encoding a base editor; and (ii) the guide RNA of any one of claims 133-135.

10 137. The pharmaceutical composition of claim 136, further comprising a lipid.

138. The pharmaceutical composition of claim 136 or claim 137, wherein the nucleic acid encoding the base editor is an mRNA.

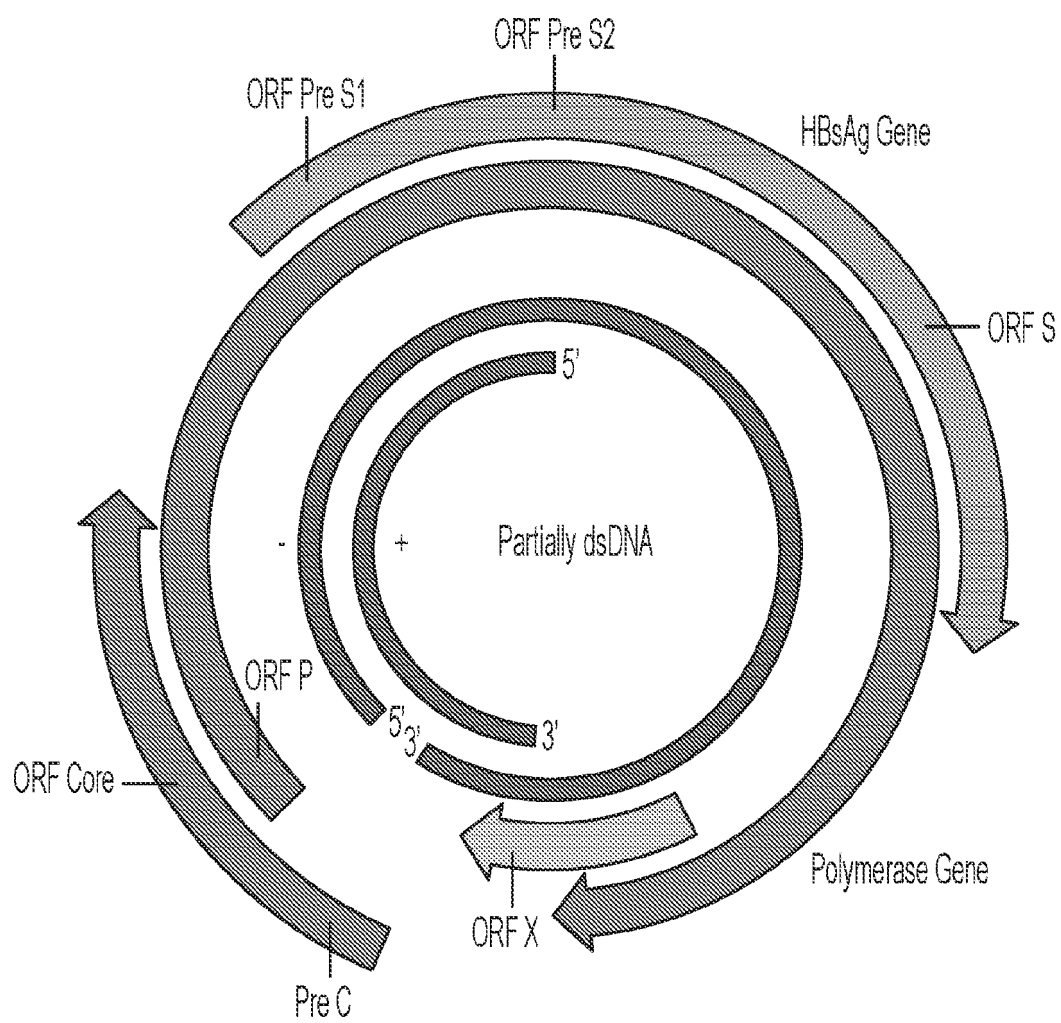


FIG. 1

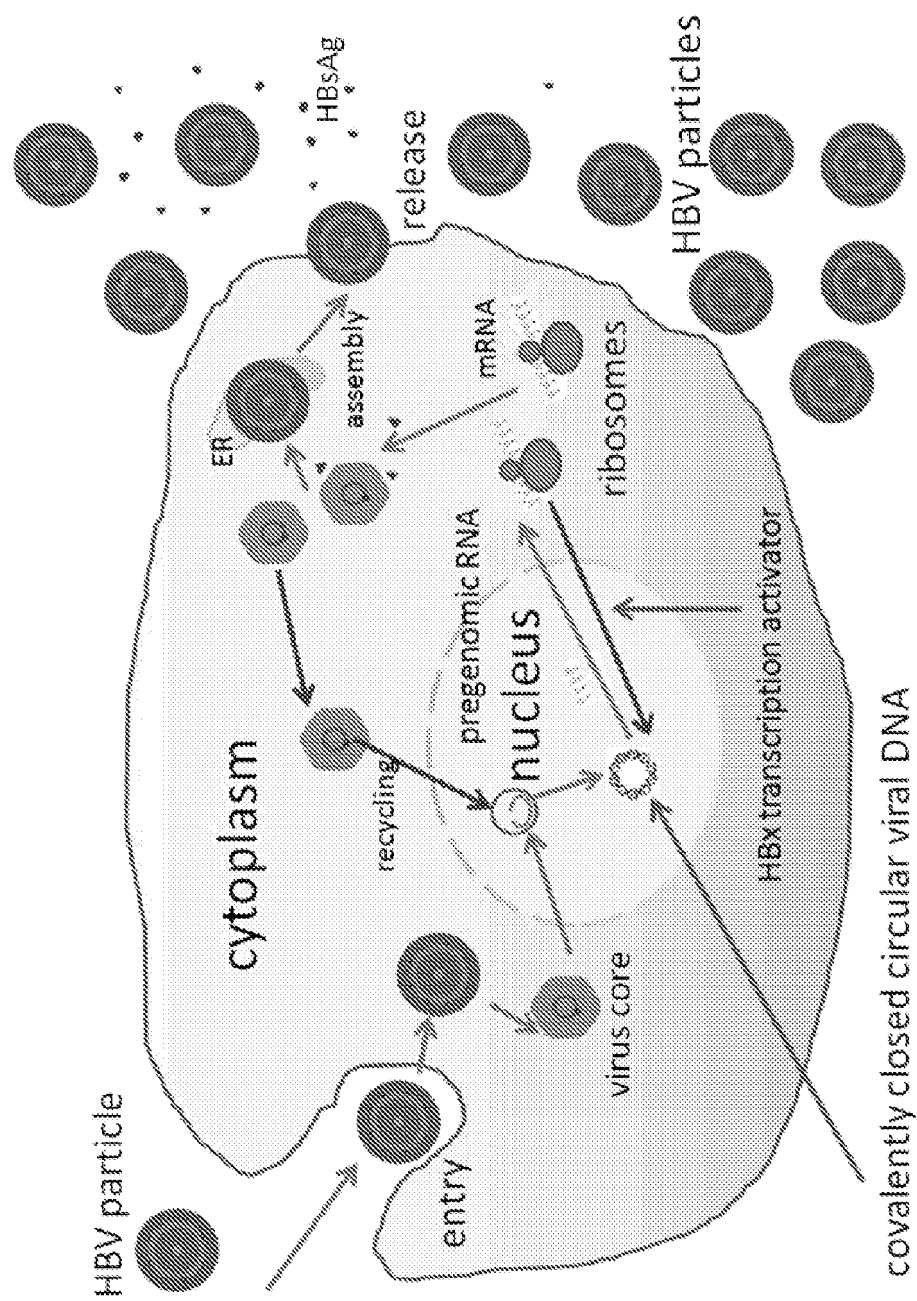
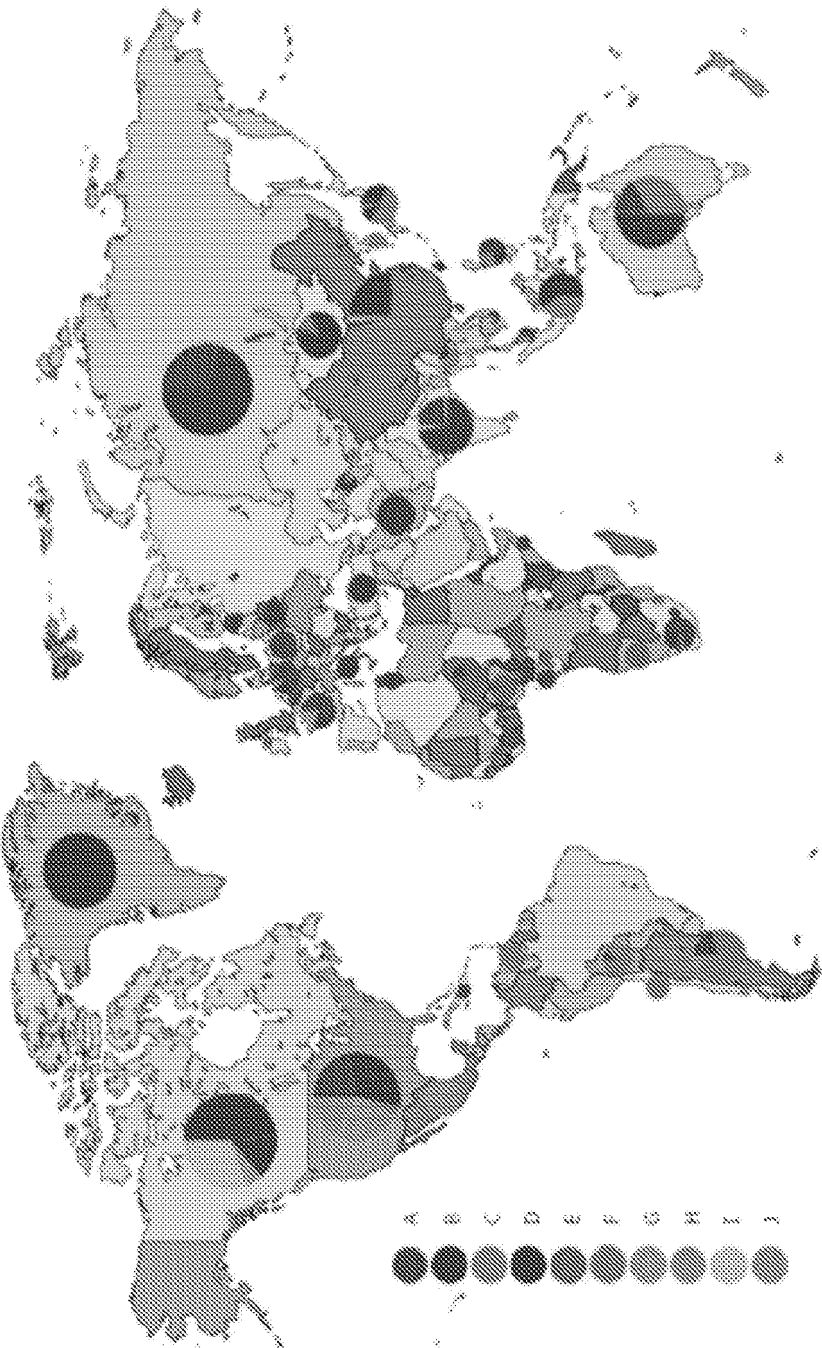


FIG. 2



World Gastroenterology Organisation Global Guideline, 2015)

FIG. 3A

Base Editing as a Strategy to Treat Chronic HBV

- Allows introducing stop codons in viral genes without generating double strand breaks
 - Although Cas9 nucleases effectively mediate HBV cccDNA disruption, the risk of cutting integrated HBV genomes has raised serious concerns
- Employs deaminase, which is known to be a natural HBV antiviral restriction factor

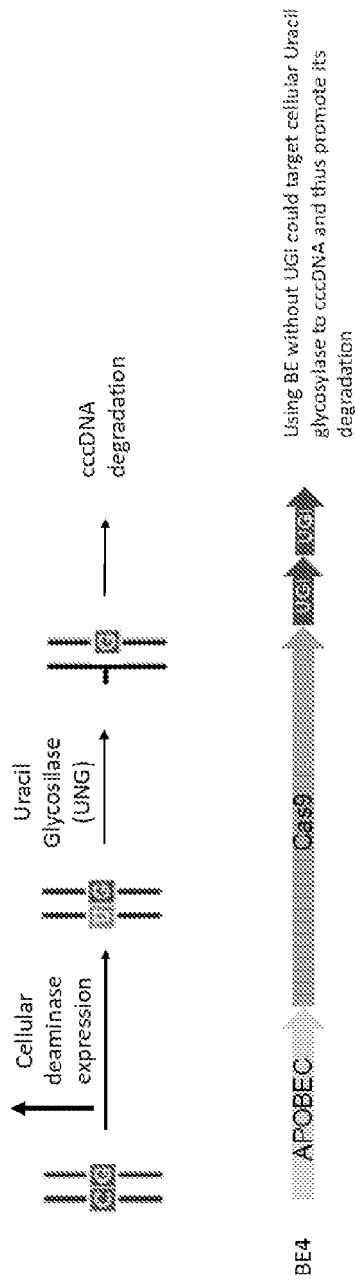


FIG. 3B

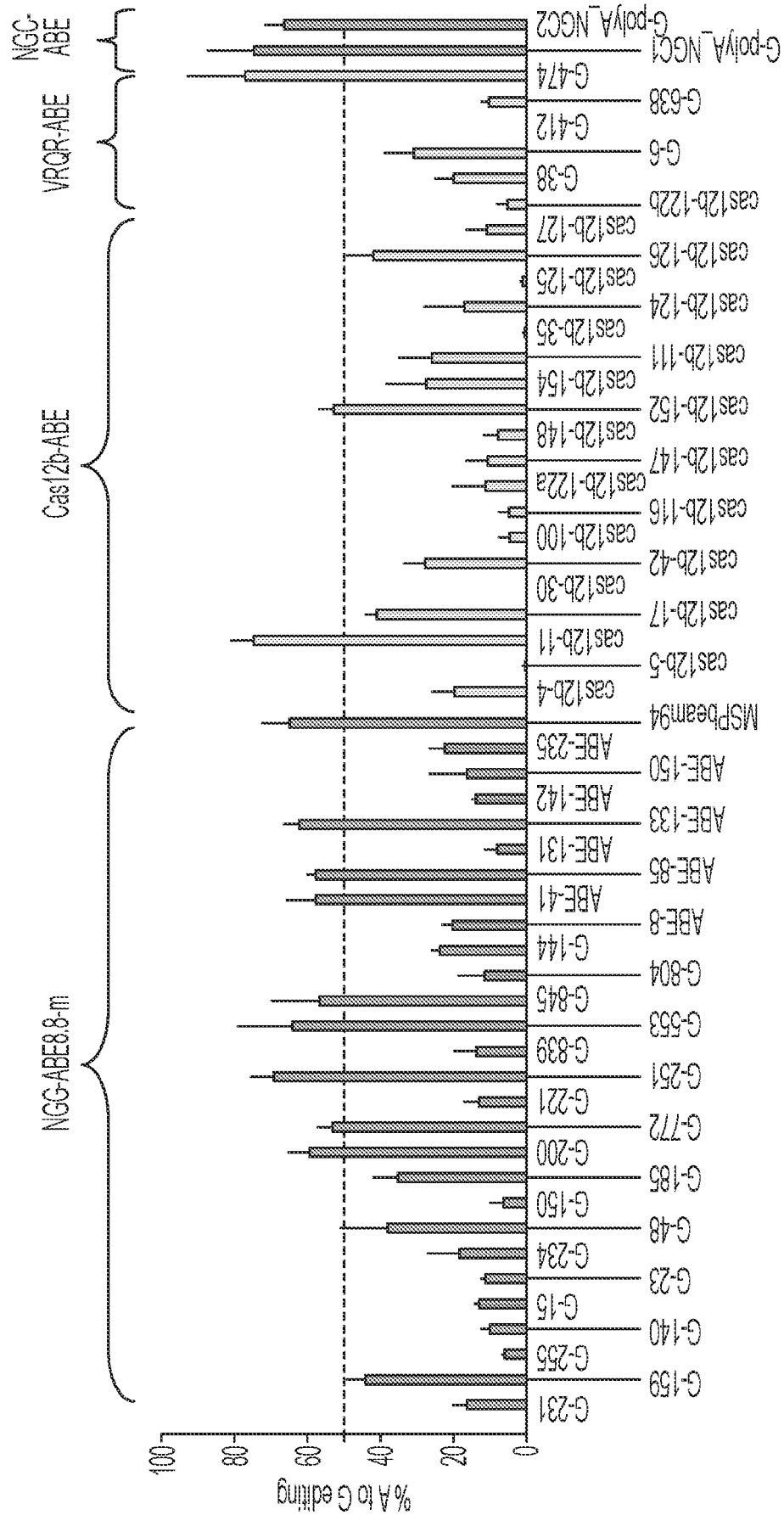


FIG. 4

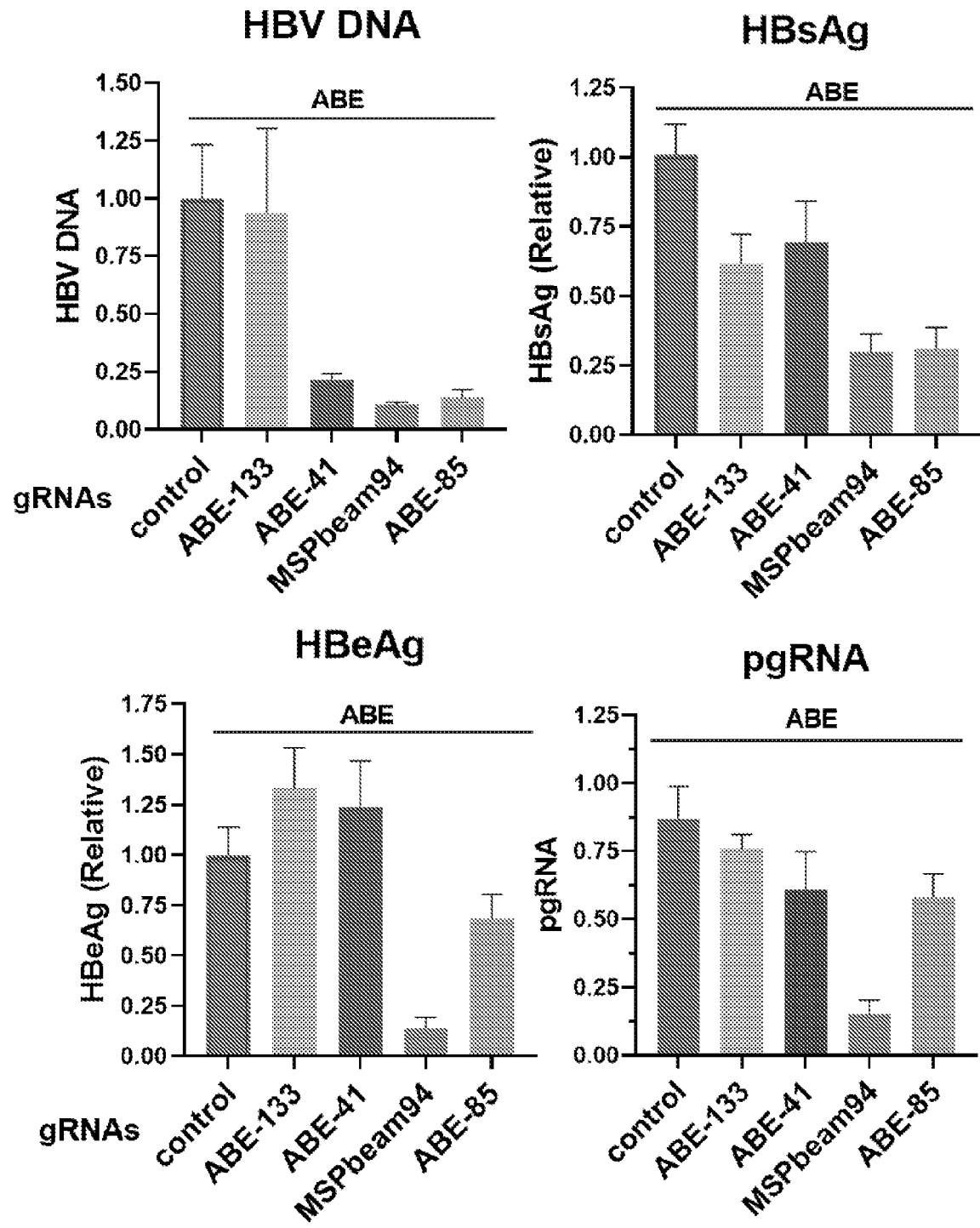


FIG. 5A

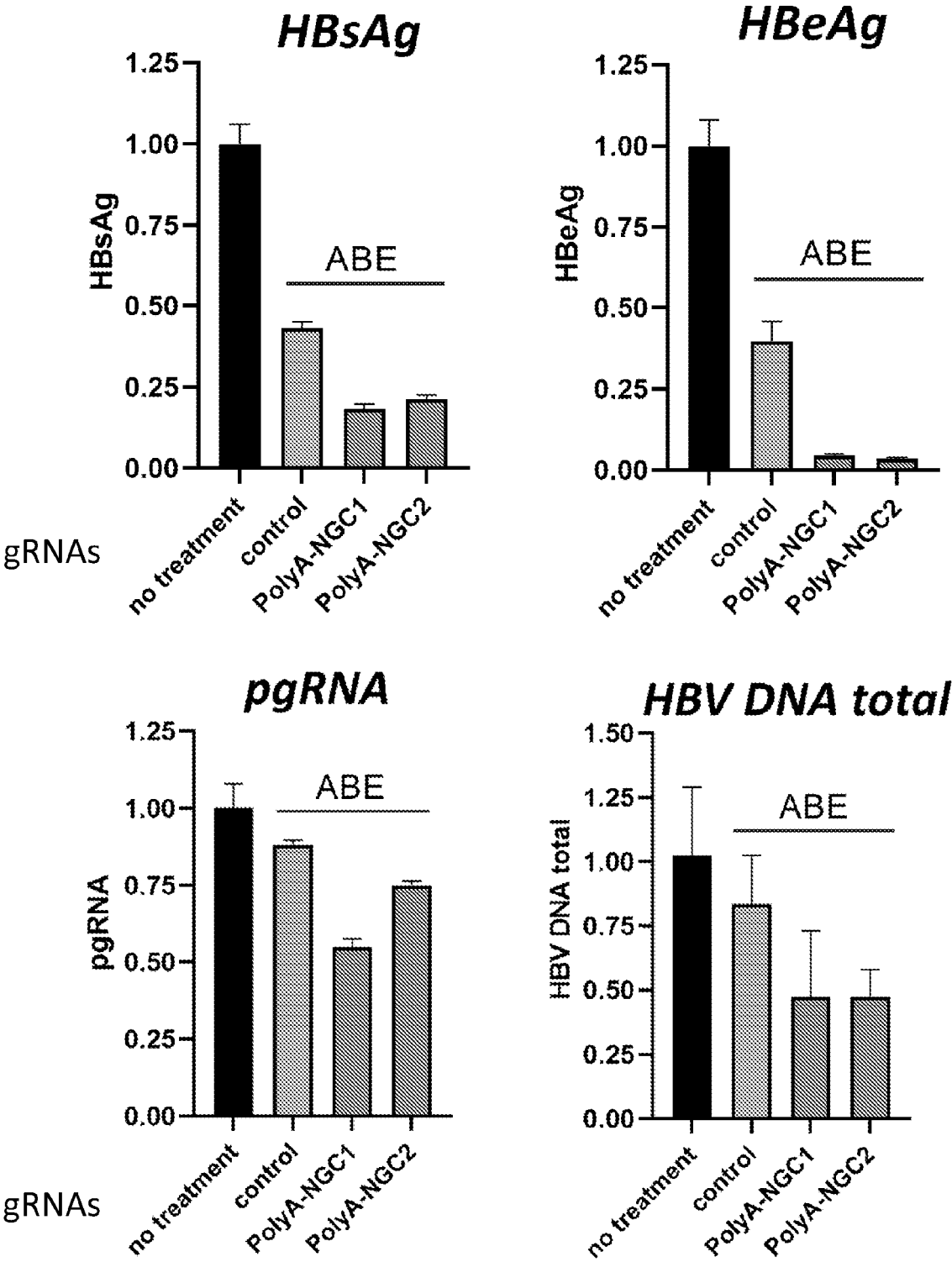


FIG. 5B

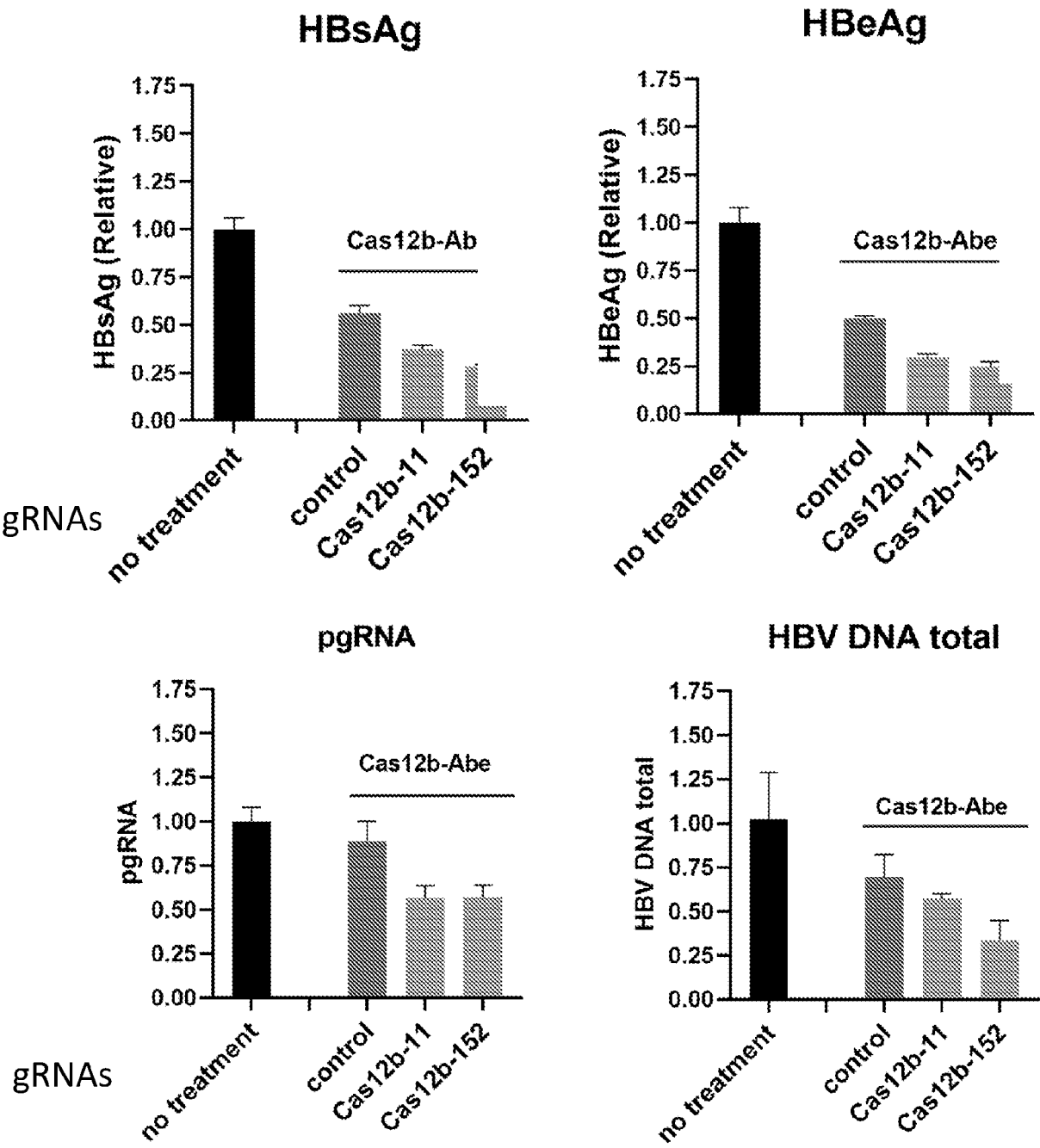


FIG. 5C

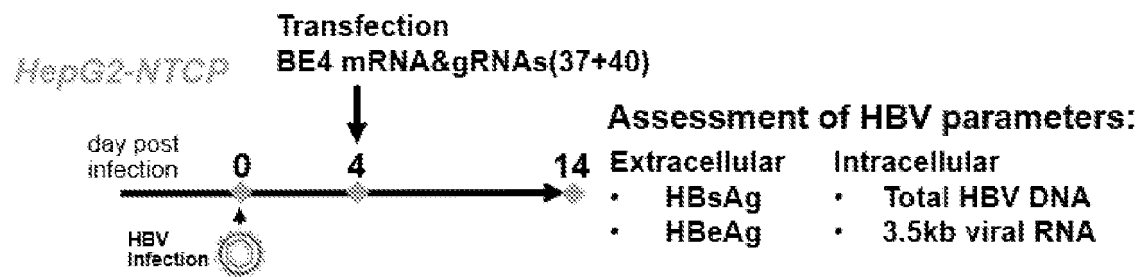


FIG. 6

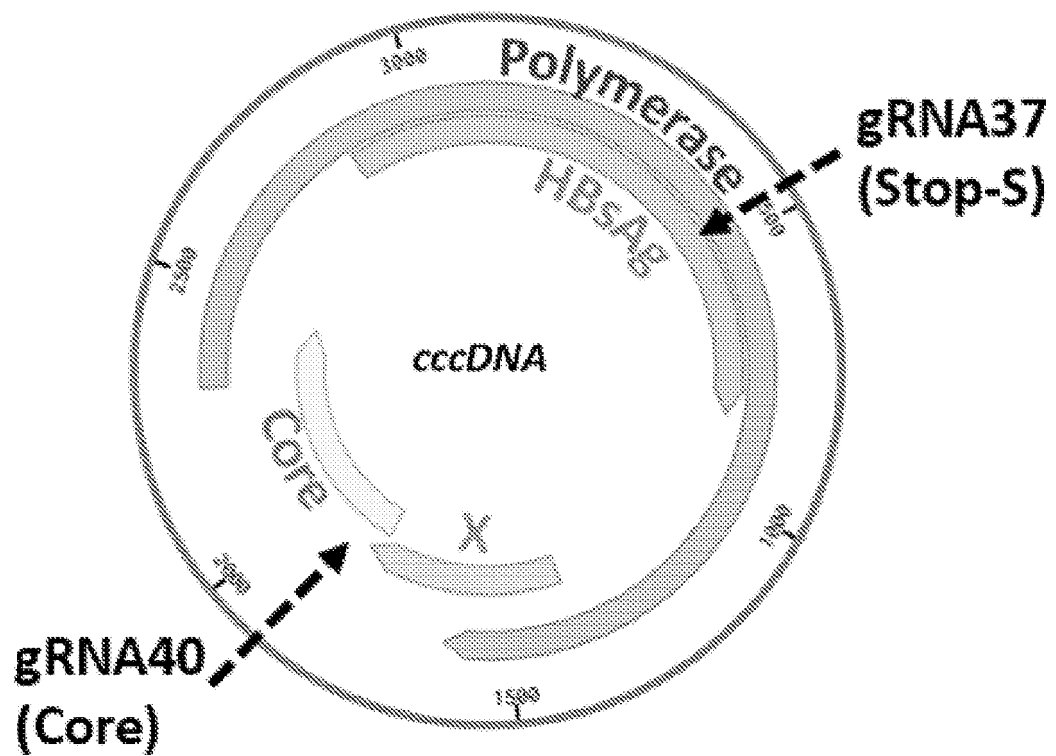


FIG. 7

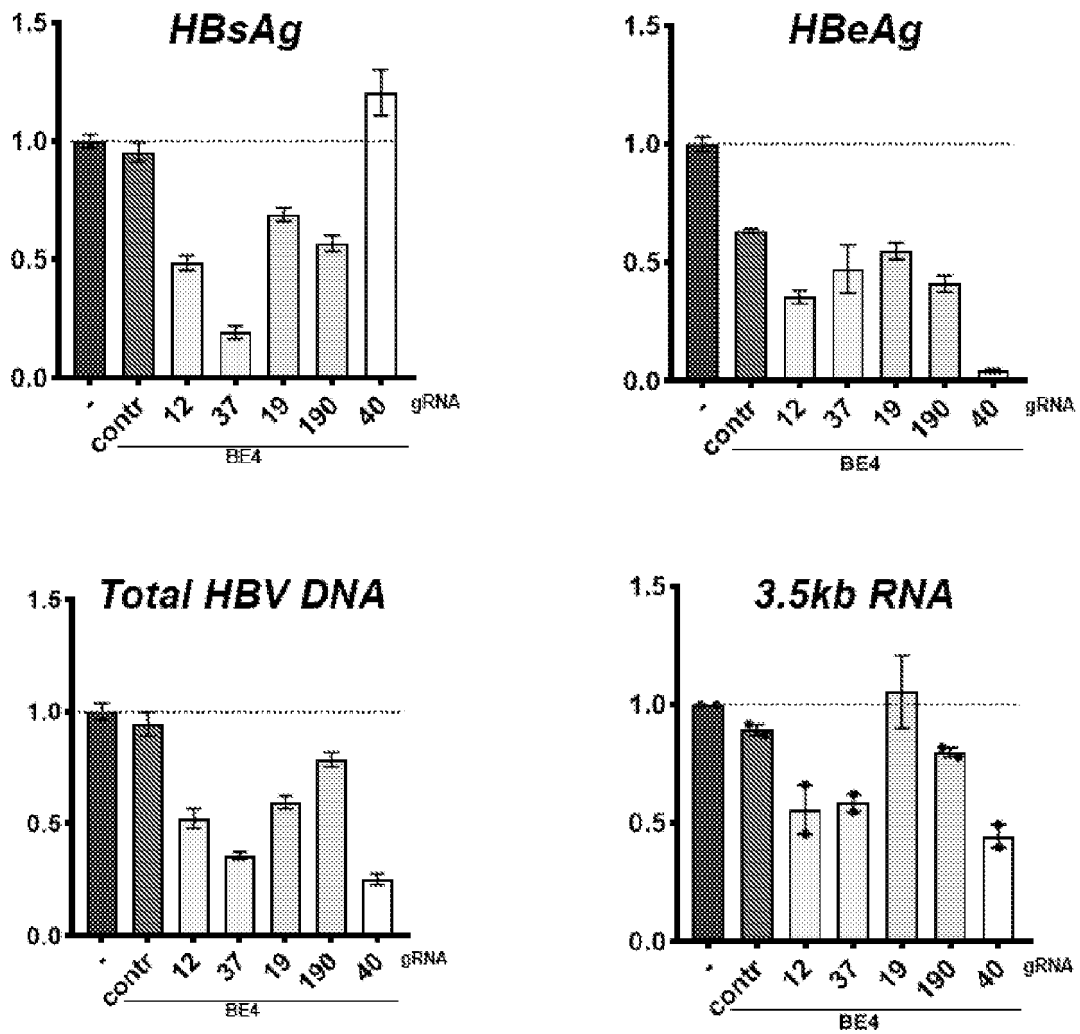


FIG. 8

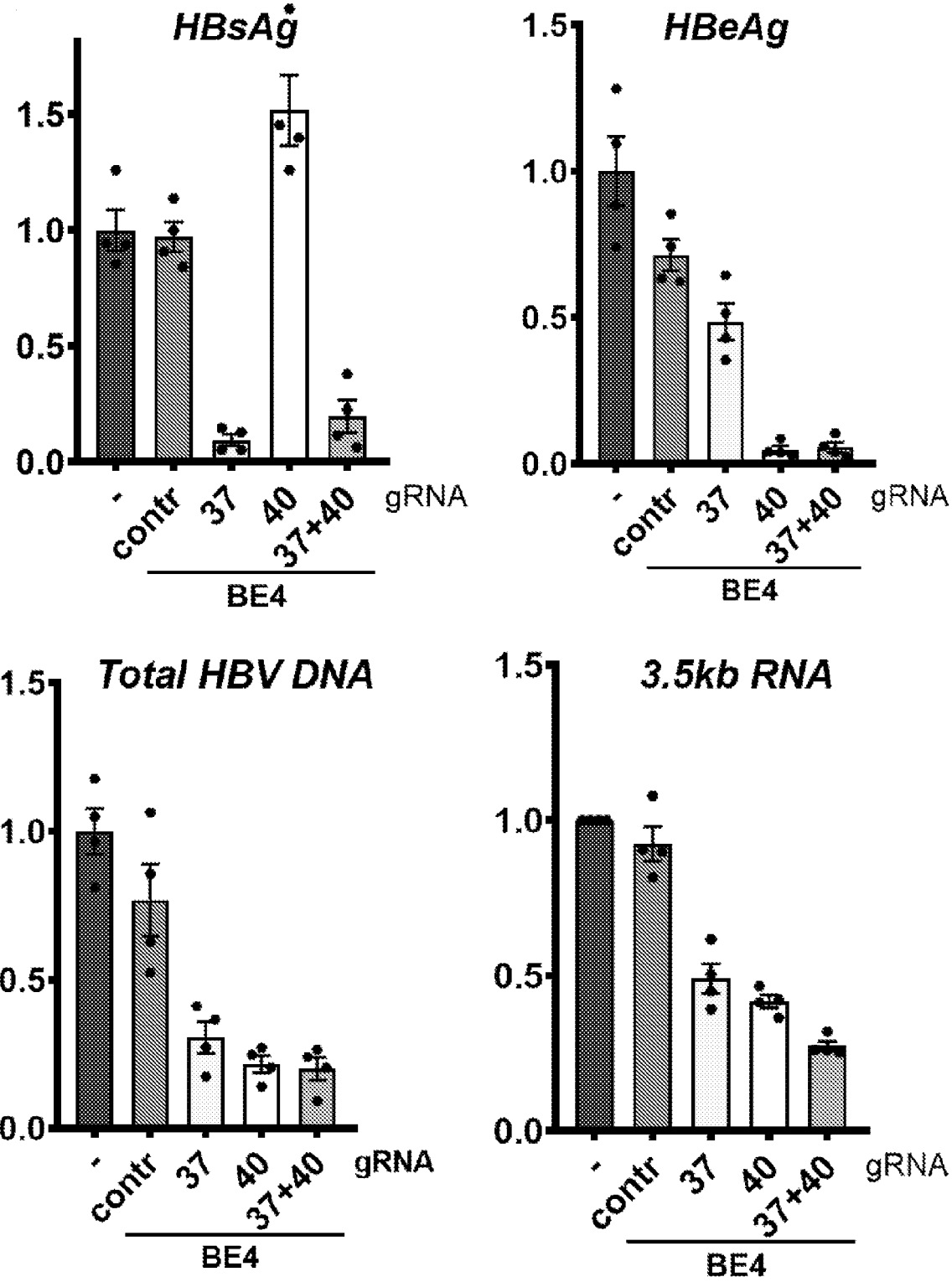


FIG. 9

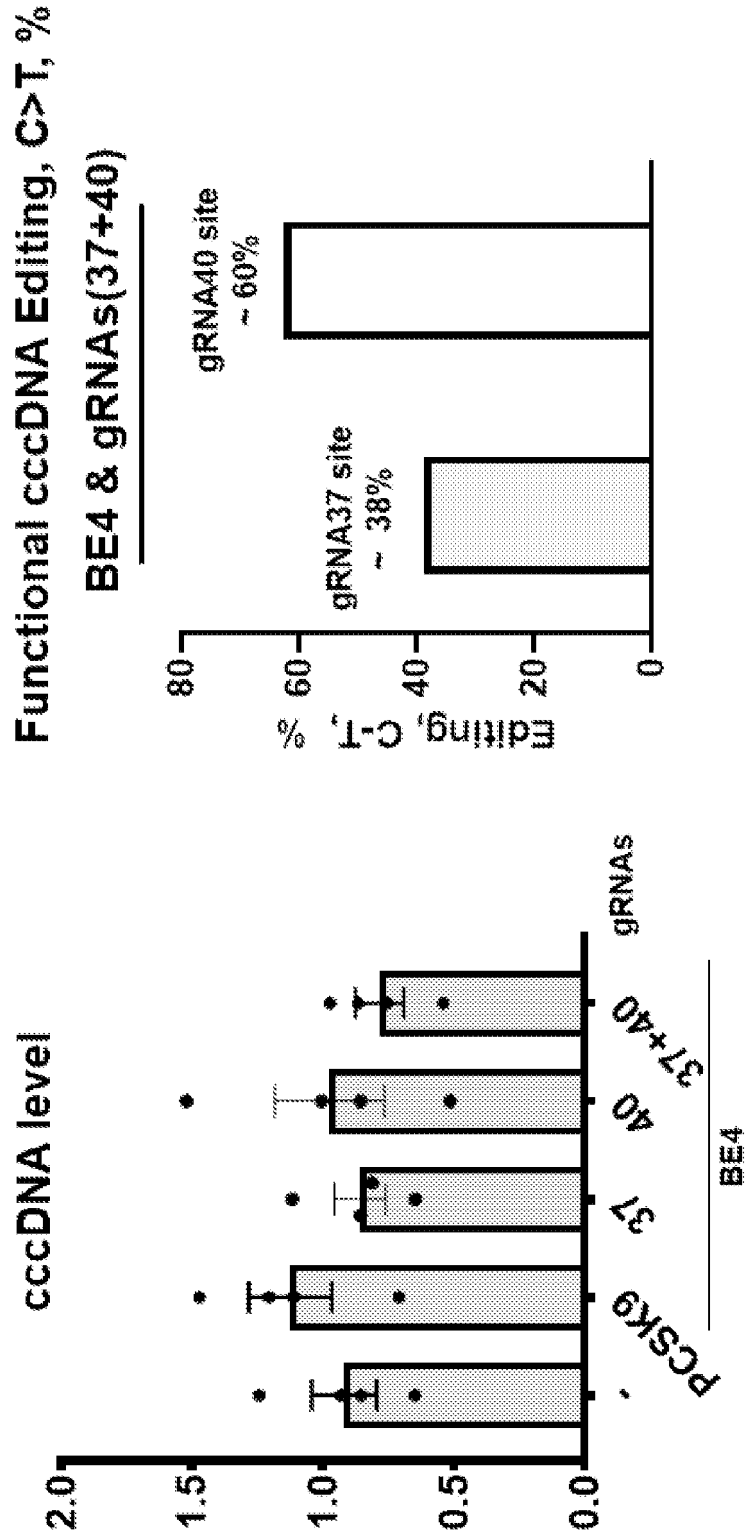


FIG. 10

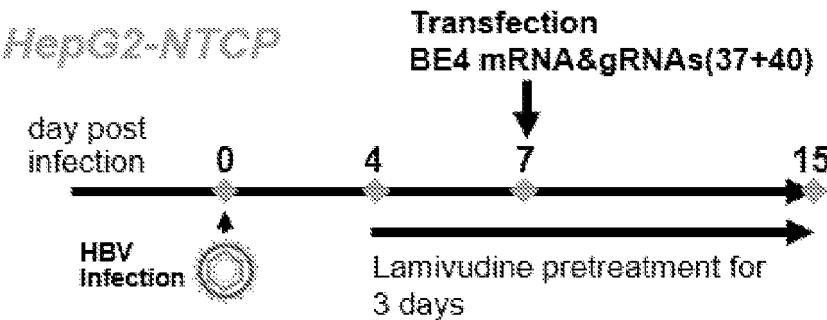


FIG. 11

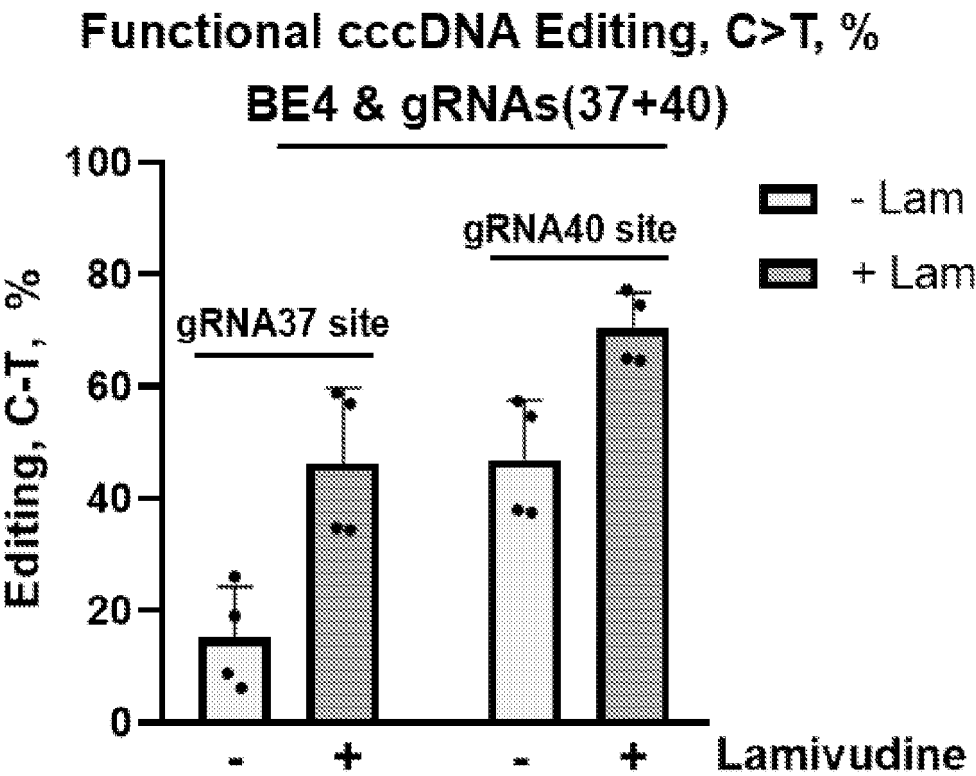


FIG. 12

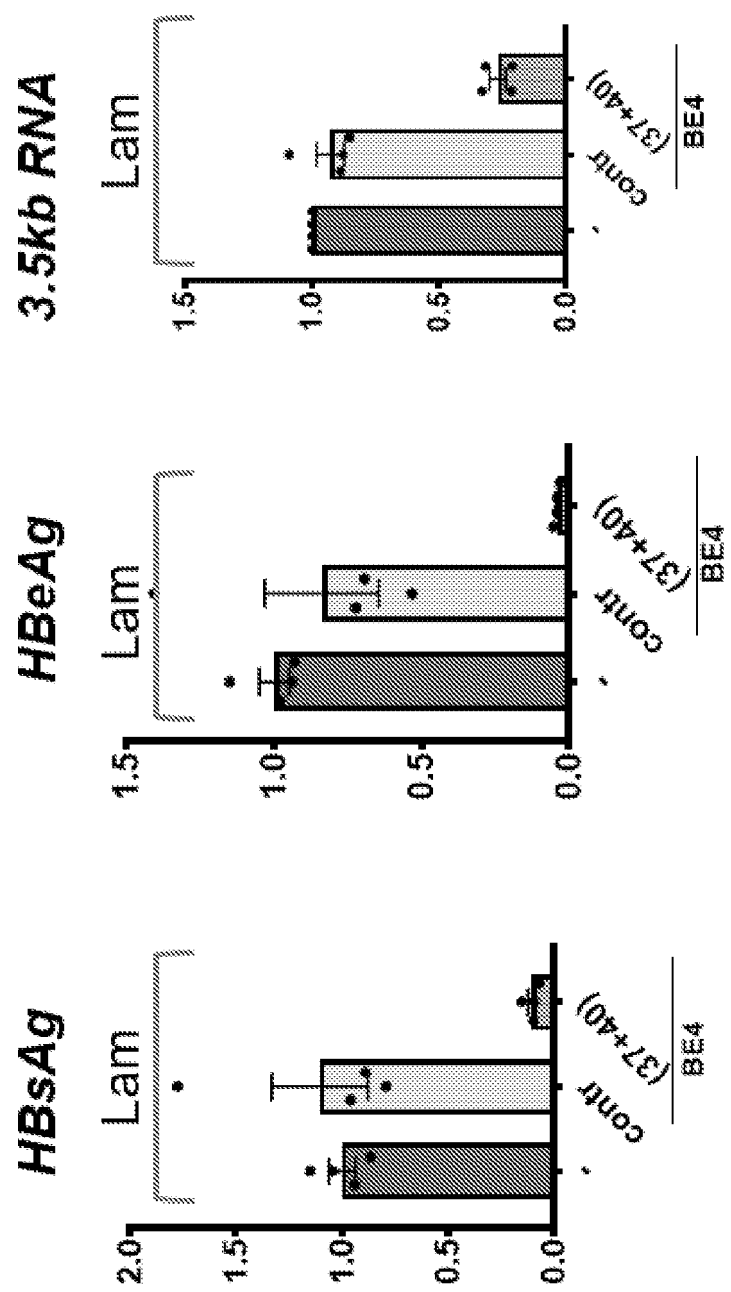


FIG. 13

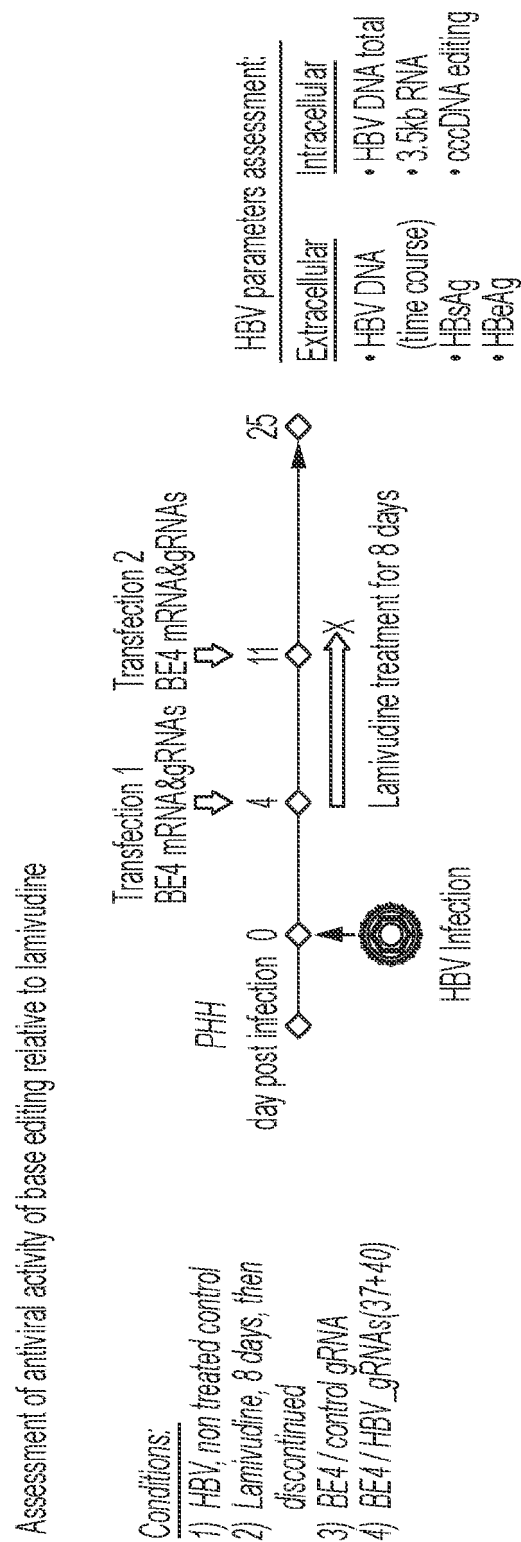


FIG. 14

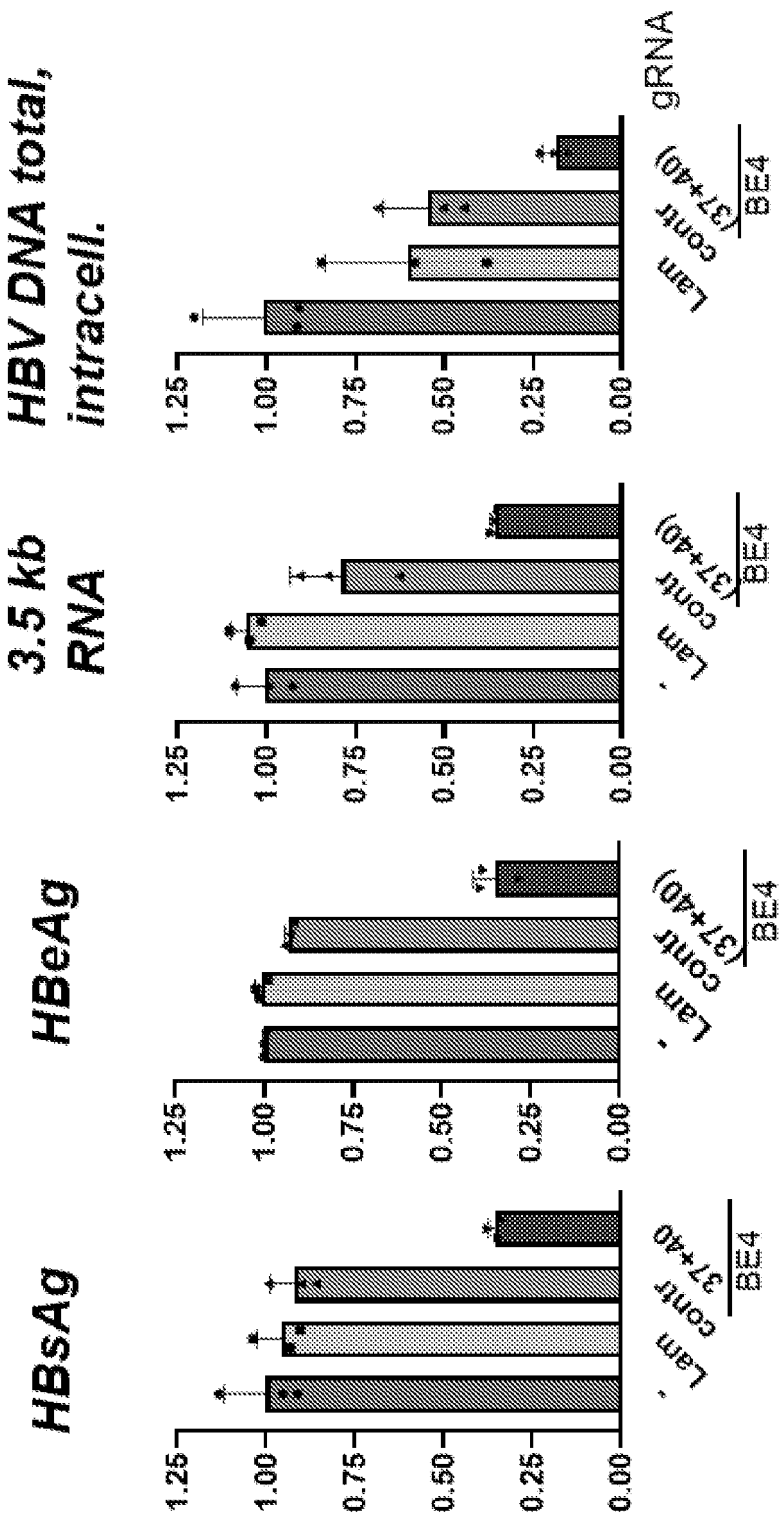


FIG. 15

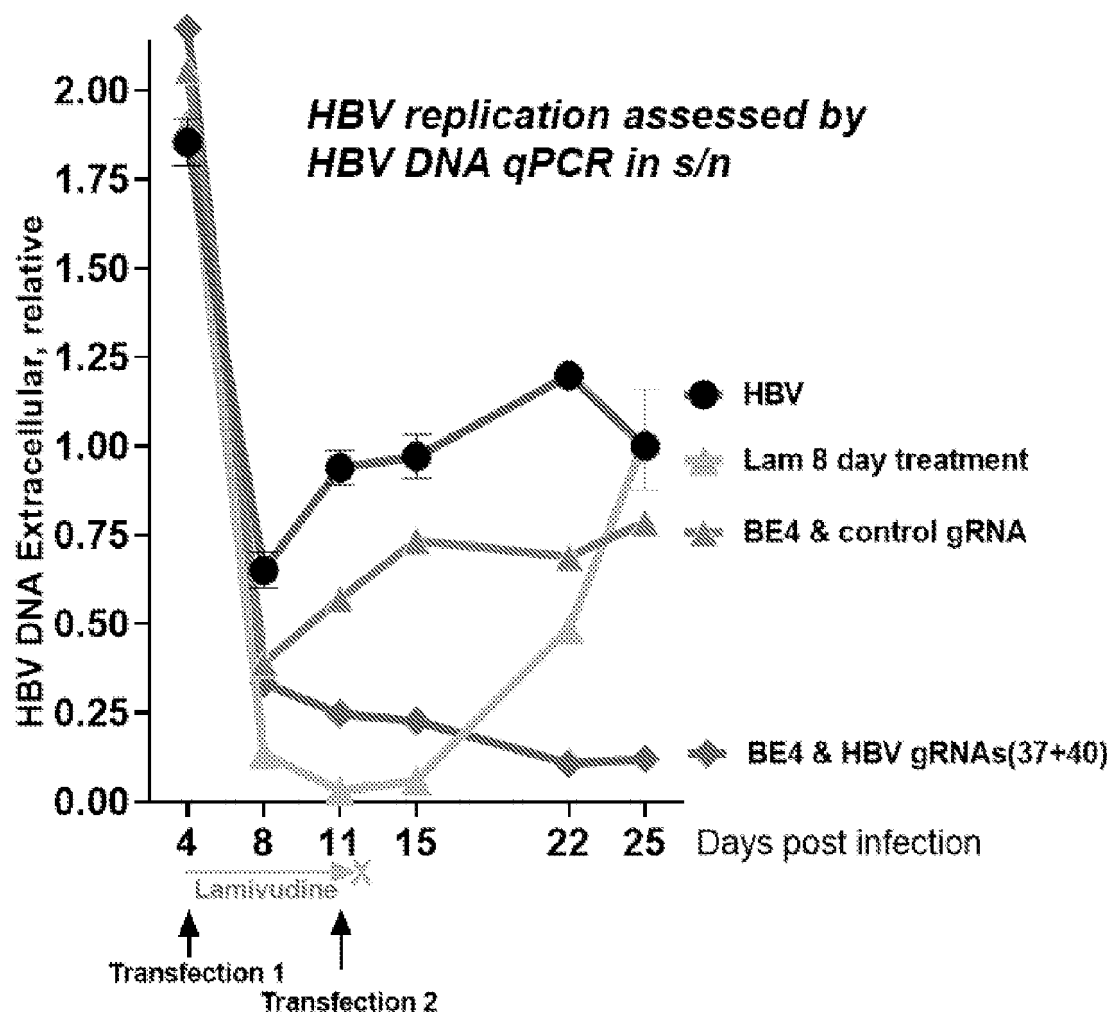


FIG. 16

Functional cccDNA Editing, C>T, %

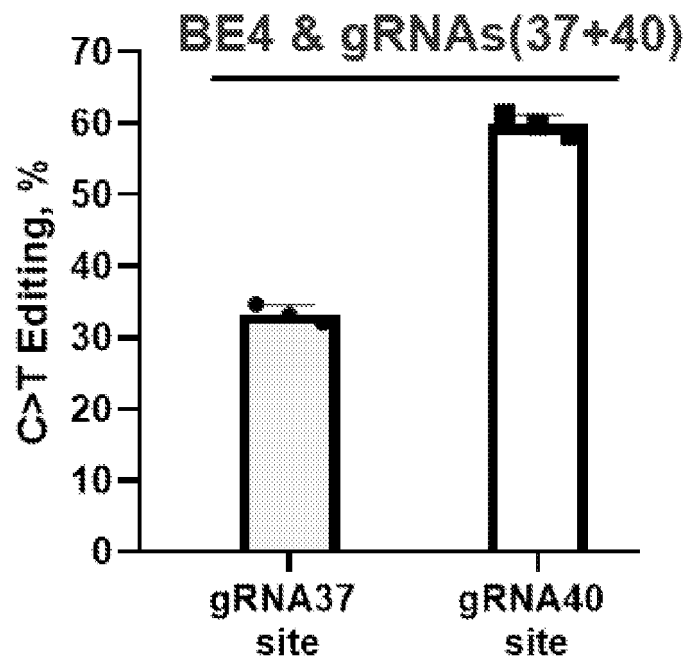


FIG. 17

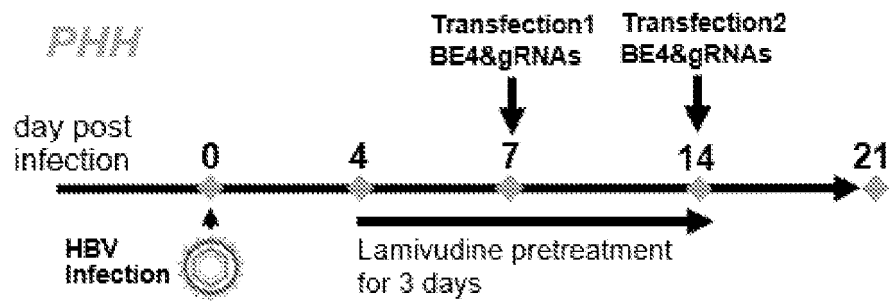


FIG. 18

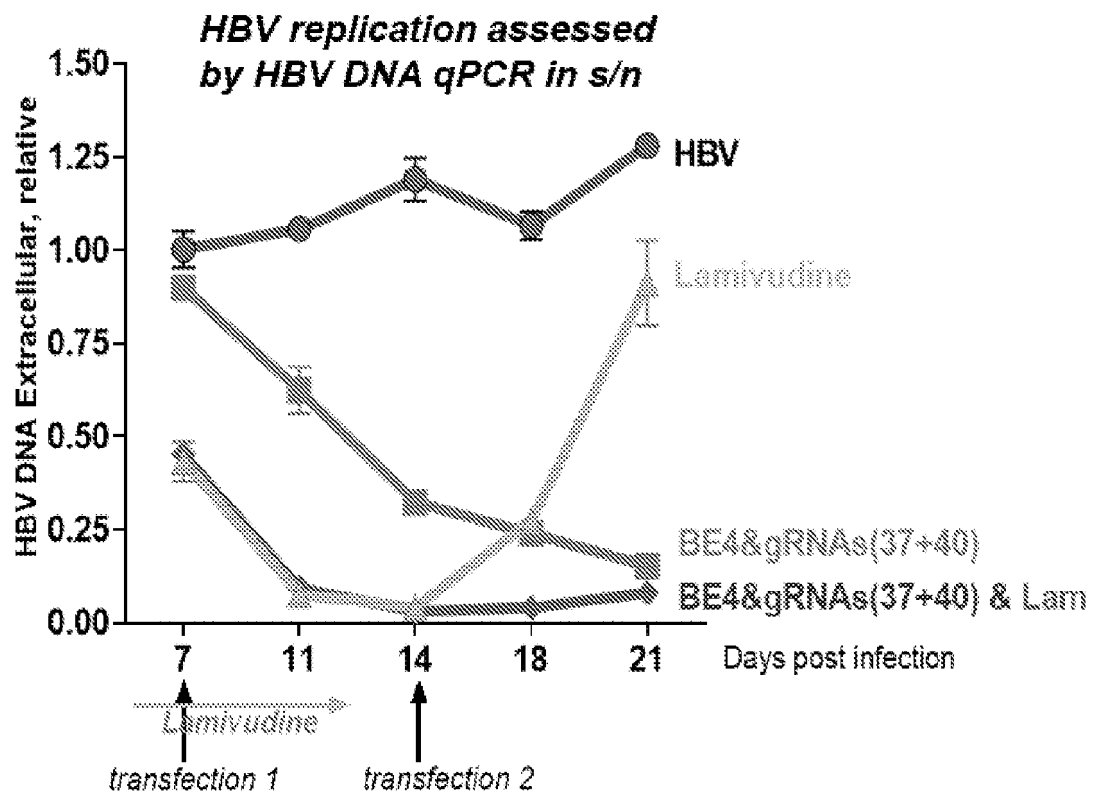


FIG. 19

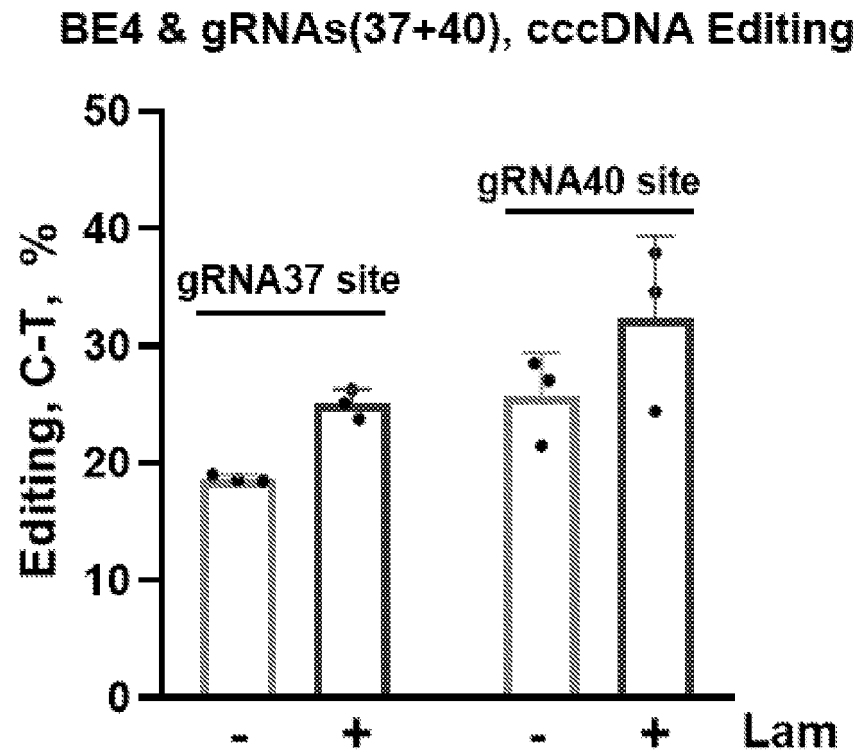


FIG. 20

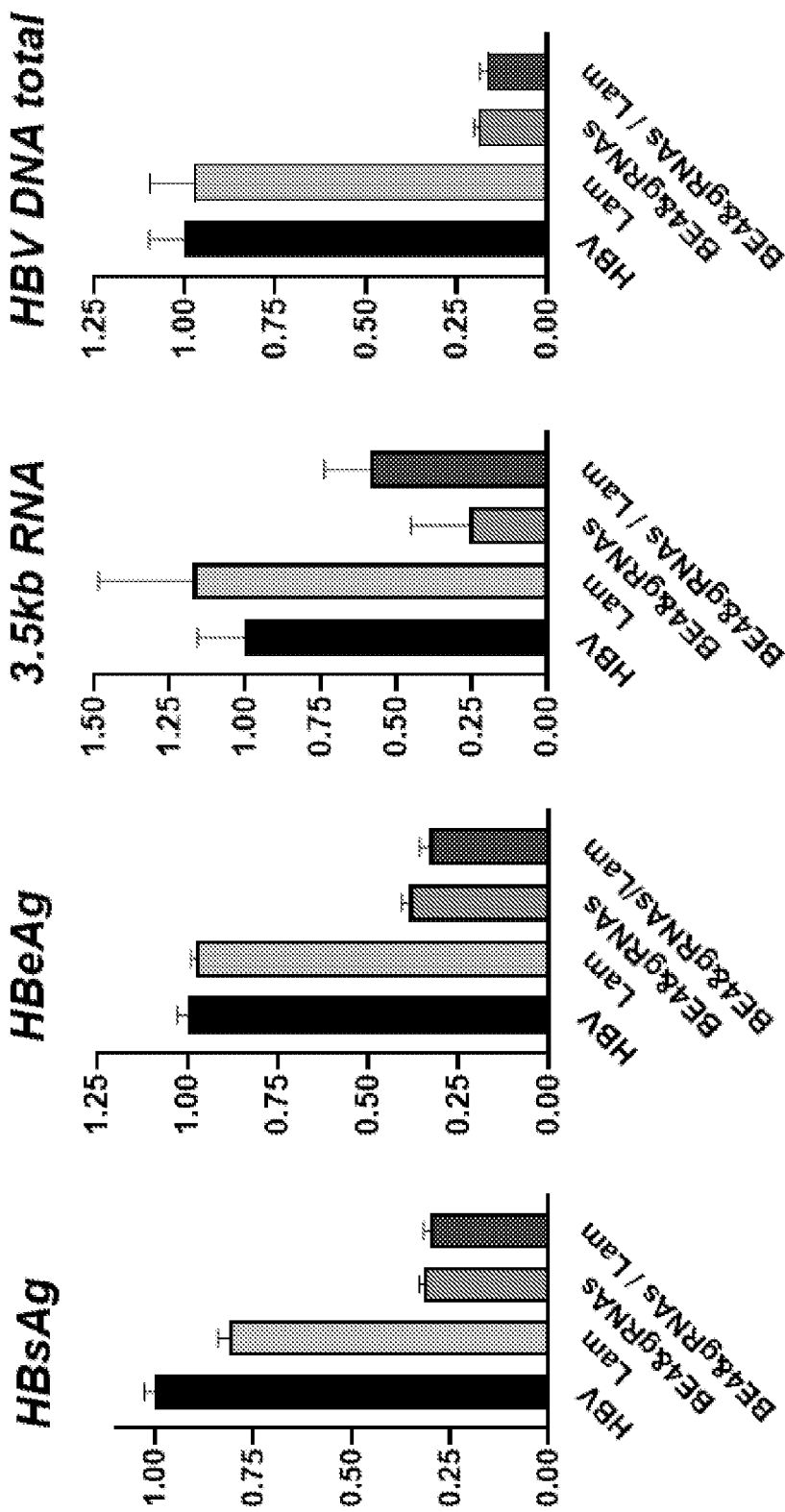


FIG. 21

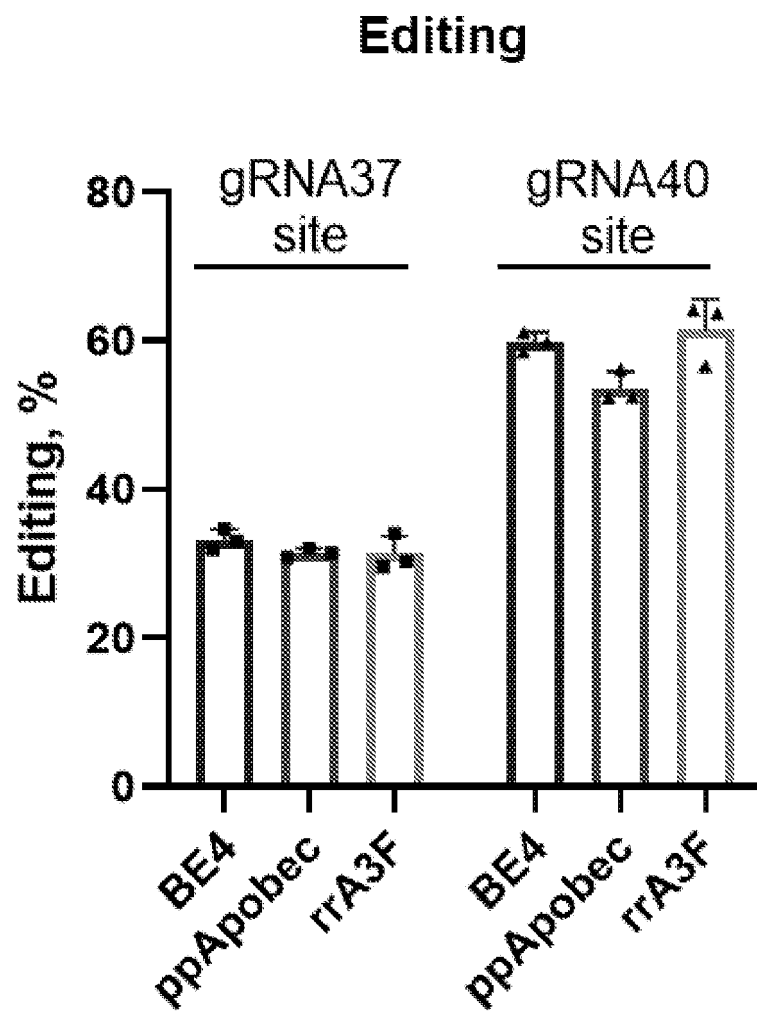


FIG. 22

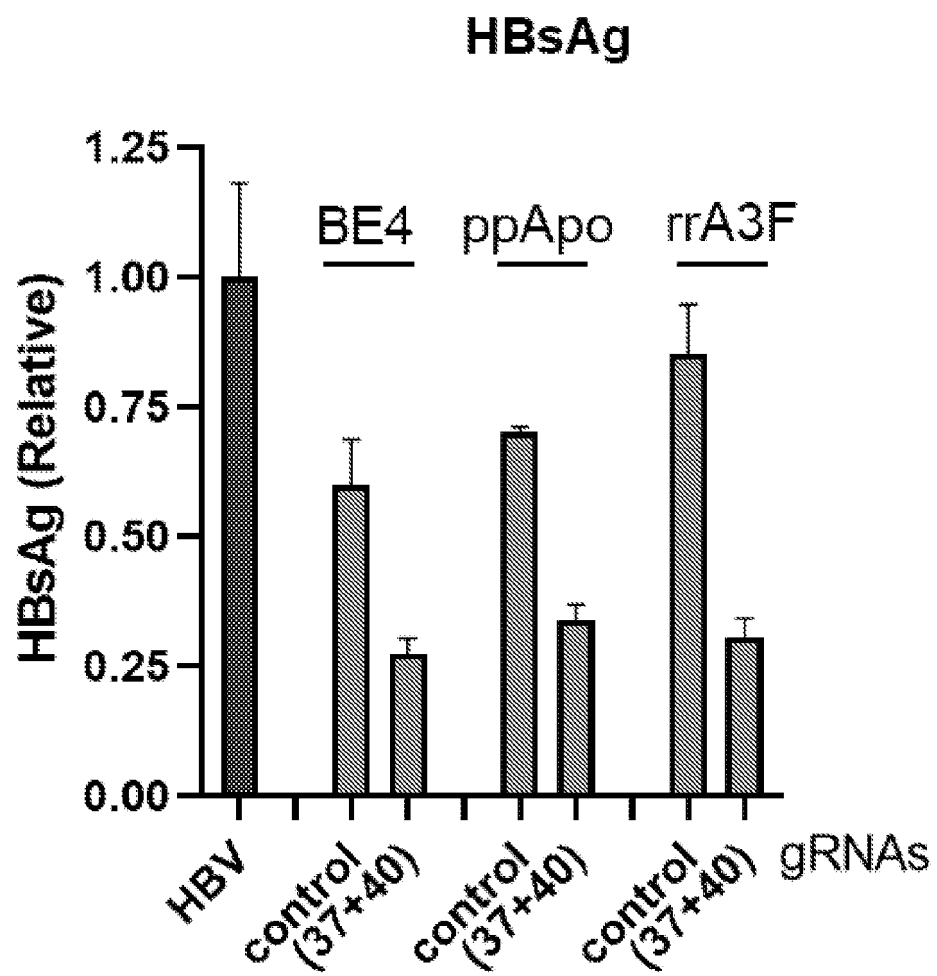


FIG. 23

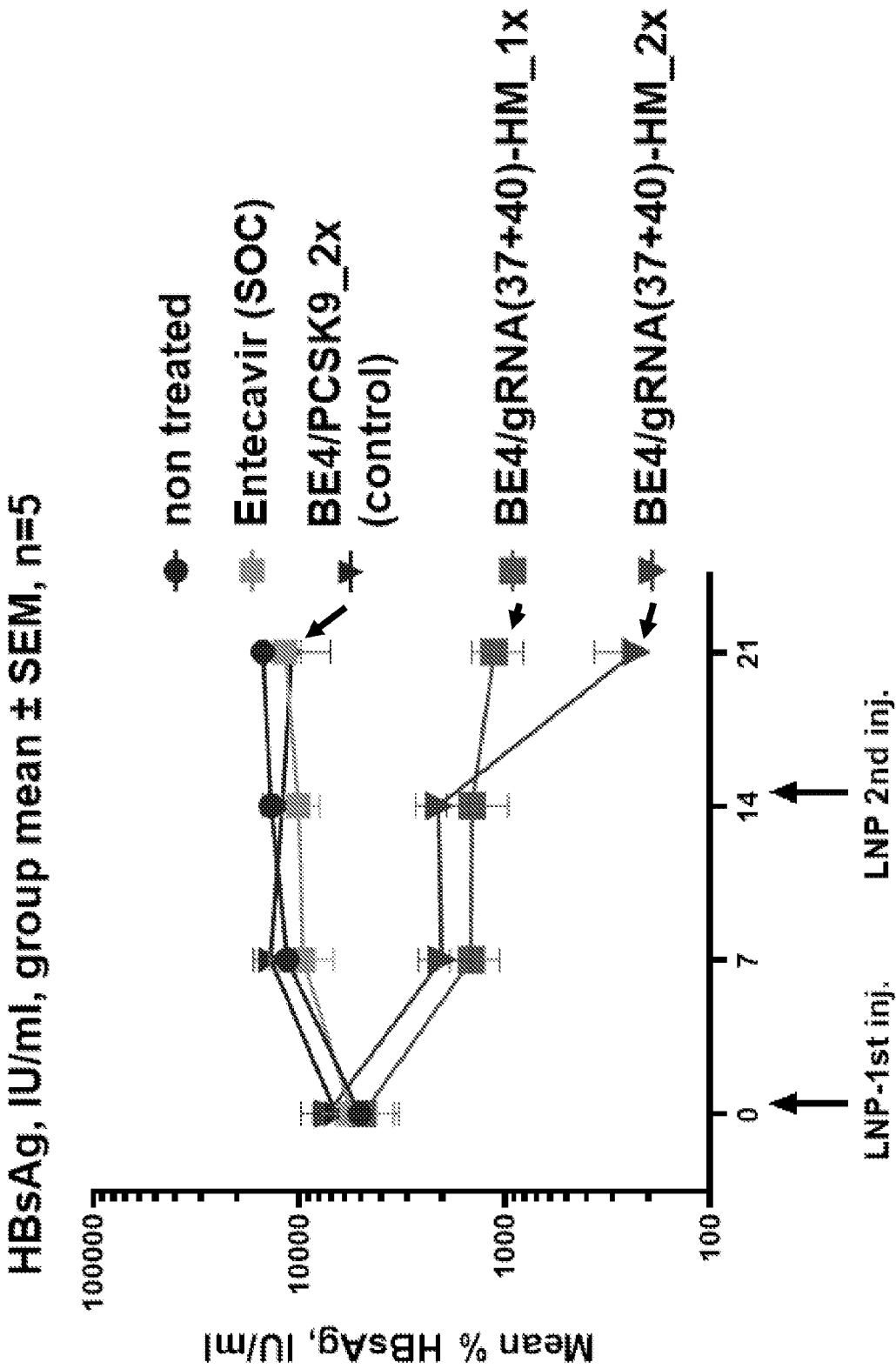


FIG. 24A

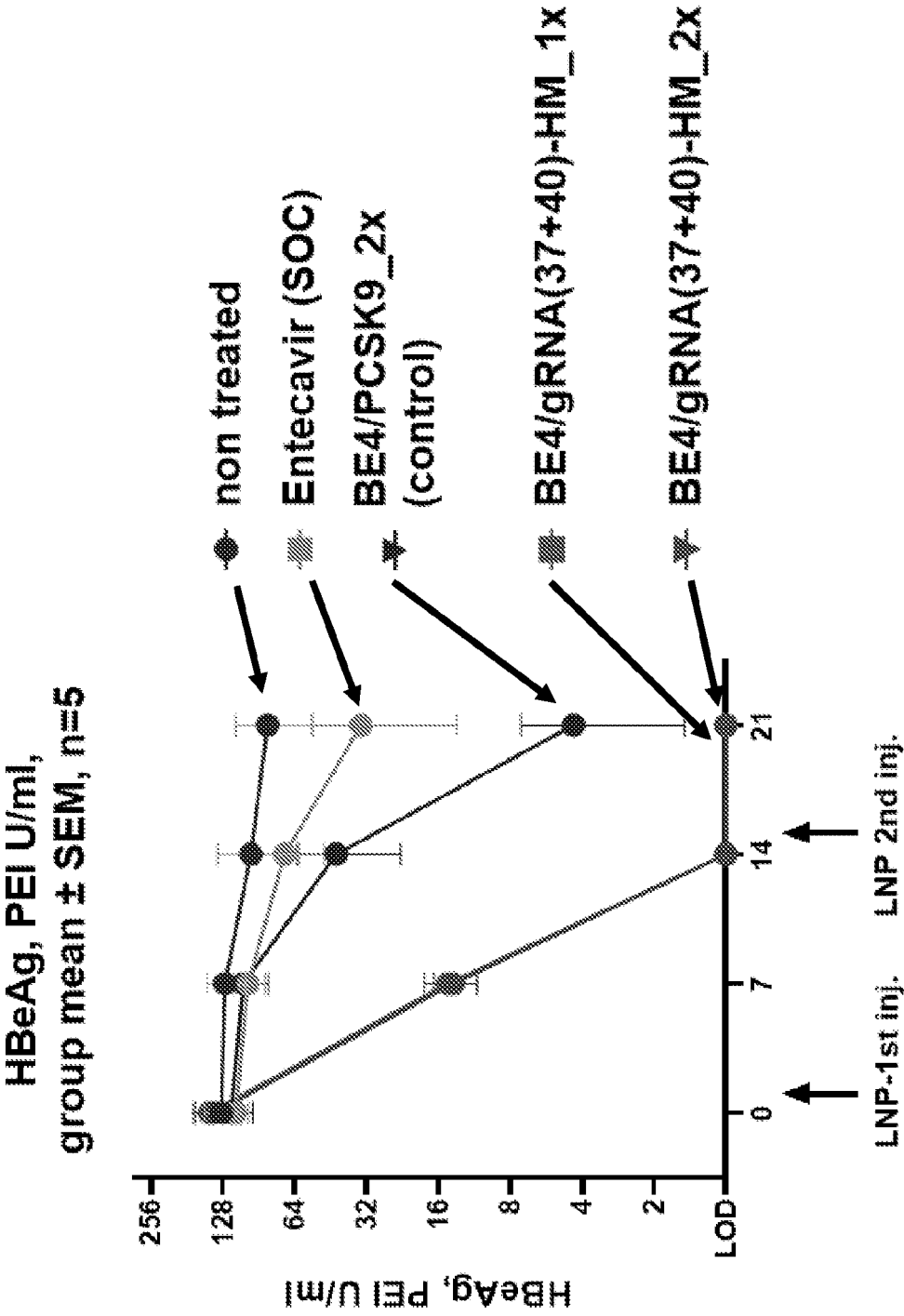


FIG. 24B

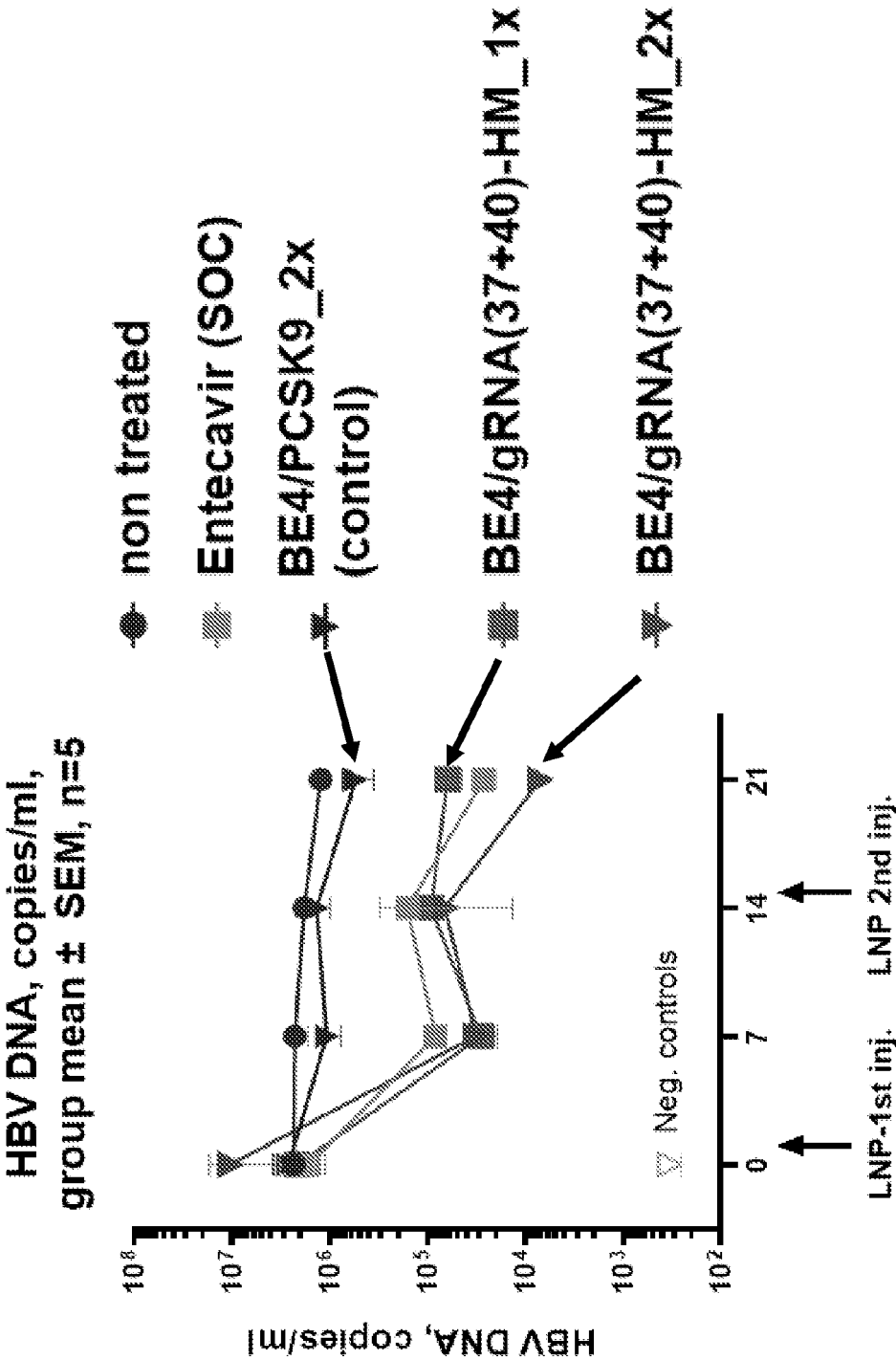


FIG. 24C

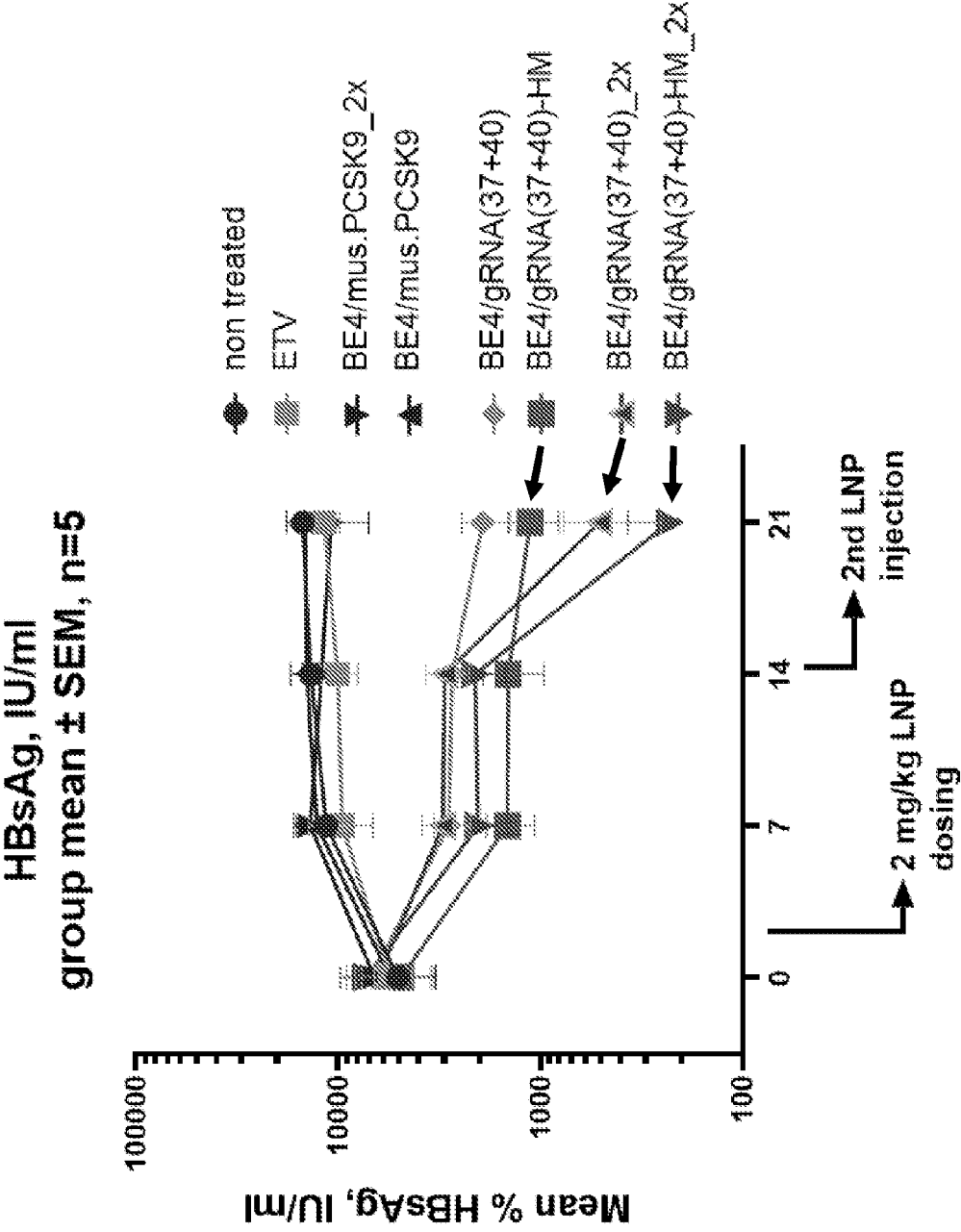


FIG. 25A

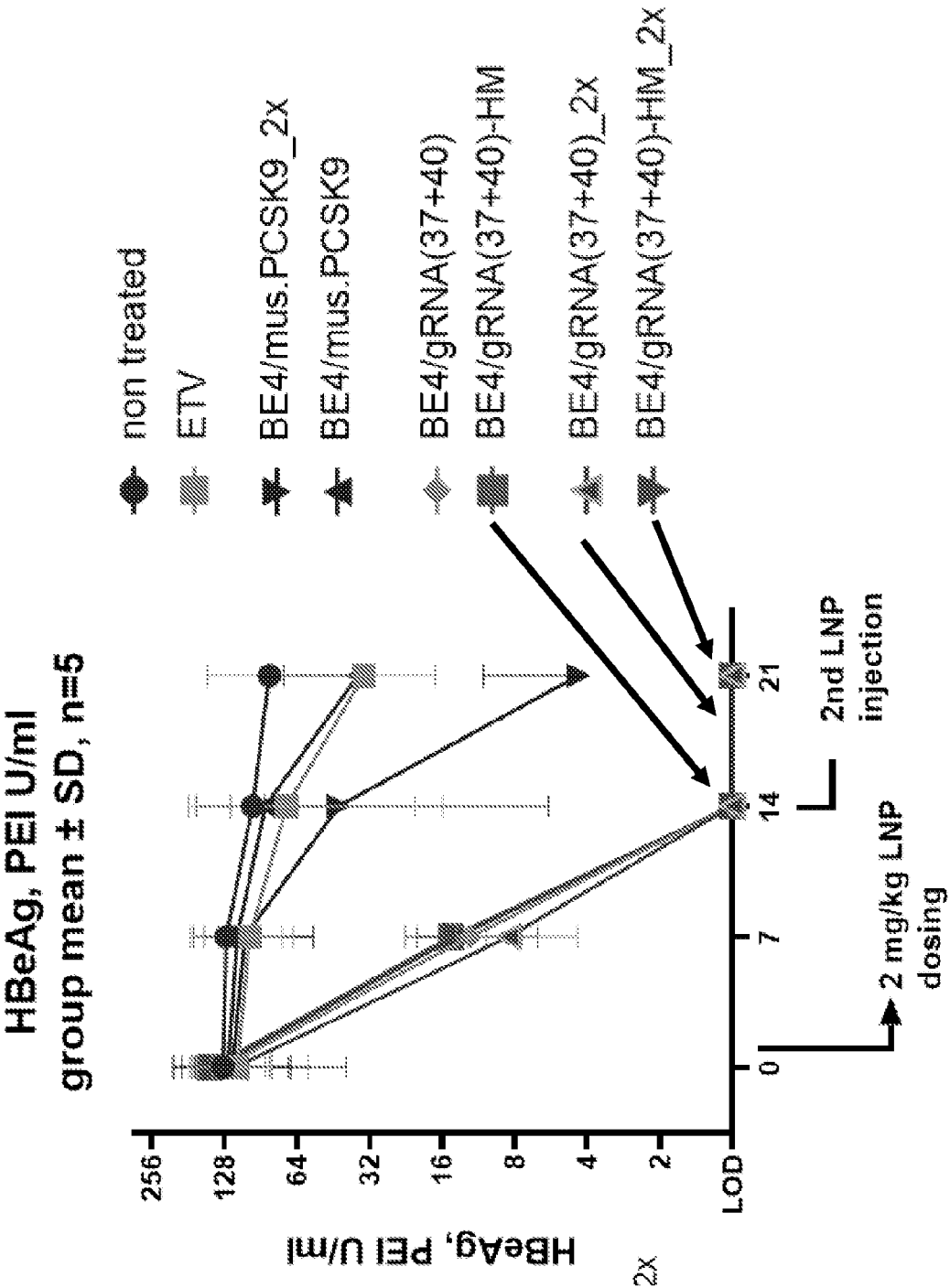


FIG. 25B

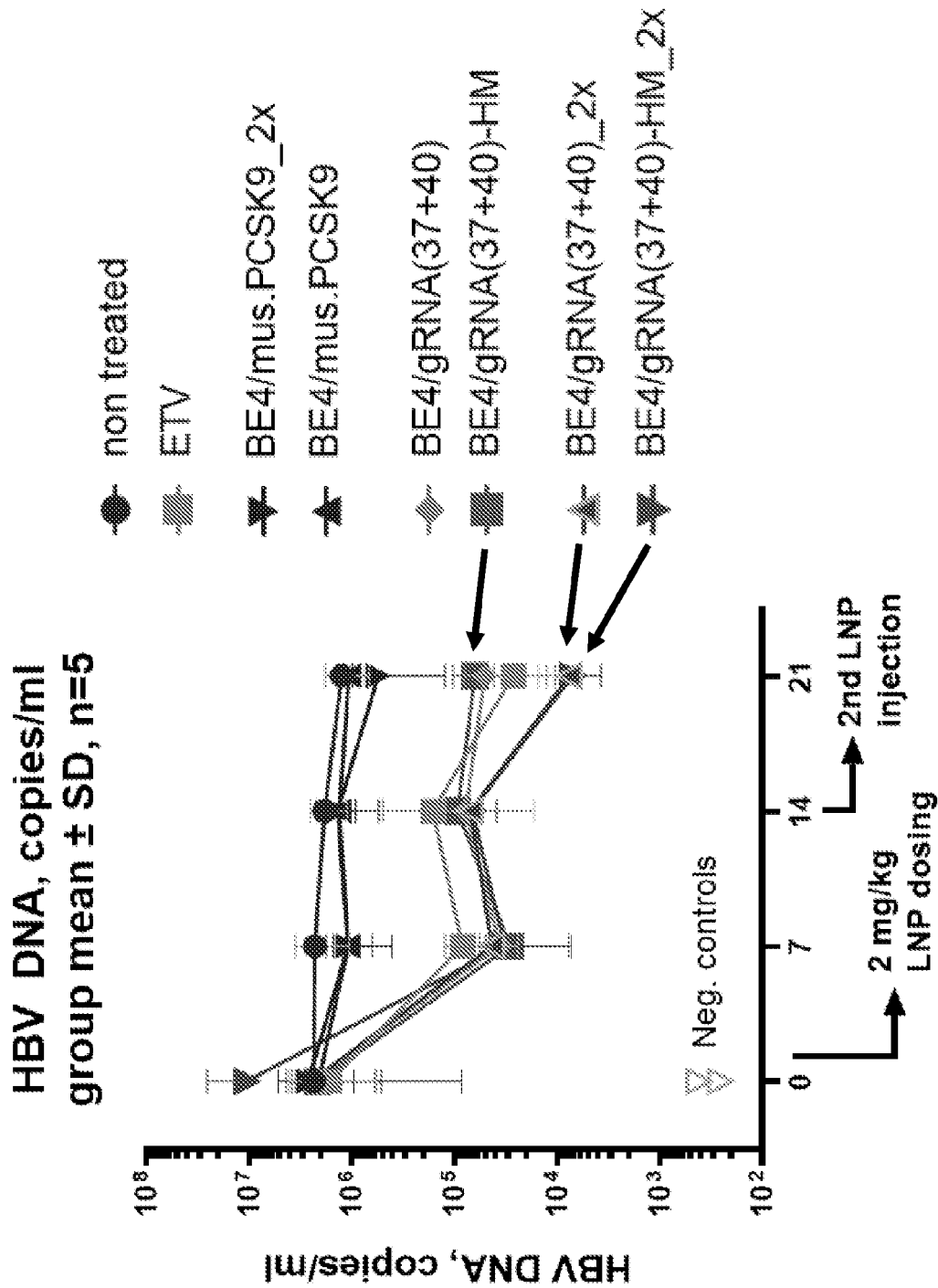


FIG. 25C

FIG. 26B

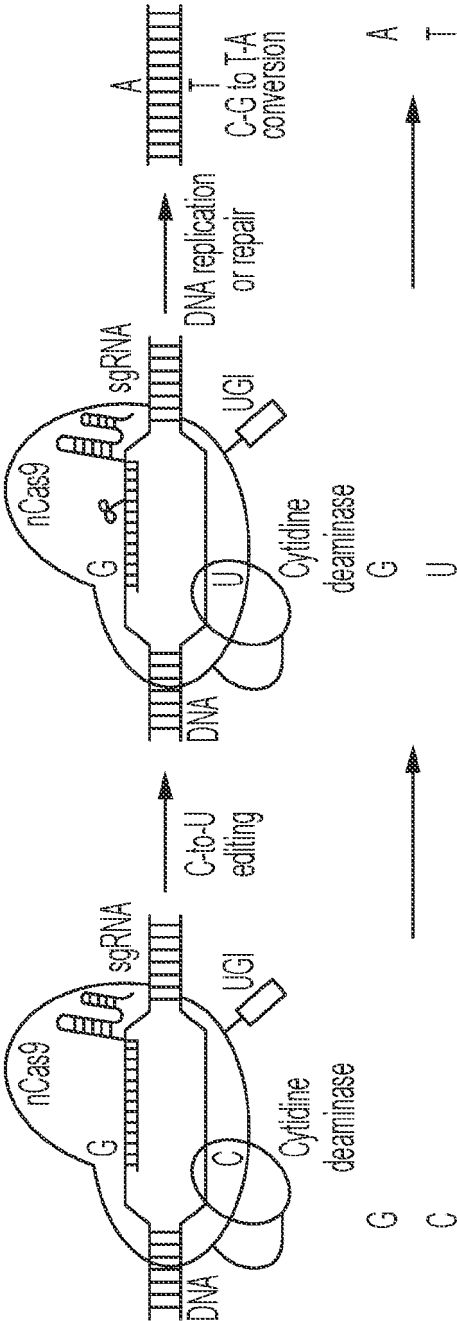


FIG. 27A-27D

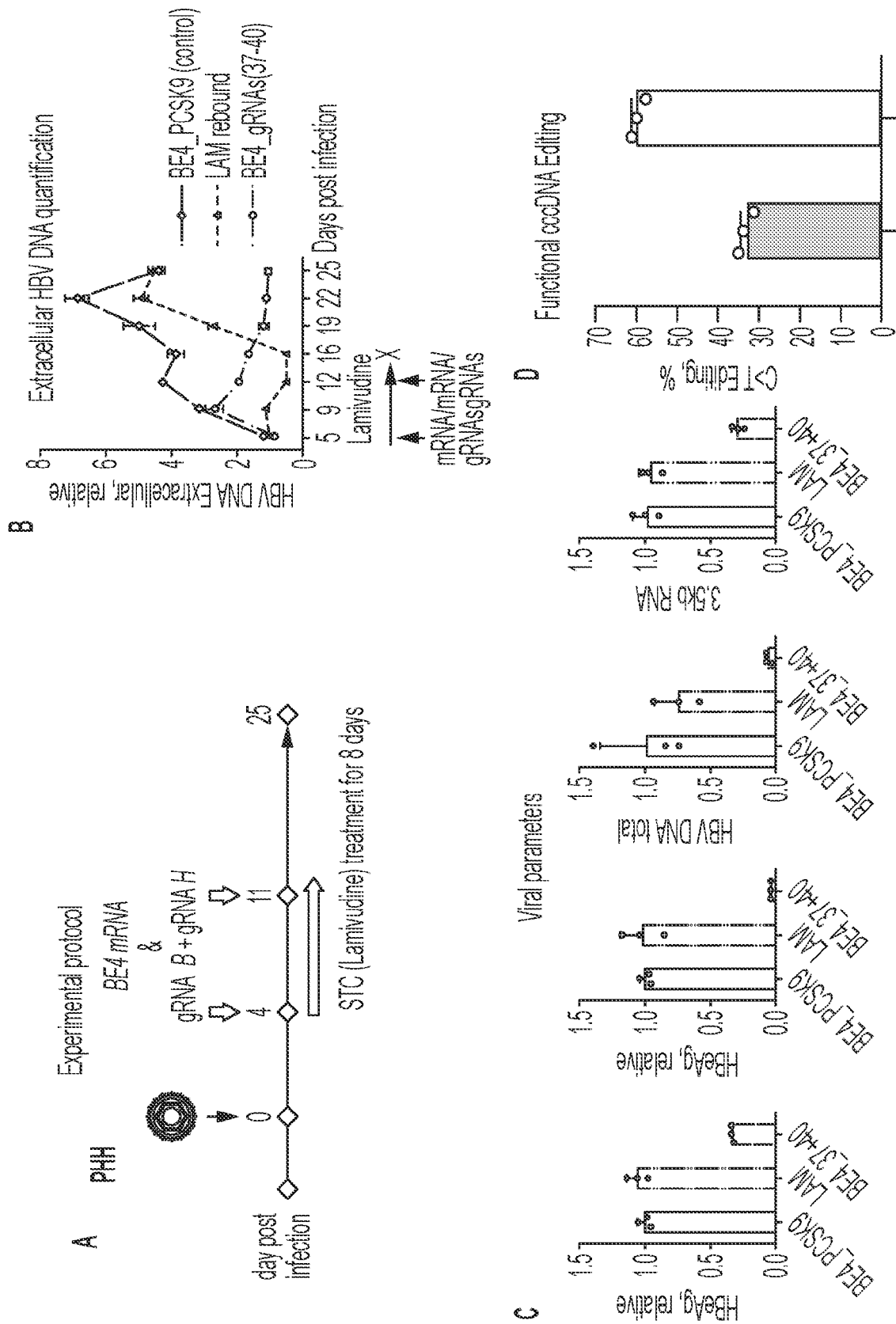


FIG. 28A and 28B

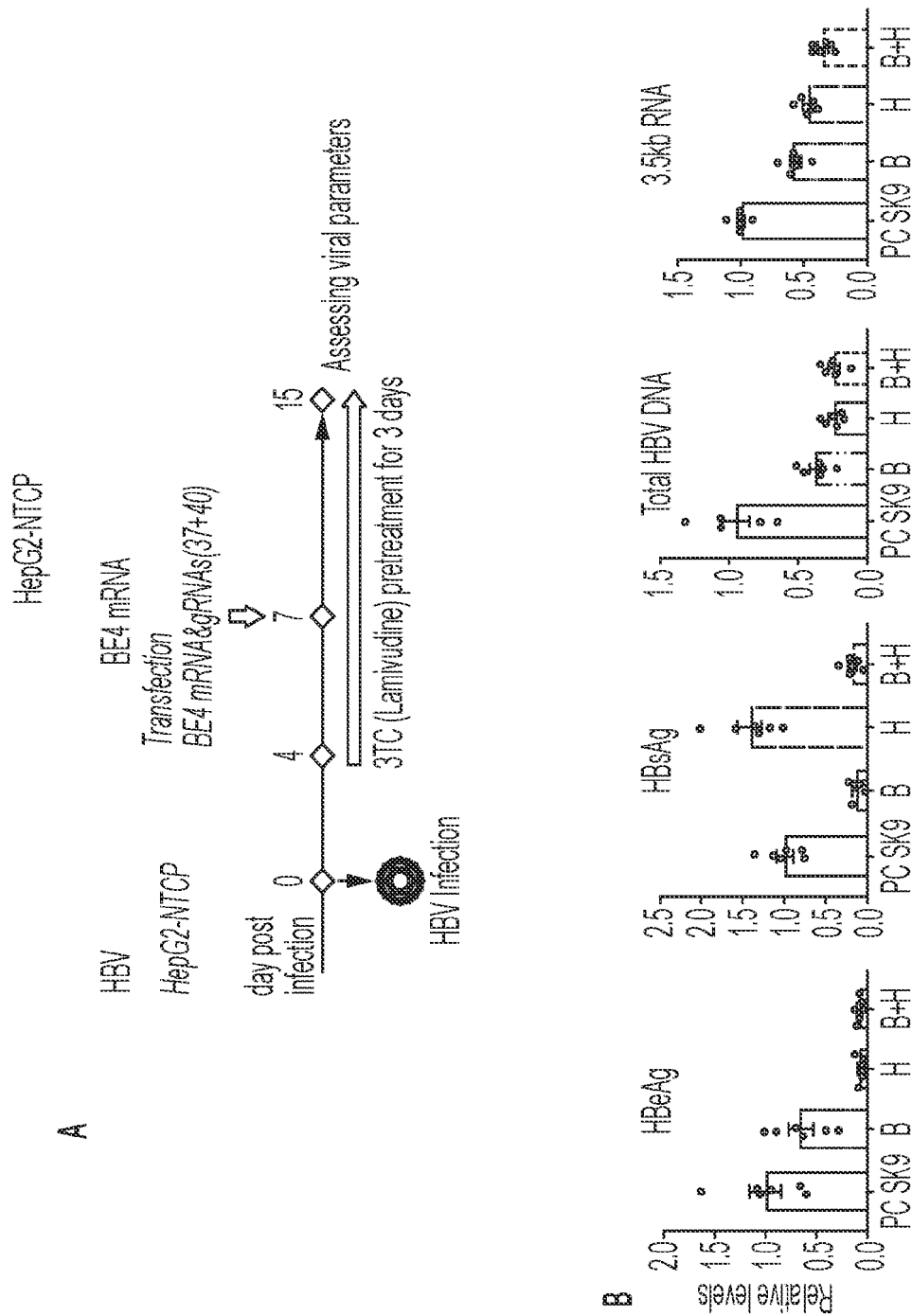


FIG. 28C-28D

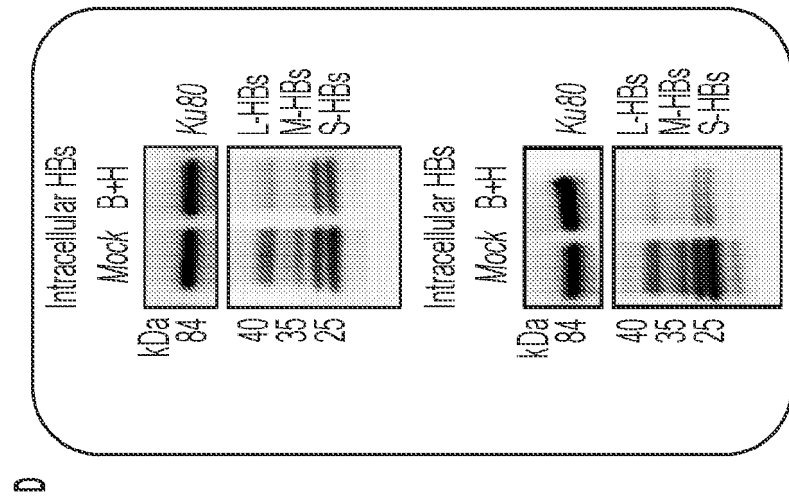
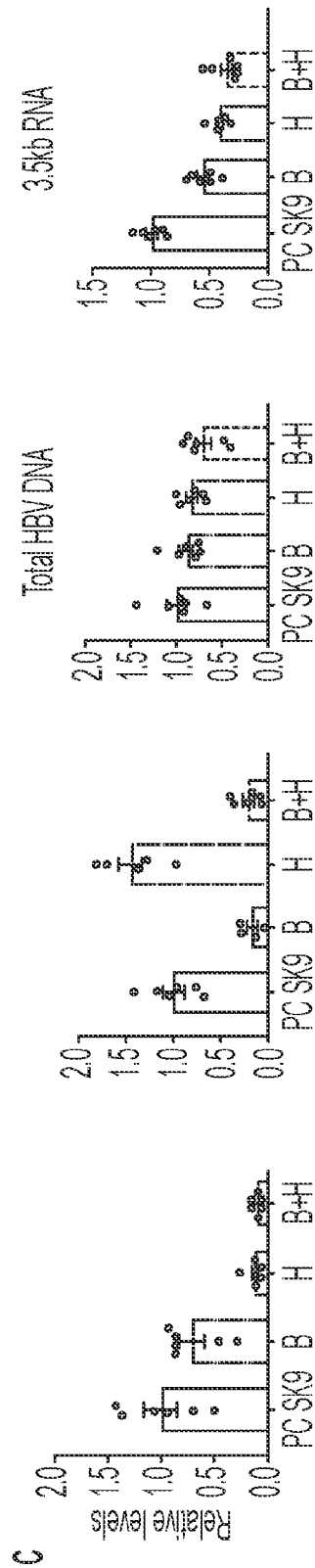


FIG. 29A-29C

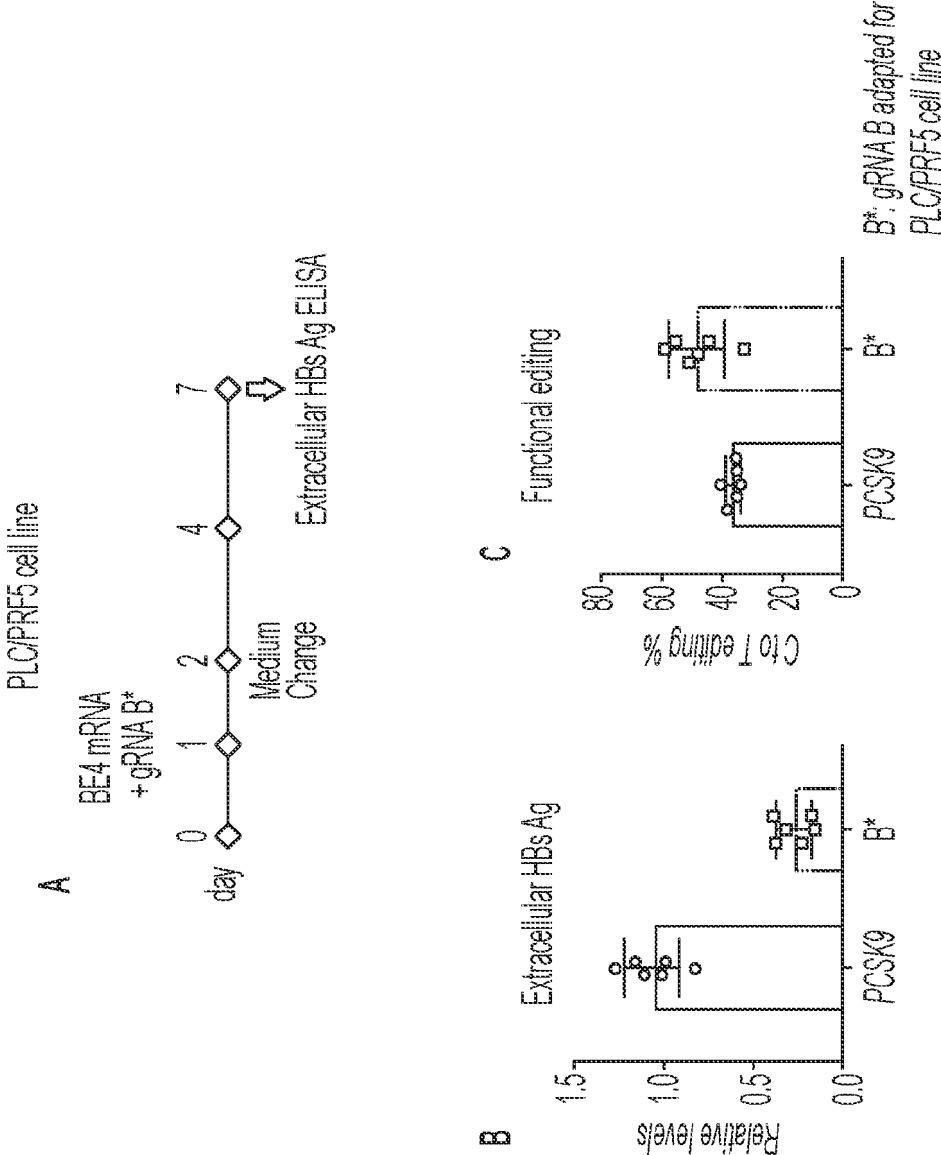
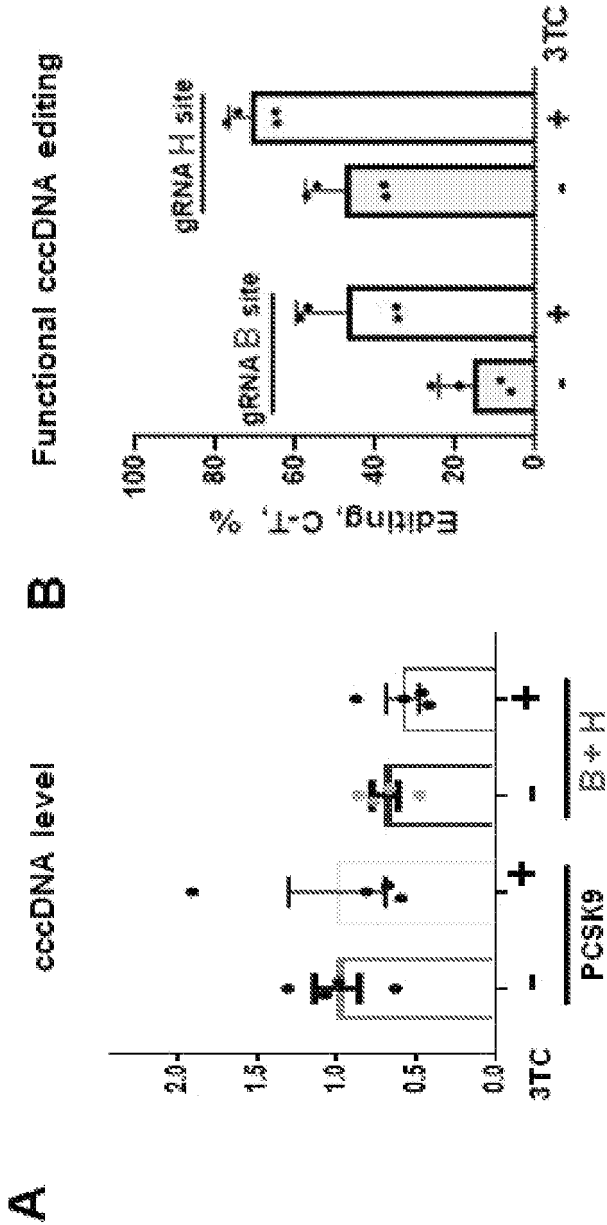


FIG. 30

HepG2-NTCP



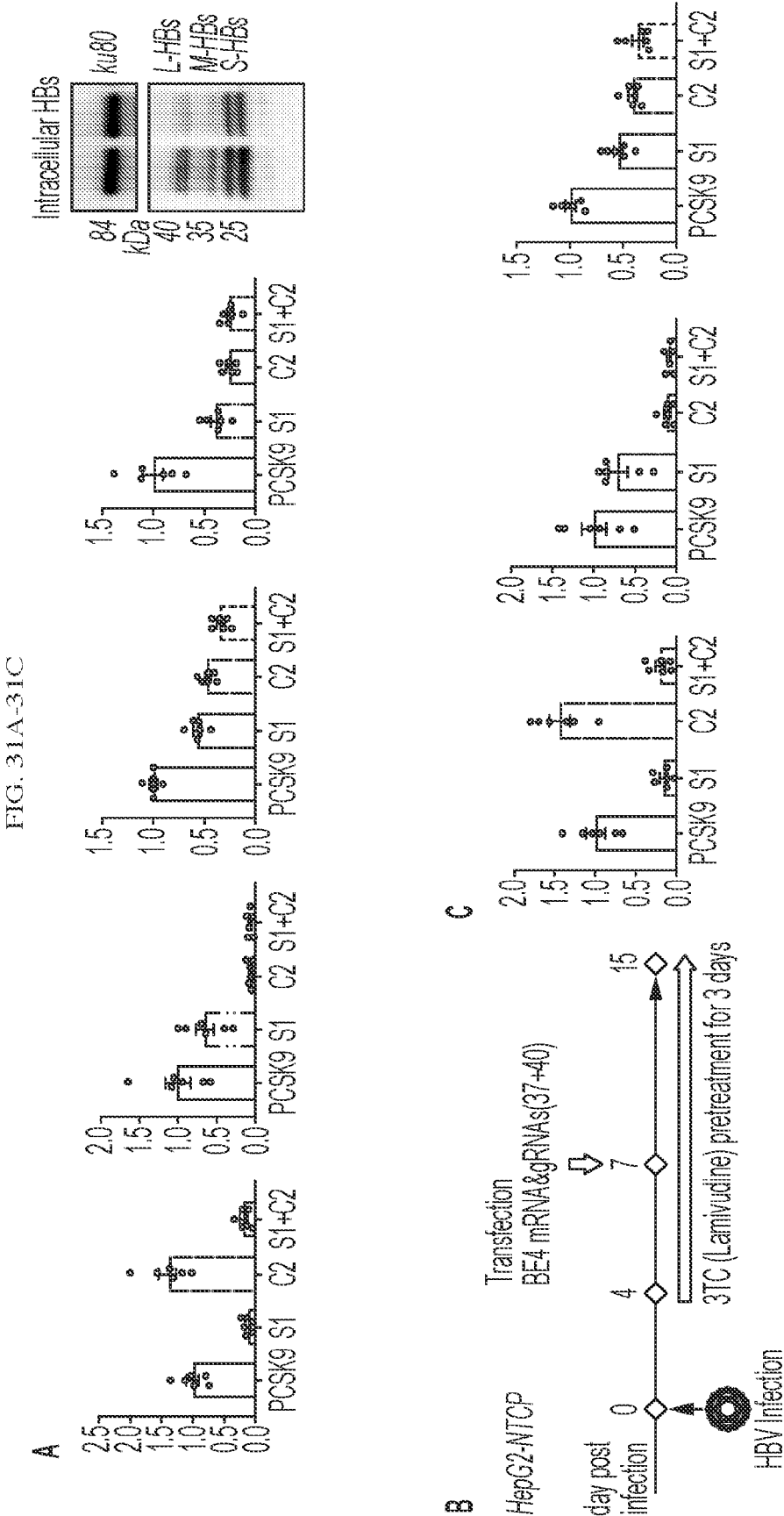


FIG. 31A-31C

FIGs. 31D and 31E

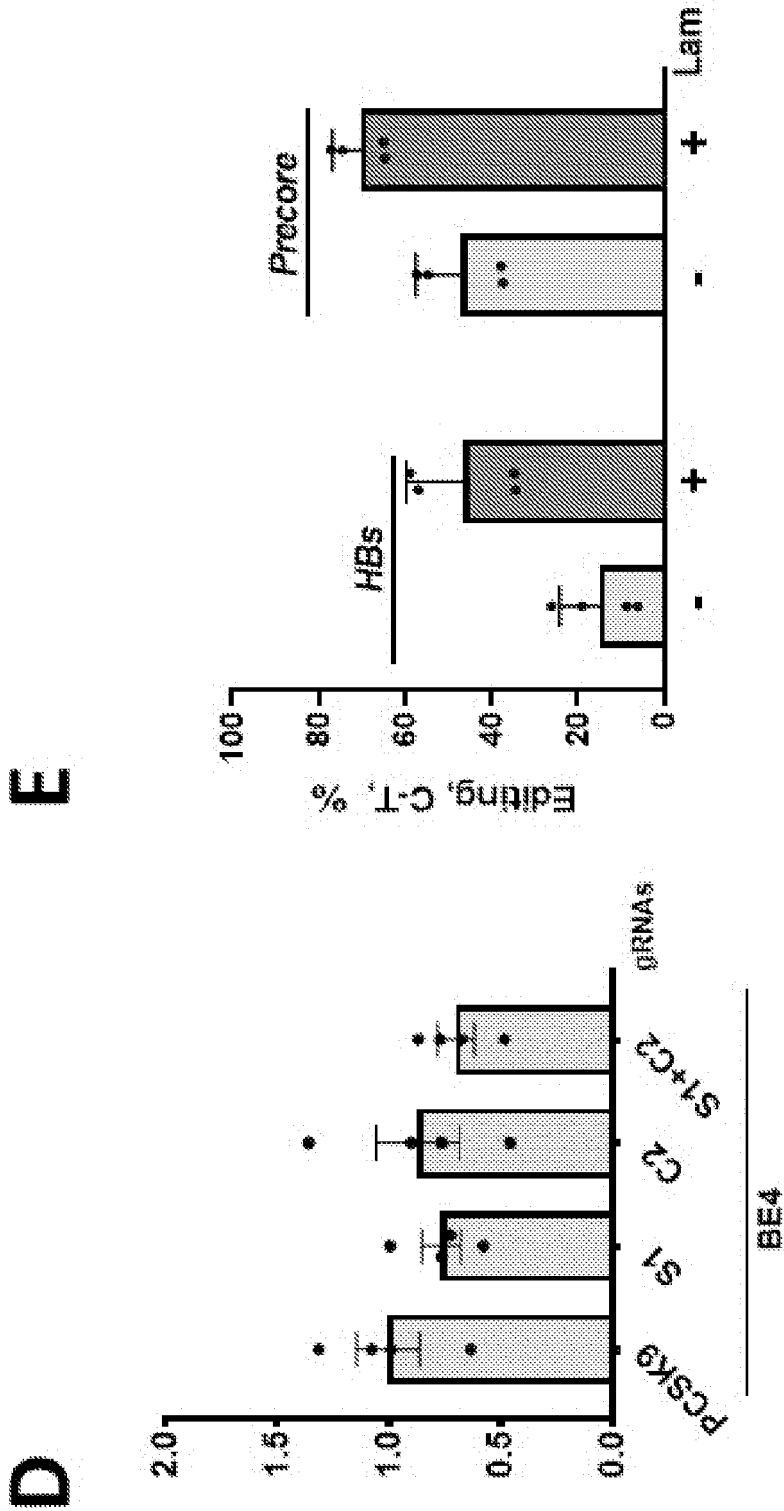


FIG. 32A and 32B

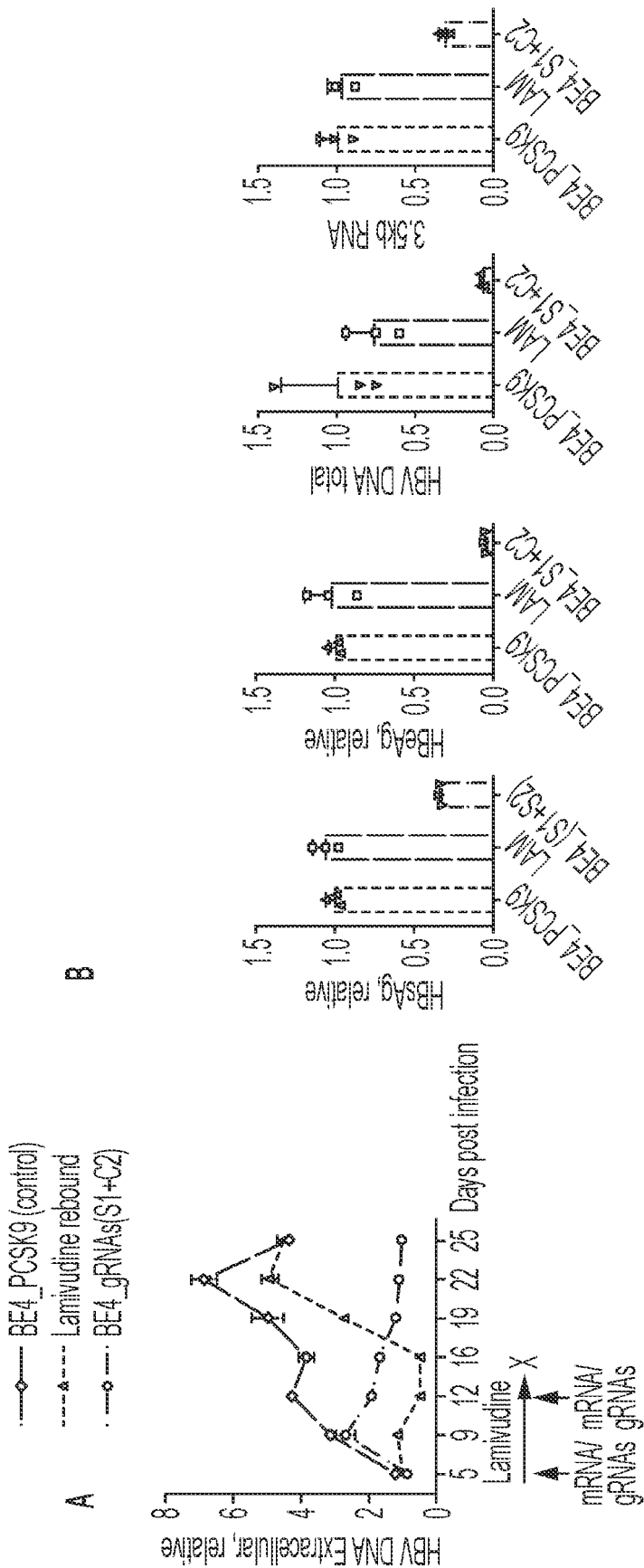


FIG. 32C

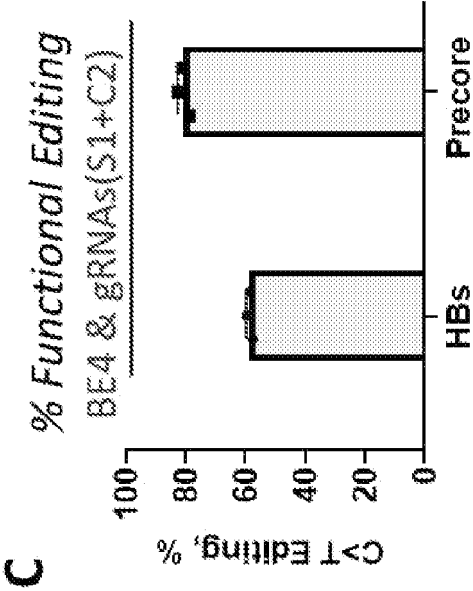


FIG. 33A and 33B

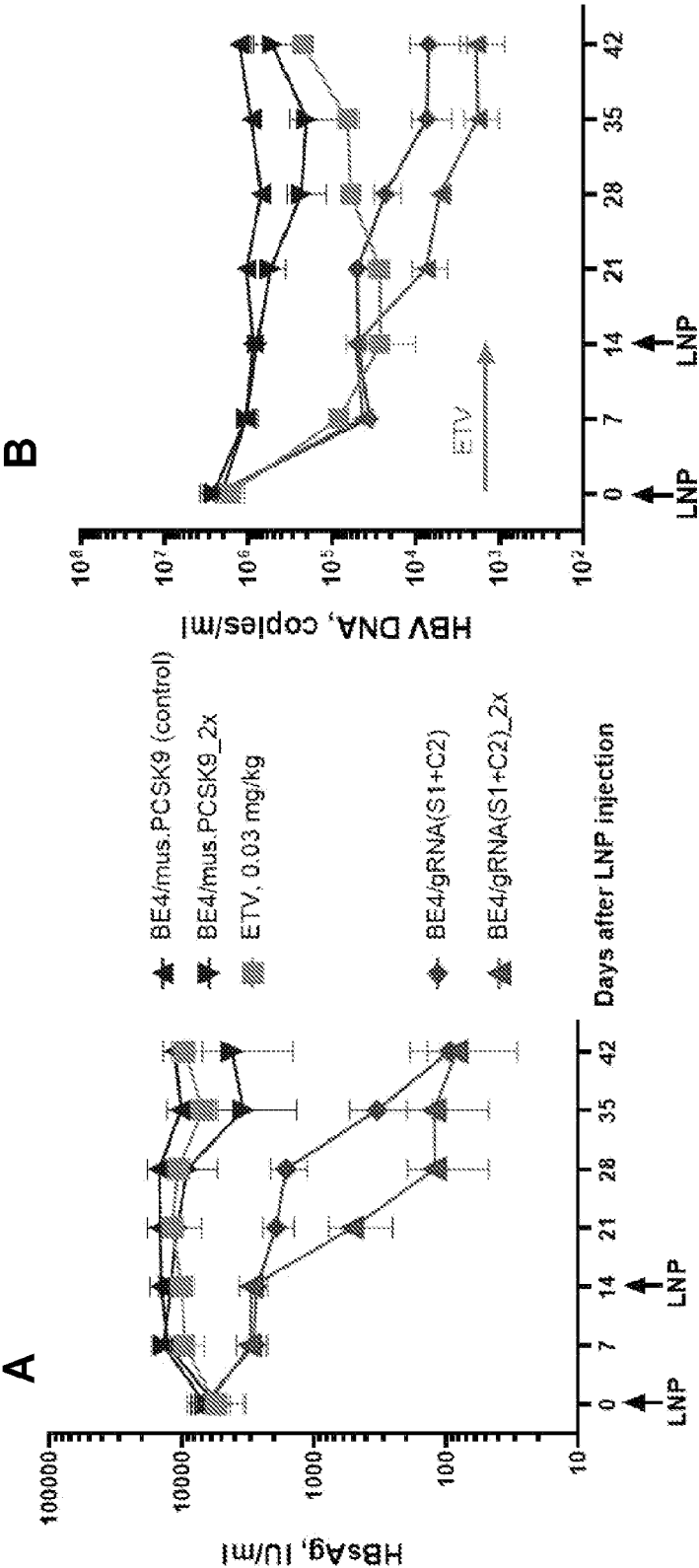


FIG. 33C

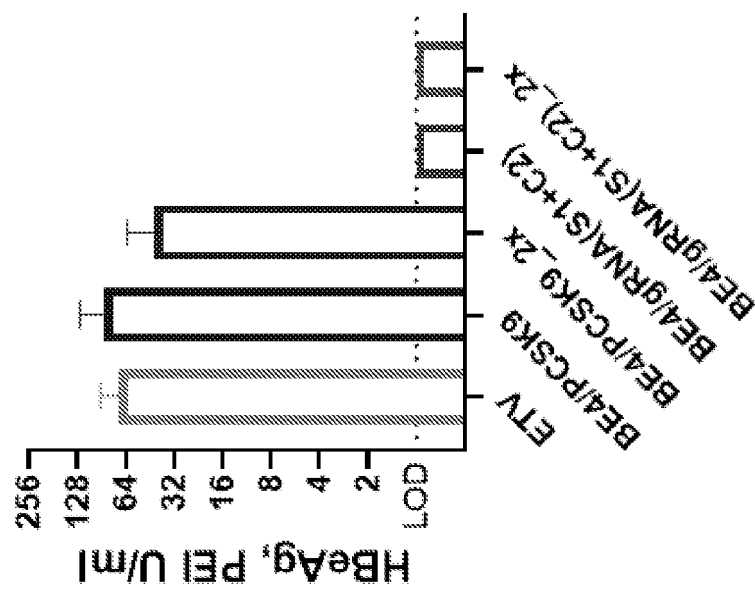
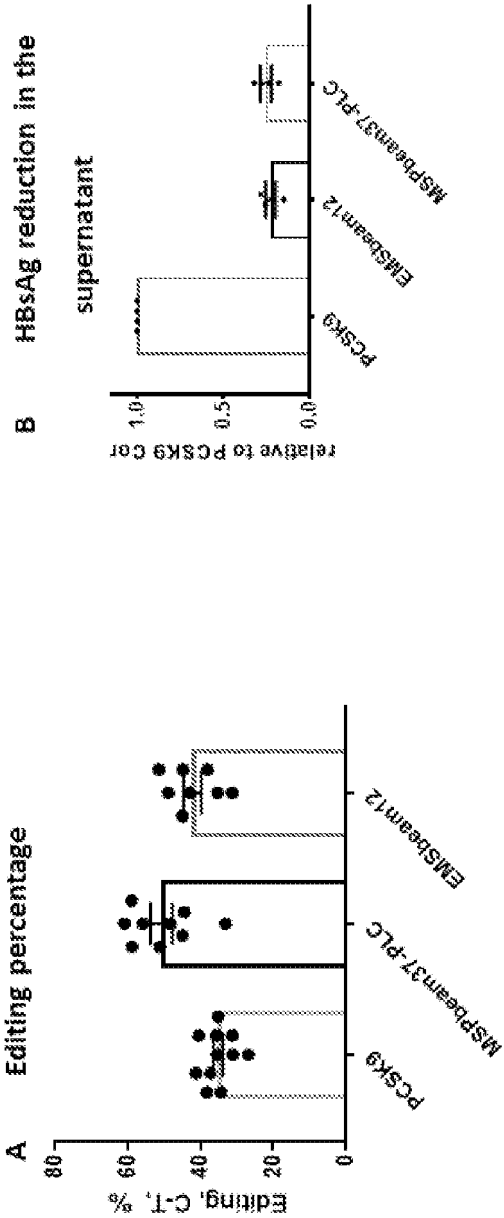


FIG. 34



MSPbeam37 protospacer: 5'-GAAAGCCAGGATGATGGGA-3'

MSPbeam37-PLC protospacer: 5'-GAAAGCCCAAGATGATGGGA-3'

(PLC/PRF/5 compatible gRNA MSPbeam37 sequence)

FIG. 35

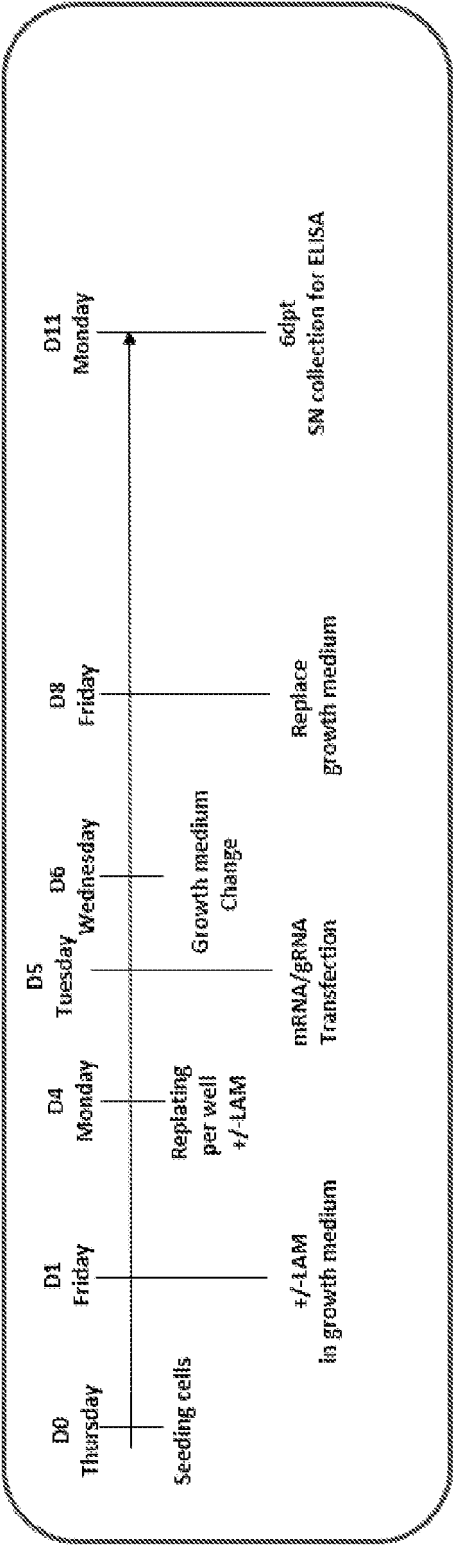
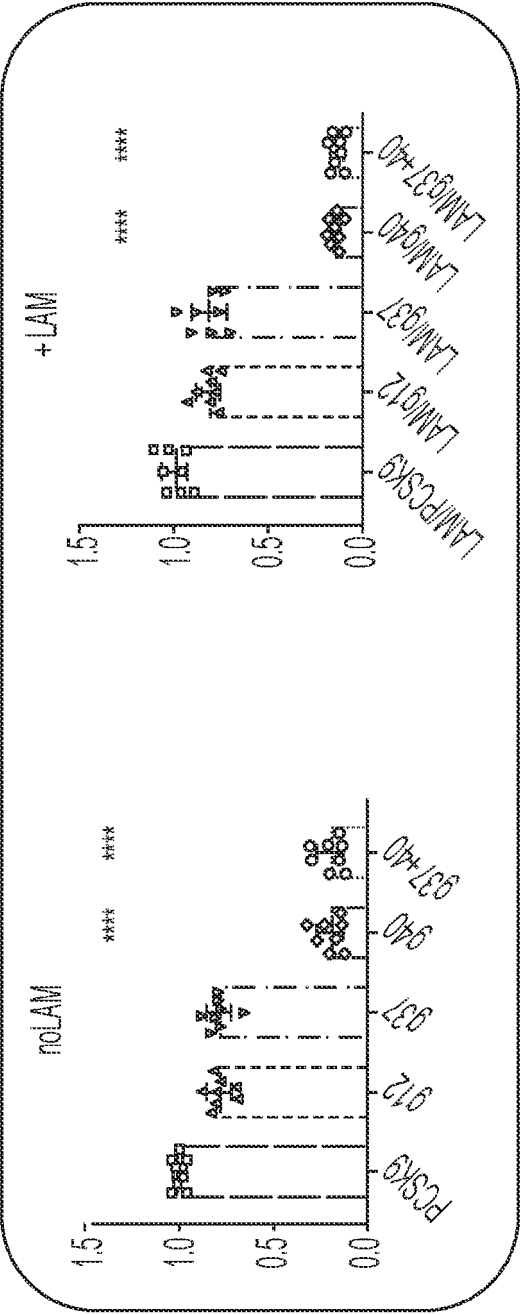


FIG. 36B

ANOVA/non-parametric test / no matching pair / multiple comparisons to PCSK9



n=8
** P ≤ 0.01
*** P ≤ 0.001
**** P ≤ 0.0001