IMPROVED PRIME EDITORS AND METHODS OF USE

GOVERNMENT SUPPORT

[0001] This invention was made with government support under grant numbers R01EB031172, R01EB022376, U01Al142756, RM1HG009490 and R35GM118062 awarded by the National Institutes of Health. The government has certain rights in the invention.

RELATED APPLICATIONS

[0002] This application claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Application U.S.S.N. 63/388,888, filed July 13, 2022, and U.S. Provisional Application U.S.S.N. 63/230,688, filed August 6, 2021, each of which is incorporated herein by reference.

INCORPORATION BY REFERENCE

[0003] This application refers to and incorporates by reference the entire contents of each of the following patent applications directed to prime editing previously filed by one or more of the present inventors: U.S. Provisional Application U.S.S.N. 62/820,813, filed March 19, 2019; U.S. Provisional Application U.S.S.N. 62/858,958, filed June 7, 2019; U.S. Provisional Application U.S.S.N. 62/889,996, filed August 21, 2019; U.S. Provisional Application U.S.S.N. 62/922,654, filed August 21, 2019; U.S. Provisional Application U.S.S.N. 62/913,553, filed October 10, 2019; U.S. Provisional Application U.S.S.N. 62/973,558, filed October 10, 2019; U.S. Provisional Application U.S.S.N. 62/931,195, filed November 5, 2019; U.S. Provisional Application U.S.S.N. 62/944,231, filed December 5, 2019; U.S. Provisional Application U.S.S.N. 62/974,537, filed December 5, 2019; U.S. Provisional Application U.S.S.N. 62/991,069, filed March 17, 2020; U.S. Provisional Application U.S.S.N. 63/100,548, filed March 17, 2020; U.S. Patent Application U.S.S.N. 17/300,668, filed September 17, 2021; International PCT Application No. PCT/US2020/023721, filed March 19, 2020; International PCT Application No. PCT/US2020/023553, filed March 19, 2020; International PCT Application No. PCT/US2020/023583, filed March 19, 2020; U.S. Patent Application U.S.S.N. 17/219,635, filed March 31; International PCT Application No. PCT/US2020/023730, filed March 19, 2020; International PCT Application No. PCT/US2020/023713, filed March 19, 2020; ; U.S. Patent Application U.S.S.N. 17/219,672, filed March 31, 2021; U.S. Patent Application U.S.S.N. 17/751,599, filed May 23, 2022; International PCT Application No. PCT/US2020/023712, filed March 19, 2020; International PCT Application No. PCT/US2020/023727, filed March 19, 2020; International PCT Application No. PCT/US2020/023724, filed March 19, 2020; U.S. Patent Application

U.S.S.N. 17/440,682, filed September 17, 2021; International PCT Application No. PCT/US2020/023725, filed March 19, 2020; International PCT Application No. PCT/US2020/023728, filed March 19, 2020; International PCT Application No. PCT/US2020/023732, filed March 19, 2020; and International PCT Application No. PCT/US2020/023723, filed March 19, 2020.

[0004] This application also refers to and incorporates by reference the entire contents of each of the following patent applications directed to prime editing previously filed by one or more of the present inventors: International PCT Application No. PCT/US2022/012054, filed January 11, 2022, U.S. Provisional Application U.S.S.N. 63/255,897, filed October 14, 2021, U.S. Provisional Application U.S.S.N. 63/231,230, filed August 9, 2021, U.S. Provisional Application U.S.S.N. 63/194,913, filed May 28, 2021, U.S. Provisional Application U.S.S.N. 63/194,865, filed May 28, 2021, U.S. Provisional Application U.S.S.N. 63/176,202, filed April 16, 2021, U.S. Provisional Application U.S.S.N. 63/176,180, filed April 16, 2021, and U.S. Provisional Application U.S.S.N. 63/136,194, filed January 11, 2021.

[0005] This application additionally refers to and incorporates by reference the entire contents of each of the following patent applications directed to prime editing previously filed by one or more of the present inventors: International PCT Application No. PCT/US2021/052097, filed September 24, 2021, U.S. Provisional Application U.S.S.N. 63/231,231, filed August 9, 2021, U.S. Provisional Application U.S.S.N. 63/091,272, filed October 13, 2020, U.S. Provisional Application U.S.S.N. 63/083,067, filed September 24, 2020, and U.S. Provisional Application U.S.S.N. 63/182,633, filed April 30, 2021.

[0006] This application additionally refers to and incorporates by reference the entire contents of each of the following patent applications directed to prime editing previously filed by one or more of the present inventors: International PCT Application No. PCT/US2021/031439, filed May 7, 2021, U.S. Provisional Application No. 63/022,397, filed May 8, 2020, and U.S. Provisional Application No. 63/116,785, filed November 20, 2020.

BACKGROUND OF THE INVENTION

[0007] The recent development of prime editing enables the insertion, deletion, or replacement of genomic DNA sequences without requiring error-prone double-strand DNA breaks. See Anzalone *et al.*, "Search-and-replace genome editing without double-strand breaks or donor DNA," *Nature*, 2019, Vol.576, pp. 149-157, the contents of which are incorporated herein by reference. Prime editing may use an engineered Cas9 nickase–reverse transcriptase fusion protein (e.g., PE1 or PE2) paired with an engineered prime editing guide

RNA (pegRNA) that not only directs Cas9 to a target genomic site, but also which encodes the information for installing the desired edit. Prime editing proceeds through a multi-step editing process: 1) the Cas9 domain binds and nicks the target genomic DNA site, which is specified by the pegRNA's spacer sequence; 2) the reverse transcriptase domain uses the nicked genomic DNA as a primer to initiate the synthesis of an edited DNA strand using an engineered extension on the pegRNA as a template for reverse transcription–this generates a single-stranded 3′ flap containing the edited DNA sequence; 3) cellular DNA repair resolves the 3′ flap intermediate by the displacement of a 5′ flap species that occurs via invasion by the edited 3′ flap, excision of the 5′ flap containing the original DNA sequence, and ligation of the new 3′ flap to incorporate the edited DNA strand, forming a heteroduplex of one edited and one unedited strand; and 4) cellular DNA repair replaces the unedited strand within the heteroduplex using the edited strand as a template for repair, completing the editing process.

[0008] Although prime editing represents a powerful tool for genomic editing, modifications that result in increasing the specificity and efficiency of the prime editing process would help advance the art. In particular, modifications that facilitate more efficient incorporation of the edited DNA strand synthesized by the prime editor into the target genomic site are desirable. It is also desirable to reduce the frequency of indel byproducts that can form as a result of prime editing. Such further modifications to prime editing would advance the art.

SUMMARY OF THE INVENTION

[0009] The present disclosure describes improved prime editor systems, including prime editor fusion proteins, which comprises an engineered Cas9 domain, an engineered reverse transcriptase domain, or a combination of an engineered Cas9 domain and an engineered reverse transcriptase domain. In the case of a prime editor system, the components of the prime editor (i.e., the Cas9 domain and the RT domain) can be provide as individual elements (i.e., uncoupled or unfused). In the case of a prime editor fusion protein, the prime editor components (i.e., the Cas9 domain and the RT domain) are provided as a fusion protein.

[0010] In various embodiments, the engineered Cas9 domain of the herein disclosed prime editor system or fusion protein can comprise a variant Cas9 sequence of SEQ ID NO: 178, SEQ ID NO: 179, or SEQ ID NO: 180, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 178, SEQ ID NO: 179, or SEQ ID NO: 180.

[0011] In various embodiments, the prime editor systems or fusion proteins provided herein may comprise a nucleic acid-programmable DNA-binding protein (napDNAbp) and a mouse mammary tumor virus (MMTV) reverse transcriptase or a variant thereof, an avian sarcoma leukosis virus (ASLV) reverse transcriptase or a variant thereof, a porcine endogenous retrovirus (PERV) reverse transcriptase or a variant thereof, an HIV-MMLV reverse transcriptase or a variant thereof, an AVIRE reverse transcriptase or a variant thereof, a baboon endogenous virus (BAEVM) reverse transcriptase or a variant thereof, a gibbon ape leukemia virus (GALV) reverse transcriptase or a variant thereof, a koala retrovirus (KORV) reverse transcriptase or a variant thereof, a Mason-Pfizer monkey virus (MPMV) reverse transcriptase or a variant thereof, a POK11ERV reverse transcriptase or a variant thereof, a simian retrovirus type 2 (SRV2) reverse transcriptase or a variant thereof, a woolly monkey sarcoma virus (WMSV) reverse transcriptase or a variant thereof, a Vp96 reverse transcriptase or a variant thereof, a Vc95 reverse transcriptase or a variant thereof, an Ec48 reverse transcriptase or a variant thereof, a Gs reverse transcriptase or a variant thereof, an Er reverse transcriptase or a variant thereof, an Ne144 reverse transcriptase or a variant thereof, a Tf1 reverse transcriptase or a variant thereof, or an Rs09415 reverse transcriptase ("CRISPR-RT") or a variant thereof.

[0012] In various other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on MMLV RT wildtype of SEQ ID NO: 33 and can include the variants of SEQ ID NOs: 172-177 or 183-184, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 172-177 or 183-184.

[0013] In still various other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Ec48 RT and can include the variants of SEQ ID NOs: 188-195, 256, and 257 or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 188-195, 256, and 257.

[0014] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Tf1 RT and can include the variants of SEQ ID NOs: 196-213, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 196-213.

[0015] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on PERV RT and can include the variants of SEQ ID NOs: 214-215 or 234-238, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 214-215 or 234-238.

[0016] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on AVIRE RT wildtype (SEQ ID NO: 216) and can include the variants of SEQ ID NOs: 217-221, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 217-221.

[0017] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on KORV RT wildtype (SEQ ID NO: 222) and can include the variants of SEQ ID NOs: 223-227, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 223-227.

[0018] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on WMSV RT wildtype (SEQ ID NO: 228) and can include the variants of SEQ ID NOs: 229-233, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 229-233.

[0019] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Ne144 RT wildtype (SEQ ID NO: 239) and can include the variants of SEQ ID NO: 240, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 240.

[0020] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Vc95 RT

wildtype (SEQ ID NO: 241) and can include the variant of SEQ ID NO: 242, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 242.

[0021] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor systems or fusion proteins can comprise a variant RT sequence based on Gs RT wildtype (SEQ ID NO: 60), or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 159-171.

[0022] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a pentamutant variant RT sequence based on AVIRE RT, KORV RT, and WMSV RT and can include the variants of SEQ ID NOs: 243-245, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 243-245.

[0023] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence of Tfl-rat4 (SEQ ID NO: 251), Tflevo3.1 (SEQ ID NO: 252), Tflevo+rat-1 (SEQ ID NO: 254), Tf1evo+rat2 (SEQ ID NO: 255), Ec48-v2 (SEQ ID NO: 256), Ec48-evo3 (SEQ ID NO: 257) , or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 251-257.

[0024] In other embodiments, the present disclosure describes improved prime editors and prime editor systems, including prime editor fusion proteins, including PEmax of SEQ ID NO: 2, which may be encoded by a nucleic acid sequence of SEQ ID NO: 1, and which may be modified with any one of the herein disclosed variant Cas9 domains or variant RT domains. The present disclosure also provides other improved prime editor variants, including fusion proteins of SEQ ID NOs: 2-8 and fusion proteins comprising evolved nucleic acid programmable DNA binding proteins of SEQ ID NOs: 9-32 and reverse transcriptases of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241. The disclosure also contemplates fusion proteins having an amino acid sequence with a sequence identity of 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 up to 100% with SEQ ID NO: 2

and any one of SEQ ID NOs: 3-8. The disclosure also contemplates evolved nucleic acid programmable DNA binding proteins having an amino acid sequence with a sequence identity of 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 up to 100% with any one of SEQ ID NOs: 9-32. Further, the disclosure contemplates reverse transcriptases having an amino acid sequence with a sequence identity of 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 up to 100% with any one of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241.

[0025] In addition, the instant specification provides for nucleic acid molecules encoding and/or expressing the evolved and/or modified prime editors as described herein, as well as expression vectors or constructs for expressing the evolved and/or modified prime editors described herein, host cells comprising said nucleic acid molecules and expression vectors, and compositions for delivering and/or administering nucleic acid-based embodiments described herein. In addition, the disclosure provides for isolated evolved and/or modified prime editors, as well as compositions comprising said isolated evolved and/or modified prime editors as described herein. Still further, the present disclosure provides for methods of making the evolved and/or modified prime editors, as well as methods of using the evolved and/or modified prime editors or nucleic acid molecules encoding the evolved and/or modified prime editors in applications including editing a nucleic acid molecule, e.g., a genome, with improved efficiency as compared to prime editor that forms the state of the art, preferably in a sequence-context agnostic manner (i.e., wherein the desired editing site does not require a specific sequence-context). In embodiments, the method of making provide herein is an improved phage-assisted continuous evolution (PACE) system which may be utilized to evolve one or more components of a prime editor (e.g., a Cas9 domain or a reverse transcriptase domain). The specification also provides methods for efficiently editing a target nucleic acid molecule, e.g., a single nucleobase of a genome, with a prime editing system described herein (e.g., in the form of an isolated evolved and/or modified prime editor as described herein or a vector or construct encoding same) and conducting prime editing, preferably in a sequence-context agnostic manner. Still further, the specification provides therapeutic methods for treating a genetic disease and/or for altering or changing a genetic trait or condition by contacting a target nucleic acid molecule, e.g., a genome, with a prime editing system (e.g., in the form of an isolated evolved and/or modified prime editor protein or a vector encoding same) and conducting prime editing to treat the genetic disease and/or change the genetic trait (e.g., eye color).

[0026] The inventors have surprisingly found that the editing efficiency of prime editing may be significantly increased (e.g., 2-fold increase, 3-fold increase, 4-fold increase, 5-fold increase, 6-fold increase, 7-fold increase, 8-fold increase, 9-fold increase, or 10-fold increase or more) when one or more components of the canonical prime editor (i.e., PE2) are modified. Modifications may include a modified amino acid sequence of one or more components (e.g., a Cas9 component, a reverse transcriptase component, or a linker).

[0027] The inventors recently developed prime editing which enables the insertion, deletion, or replacement of genomic DNA sequences without requiring error-prone double-strand DNA breaks. Prime editing may use an engineered Cas9 nickase–reverse transcriptase fusion protein (e.g., PE1 or PE2) paired with an engineered prime editing guide RNA (pegRNA) that both directs Cas9 to the target genomic site and encodes the information for installing the desired edit. Prime editing proceeds through a multi-step editing process: 1) the Cas9 domain binds and nicks the target genomic DNA site, which is specified by the pegRNA's spacer sequence; 2) the reverse transcriptase domain uses the nicked genomic DNA as a primer to initiate the synthesis of an edited DNA strand using an engineered extension on the pegRNA as a template for reverse transcription–this generates a single-stranded 3′ flap containing the edited DNA sequence; 3) cellular DNA repair resolves the 3′ flap intermediate by the displacement of a 5' flap species that occurs via invasion by the edited 3′ flap, excision of the 5′ flap containing the original DNA sequence, and ligation of the new 3′ flap to incorporate the edited DNA strand, forming a heteroduplex of one edited and one unedited strand; and 4) cellular DNA repair replaces the unedited strand within the heteroduplex using the edited strand as a template for repair, completing the editing process.

[0028] Efficient incorporation of the desired edit requires that the newly synthesized 3′ flap contains a portion of sequence that is homologous to the genomic DNA site. This homology enables the edited 3′ flap to compete with the endogenous DNA strand (the corresponding 5' flap) for incorporation into the DNA duplex. Because the edited 3' flap will contain less sequence homology than the endogenous 5′ flap, the competition is expected to favor the 5′ flap strand. Thus, a potential limiting factor in the efficiency of prime editing may be the failure of the 3′ flap, which contains the edit, to effectively invade and displace the 5′ flap strand. Moreover, successful 3′ flap invasion and removal of the 5′ flap only incorporates the edit on one strand of the double-stranded DNA genome. Permanent installation of the edit requires cellular DNA repair to replace the unedited complementary DNA strand using the edited strand as a template. While the cell can be made to favor replacement of the unedited strand over the edited strand (step 4 above) by the introduction of a nick in the unedited

strand adjacent to the edit using a secondary sgRNA (i.e., the PE3 system), this process still relies on a second stage of DNA repair.

[0029] The napDNAbp and the polymerase of the prime editor may be joined together to form a fusion protein. In some embodiments, the napDNAbp and the polymerase of the prime editor are joined by a linker to form a fusion protein. In certain embodiments, the linker comprises an amino acid sequence of any one of SEQ ID Nos: 79-93, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 79-93. In some embodiments, the linker is 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, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0030] In other embodiments, the linkers may include in certain embodiments SGGSx2-NLS^{SV40}-SGGSx2, which corresponds to the amino acid sequence SGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGS (SEQ ID NO: 79).

[0031] The components used in the method (*e.g.*, the prime editor, the pegRNA) may be encoded on a DNA vector. In some embodiments, the prime editor, the pegRNA are encoded on one or more DNA vectors. In certain embodiments, the one or more DNA vectors comprise AAV or lentivirus DNA vectors. In some embodiments, the AAV vector is serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

[0032] The prime editors utilized in the presently disclosed methods may also be further joined to additional components. In certain embodiments, the second linker is a self-hydrolyzing linker. In certain embodiments, the second linker comprises an amino acid sequence of any one of SEQ ID Nos: 79-93, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 79-93. In some embodiments, the second linker is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0033] In some embodiments, the one or more modifications to the nucleic acid molecule installed at the target site comprise one or more transitions, one or more transversions, one or more insertions, one or more deletions, or one more inversions. In certain embodiments, the one or more transitions are selected from the group consisting of: (a) T to C; (b) A to G; (c) C to T; and (d) G to A. In certain embodiments, the one or more transversions are selected from the group consisting of: (a) T to A; (b) T to G; (c) C to G; (d) C to A; (e) A to T; (f) A to C; (g) G to C; and (h) G to T. In certain embodiments, the one or more modifications comprises changing (1) a G:C basepair to a T:A basepair, (2) a G:C basepair to an A:T basepair, (3) a

G:C basepair to a C:G basepair, (4) a T:A basepair to a G:C basepair, (5) a T:A basepair to an A:T basepair, (6) a T:A basepair to a C:G basepair, (7) a C:G basepair to a G:C basepair, (8) a C:G basepair to a T:A basepair, (9) a C:G basepair to an A:T basepair, (10) an A:T basepair to a T:A basepair, (11) an A:T basepair to a G:C basepair, or (12) an A:T basepair to a C:G basepair. In some embodiments, the one or more modifications comprises an insertion or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

[0034] The methods of the present disclosure may be used for making corrections to one or more disease-associated genes. In some embodiments, the one or more modifications comprises a correction to a disease-associated gene. In certain embodiments, the disease-associated gene is associated with a polygenic disorder selected from the group consisting of: heart disease; high blood pressure; Alzheimer's disease; arthritis; diabetes; cancer; and obesity. In certain embodiments, the disease-associated gene is associated with a monogenic disorder selected from the group consisting of: Adenosine Deaminase (ADA) Deficiency; Alpha-1 Antitrypsin Deficiency; Cystic Fibrosis; Duchenne Muscular Dystrophy; Galactosemia; Hemochromatosis; Huntington's Disease; Maple Syrup Urine Disease; Marfan Syndrome; Neurofibromatosis Type 1; Pachyonychia Congenita; Phenylkeotnuria; Severe Combined Immunodeficiency; Sickle Cell Disease; Smith-Lemli-Opitz Syndrome; a trinucleotide repeat disorder; a prion disease; and Tay-Sachs Disease.

[0035] In another aspect, the present disclosure provides compositions for editing a nucleic acid molecule by prime editing. In some embodiments, the composition comprises a prime editor, a pegRNA, wherein the composition is capable of installing one or more modifications to the nucleic acid molecule at a target site.

[0036] The composition may increase the efficiency of prime editing and/or decrease the frequency of indel formation. In some embodiments, the prime editing efficiency is increased by at least 1.5-fold, at least 2.0-fold, at least 2.5-fold, at least 3.0-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, at least 6.0-fold, at least 6.5-fold, at least 7.0-fold, at least 7.5-fold, at least 8.0-fold, at least 8.5-fold, at least 9.0-fold, at least 9.5-fold, or at least 10.0-fold as compared to editing with PE2. In some embodiments, the frequency of indel formation is decreased by at least 1.5-fold, at least 2.0-fold, at least 2.5-fold, at least 3.0-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, at least 6.0-fold, at least 6.5-fold, at least 7.0-fold, at least 7.5-fold, at least 8.0-fold, at least 8.5-fold, at least 9.0-fold, at least 9.5-fold, or at least 10.0-fold as compared to editing with PE2.

[0037] The prime editors utilized in the compositions of the present disclosure comprise multiple components. In some embodiments, the prime editor comprises a napDNAbp and a polymerase. In some embodiments, the napDNAbp is a nuclease active Cas9 domain, a nuclease inactive Cas9 domain, or a Cas9 nickase domain or variant thereof. In certain embodiments, the napDNAbp is selected from the group consisting of: Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, and Argonaute and optionally has a nickase activity. In certain embodiments, the napDNAbp comprises an amino acid sequence of any one of SEQ ID Nos: 9-32, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 9-32. In certain embodiments, the napDNAbp comprises an amino acid sequence of SEQ ID NO: 10 (i.e., the napDNAbp of PE1 and PE2) or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with SEQ ID NO: 10. In some embodiments, the polymerase is a DNA-dependent DNA polymerase or an RNA-dependent DNA polymerase. In some embodiments, the polymerase is a reverse transcriptase. In certain embodiments, the reverse transcriptase comprises an amino acid sequence of any one of SEQ ID Nos: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241 or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241.

[0038] The napDNAbp and the polymerase of the prime editor may be joined together to form a fusion protein. In some embodiments, the napDNAbp and the polymerase of the prime editor are joined by a linker to form a fusion protein. In certain embodiments, the linker comprises an amino acid sequence of any one of SEQ ID Nos: 79-93, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 79-93. In some embodiments, the linker is 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, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0039] The components used in the compositions disclosed herein may be encoded on a DNA vector. In some embodiments, the prime editor, the pegRNA, are encoded on one or more DNA vectors. In certain embodiments, the one or more DNA vectors comprise AAV or lentivirus DNA vectors. In some embodiments, the AAV vector is serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

[0040] The prime editors utilized in the presently disclosed compositions may also be further joined to additional components. In some embodiments, the prime editor as a fusion protein is further joined by a second linker. In certain embodiments, the second linker is a self-

hydrolyzing linker. In certain embodiments, the second linker comprises an amino acid sequence of any one of SEQ ID Nos: 79-93, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 79-93. In some embodiments, the second linker is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0041] In some embodiments, the one or more modifications to the nucleic acid molecule installed at the target site comprise one or more transitions, one or more transversions, one or more insertions, one or more deletions, or one more inversions. In certain embodiments, the one or more transitions are selected from the group consisting of: (a) T to C; (b) A to G; (c) C to T; and (d) G to A. In certain embodiments, the one or more transversions are selected from the group consisting of: (a) T to A; (b) T to G; (c) C to G; (d) C to A; (e) A to T; (f) A to C; (g) G to C; and (h) G to T. In certain embodiments, the one or more modifications comprises changing (1) a G:C basepair to a T:A basepair, (2) a G:C basepair to an A:T basepair, (3) a G:C basepair to a C:G basepair, (4) a T:A basepair to a G:C basepair, (5) a T:A basepair to an A:T basepair, (6) a T:A basepair to a C:G basepair, (7) a C:G basepair to a G:C basepair, (8) a C:G basepair to a T:A basepair, (9) a C:G basepair to an A:T basepair, (10) an A:T basepair to a T:A basepair, (11) an A:T basepair to a G:C basepair, or (12) an A:T basepair to a C:G basepair. In some embodiments, the one or more modifications comprises an insertion or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

[0042] The compositions of the present disclosure may be used for making corrections to one or more disease-associated genes. In some embodiments, the one or more modifications comprises a correction to a disease-associated gene. In certain embodiments, the disease-associated gene is associated with a polygenic disorder selected from the group consisting of: heart disease; high blood pressure; Alzheimer's disease; arthritis; diabetes; cancer; and obesity. In certain embodiments, the disease-associated gene is associated with a monogenic disorder selected from the group consisting of: Adenosine Deaminase (ADA) Deficiency; Alpha-1 Antitrypsin Deficiency; Cystic Fibrosis; Duchenne Muscular Dystrophy; Galactosemia; Hemochromatosis; Huntington's Disease; Maple Syrup Urine Disease; Marfan Syndrome; Neurofibromatosis Type 1; Pachyonychia Congenita; Phenylkeotnuria; Severe Combined Immunodeficiency; Sickle Cell Disease; Smith-Lemli-Opitz Syndrome; a trinucleotide repeat disorder; a prion disease; and Tay-Sachs Disease.

[0043] In another aspect, this disclosure provides polynucleotides for editing a DNA target site by prime editing. In some embodiments, the polynucleotide comprises a nucleic acid sequence encoding a napDNAbp, a polymerase, wherein the napDNAbp and polymerase is capable in the presence of a pegRNA of installing one or more modifications in the DNA target site.

[0044] The prime editors utilized in the polynucleotides of the present disclosure comprise multiple components (*e.g.*, a napDNAbp and a polymerase). In some embodiments, the napDNAbp is a nuclease active Cas9 domain, a nuclease inactive Cas9 domain, or a Cas9 nickase domain or variant thereof. In certain embodiments, the napDNAbp is selected from the group consisting of: Cas9, Cas12e, Cas12d, Cas12a, Cas12b1, Cas13a, Cas12c, and Argonaute and optionally has a nickase activity. In certain embodiments, the napDNAbp comprises an amino acid sequence of any one of SEQ ID Nos: 2-8, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 2-8. In certain embodiments, the napDNAbp comprises an amino acid sequence of SEQ ID NO: 10 (i.e., the napDNAbp of PE1 and PE2) or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with SEQ ID NO: 10. In some embodiments, the polymerase is a DNA-dependent DNA polymerase or an RNA-dependent DNA polymerase. In some embodiments, the polymerase is a reverse transcriptase. In certain embodiments, the reverse transcriptase comprises an amino acid sequence of any one of SEQ ID Nos: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241 or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241.

[0045] The napDNAbp and the polymerase of the prime editor may be joined together to form a fusion protein. In some embodiments, the napDNAbp and the polymerase of the prime editor are joined by a linker to form a fusion protein. In certain embodiments, the linker comprises an amino acid sequence of any one of SEQ ID Nos: 9-32, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 9-32. In some embodiments, the linker is 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, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0046] The polynucleotides disclosed herein may comprise vectors. In some embodiments, the polynucleotide is a DNA vector. In certain embodiments, the DNA vector is an AAV or

lentivirus DNA vector. In some embodiments, the AAV vector is serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

[0047] The prime editors encoded by the presently disclosed polynucleotides may also be further joined to additional components. In certain embodiments, the second linker comprises a self-hydrolyzing linker. In certain embodiments, the second linker comprises an amino acid sequence of any one of SEQ ID Nos: 79-93, or an amino acid sequence having at least an 80%, 85%, 90%, 95%, or 99% sequence identity with any one of SEQ ID Nos: 79-93. In some embodiments, the second linker is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 amino acids in length.

[0048] In some embodiments, the one or more modifications to the nucleic acid molecule installed at the target site comprise one or more transitions, one or more transversions, one or more insertions, one or more deletions, or one more inversions. In certain embodiments, the one or more transitions are selected from the group consisting of: (a) T to C; (b) A to G; (c) C to T; and (d) G to A. In certain embodiments, the one or more transversions are selected from the group consisting of: (a) T to A; (b) T to G; (c) C to G; (d) C to A; (e) A to T; (f) A to C; (g) G to C; and (h) G to T. In certain embodiments, the one or more modifications comprises changing (1) a G:C basepair to a T:A basepair, (2) a G:C basepair to an A:T basepair, (3) a G:C basepair to a C:G basepair, (4) a T:A basepair to a G:C basepair, (5) a T:A basepair to an A:T basepair, (6) a T:A basepair to a C:G basepair, (7) a C:G basepair to a G:C basepair, (8) a C:G basepair to a T:A basepair, (9) a C:G basepair to an A:T basepair, (10) an A:T basepair to a T:A basepair, (11) an A:T basepair to a G:C basepair, or (12) an A:T basepair to a C:G basepair. In some embodiments, the one or more modifications comprises an insertion or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

[0049] The polynucleotides of the present disclosure may be used for making corrections to one or more disease-associated genes. In some embodiments, the one or more modifications comprises a correction to a disease-associated gene. In certain embodiments, the disease-associated gene is associated with a polygenic disorder selected from the group consisting of: heart disease; high blood pressure; Alzheimer's disease; arthritis; diabetes; cancer; and obesity. In certain embodiments, the disease-associated gene is associated with a monogenic disorder selected from the group consisting of: Adenosine Deaminase (ADA) Deficiency; Alpha-1 Antitrypsin Deficiency; Cystic Fibrosis; Duchenne Muscular Dystrophy; Galactosemia; Hemochromatosis; Huntington's Disease; Maple Syrup Urine Disease; Marfan

Syndrome; Neurofibromatosis Type 1; Pachyonychia Congenita; Phenylkeotnuria; Severe Combined Immunodeficiency; Sickle Cell Disease; Smith-Lemli-Opitz Syndrome; a trinucleotide repeat disorder; a prion disease; and Tay-Sachs Disease.

[0050] In another aspect, the present disclosure provides cells. In some embodiments, the cell comprises any of the polynucleotides described herein.

[0051] In another aspect, the present disclosure provides pharmaceutical compositions. In some embodiments, the pharmaceutical composition comprises any of the compositions disclosed herein. In some embodiments, the pharmaceutical composition comprises any of the compositions disclosed herein and a pharmaceutically acceptable excipient. In some embodiments, the pharmaceutical composition comprises any of the polynucleotides disclosed herein. In some embodiments, the pharmaceutical composition comprises any of the polynucleotides disclosed herein and a pharmaceutically acceptable excipient.

[0052] In another aspect, the present disclosure provides kits. In some embodiments, the kit comprises any of the compositions disclosed herein, a pharmaceutical excipient, and instructions for editing a DNA target site by prime editing. In some embodiments, the kit comprises any of the polynucleotides disclosed herein, a pharmaceutical excipient, and instructions for editing a DNA target site by prime editing.

BRIEF DESCRIPTION OF THE DRAWINGS

[0053] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present disclosure, which can be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0054] FIG. 1 provides a schematic showing the optimization of PE2 protein. SEQ ID NO: 80 is shown.

[0055] FIG. 2 shows the fold change in the frequency of the intended edit using PE2 and various other PE constructs in HEK293T cells (low plasmid dose) at a range of gene targets (*HEK3, EMX1, RNF2, FANCF, FUNX1, DNMT1, VEGFA, HEK4, PRNP, APOE, CXCR4, HEK3*).

[0056] FIG. 3 shows the fold change in the frequency of the intended edit using PE3 and various prime editor constructs in HeLa cells at a range of gene targets (*HEK3, FANCF, RUNX1, VEGFA*).

[0057] FIG. 4 shows a comparison of prime editing in HEK293T vs. HeLa editing using various PE constructs.

[0058] FIG. 5 shows NLS architecture optimization of PE3 in HeLa cells.

[0059] FIG. 6 provides a schematic showing the final PEmax construct, which corresponds to SEQ ID NO: 2.

[0060] FIG. 7 shows that PEmax increases indels in addition to the intended edit.

[0061] FIGs. 8A-8C show the development of PEmax. FIGs. 8A and 8B show screening of prime editor variants to maximize editing efficiency in HeLa cells. All PE architectures carry a Cas9 H840A mutation. NLS^{SV40} indicates the bipartite SV40 NLS. $*NLS^{SV40}$ contains a 1-aa deletion outside the PKKKRKV (SEQ ID NO: 94) NLSSV40 consensus sequence. All individual values of $n = 3$ independent biological replicates are shown. FIG. 8C shows a comparison of PE3max (PE3 editing system with PEmax protein) and PE3 (PE3 editing system with PE2 protein) in HeLa cells (mean of $n = 3$ independent biological replicates).

[0062] FIG. 9 shows that PEmax architecture enhances editing at disease-relevant gene targets and cell types. FIG. 9 provides a schematic of PE2 and PEmax editor architectures. $bpNLS^{SV40}$, bipartite SV40 NLS. MMLV RT, Moloney Murine Leukemia Virus reverse transcriptase pentamutant. GS codon, Genscript human codon optimized.

[0063] FIG. 10 provides a schematic of the prime editor phage-assisted continuous evolution (PACE) circuit. The PACE circuit is useful for disease-specific evolutions, evolution of different prime editor domains, and whole-editor evolutions.

[0064] FIG. 11 shows the editing efficiency of evolved Gs mutants in HEK293T cells.

[0065] FIG. 12 shows the editing efficiency of evolved PE2 reverse transcriptase (RT) mutants in HEK293T cells at low dose (75 ng editor). The evolved mutants result in outsized benefit at low doses.

[0066] FIG. 13 provides a schematic of the PACE circuit for Cas9 and reverse transcriptase evolution.

[0067] FIG. 14 shows the editing efficiency of Cas9 mutant prime editors in HEK293T cells.

[0068] FIG. 15 shows the editing efficiency of evolved prime editor mutants in N2A cells.

[0069] FIG. 16 shows that unique reverse transcriptase enzymes show detectable prime editing activity at the *RNF2* and *HEK3* sites in HEK293T cells. M-MLV* is the engineered pentamutant variant of the M-MLV RT.

[0070] FIG. 17 shows that retroviral reverse transcriptases exhibit prime editing activity. Unique retroviral reverse transcriptase (RT) enzymes exhibit prime editing activity in HEK293T cells in the *FANCF* and *HEK3* loci. MMTV, PERV, AVIRE, KORV, and WMSV perform better than the wild-type (WT) M-MLV enzyme.

[0071] FIG. 18 shows a comparison of the PERV pentamutant and PE2. A pentamutant, engineered version of the PERV retroviral RT (21.6) shows improved performance over the WT enzyme. 21.6 has comparable editing to the pentamutant, engineered version of M-MLV RT (PE2) for FANCF +5 G to T, HEK3 +1 His ins and HEK3 +1 FLAG ins edits but lower editing for VEGFA +2 G to A, RNF2 +1 C to A, EMX1 +5 G to T, and DNMT1 1-15 deletion edits.

[0072] FIG. 19 shows that the yeast retrotransposon RT enzyme, Tf1 RT, exhibits prime editing activity in HEK293T cells. A yeast retrotransposon RT enzyme, Tf1, exhibits prime editing activity in HEK293T cells. Tf1 has higher editing than the WT M-MLV reverse transcriptase but lower activity than the pentamutant engineered enzyme (PE2).

[0073] FIG. 20 shows that mutants S297Q and K118R improve editing activity. A structure-guided rationally designed variant of Tf1 (with S297Q and K118R mutations) shows improved editing over the WT enzyme. The double mutant is 1.3-4.2 fold better than the WT enzymes at the four sites tested. PE2 outperforms the rationally designed mutant. Increasing contacts of the RT with the RNA-DNA substrate improves PE outcomes.

[0074] FIG. 21 shows editing efficiencies of Tf1 20 bp PANCE mutants in HEK293T cells. Tf1 variants (evolved using PANCE) 5.27, 5.59, and 5.60 show improved editing compared with the WT enzyme Tf1 variant in HEK293T cells. Variants 5.59 and 5.60 have comparable editing to PE2 in the sites tested.

[0075] FIG. 22 shows editing efficiencies of evolved Tf1 mutants in N2a cells. Editing using Tf1 variants (evolved using PACE or PANCE) 5.27, 5.47, 5.59, and 5.60 in mouse Neuro2a cells is shown. WT and evolved Tf1 variants (5.47 and 5.60) exhibit higher editing than PE2 at the *Dnmt1* locus.

[0076] FIG. 23 shows that unique small bacterial reverse transcriptase enzymes exhibit prime editing activity in HEK293T cells.

[0077] FIG. 24 shows editing efficiencies of Ec48 20 bp PANCE mutants in HEK293T cells. Ec48 variants (evolved using PANCE) 3.8, 3.35, 3.36, and 3.38 show improved editing compared with the WT Ec48 enzyme in HEK293T cells.

[0078] FIG. 25 shows editing efficiencies of evolved Ec48 mutants in N2a cells. Ec48 variants (evolved using PACE or PANCE) 3.8, 3.23, 3.35, 3.36, 3.37, and 3.38 were used in mouse Neuro2a cells. Evolved Ec48 variants exhibit comparable editing to PE2 at the *Dnmt1* locus.

[0079] FIG. 26 provides the structural components of PEmax from the N-terminal to C-terminal direction.

[0080] FIG. 27A illustrates strategies for improving prime editors, e.g., PE2, which includes (a) PACE-evolving of the Cas9 domain, (b) PACE-evolving of the RT domain, and (c) replacement of RT domain with alternate RT domains.

[0081] FIG. 27B provides a list of prime editor embodiments disclosed herein comprising a PACE-evolved Cas9 domain and an MMLV domain or variant thereof. The amino acid substitutions (e.g., "T128N") refer to the amino acid positions of the wild type MMLV protein of SEQ ID NO: 33.

[0082] FIG. 28 provides a list of alternate reverse transcriptase domains described herein in Example 2 that can be used in place of MMLV domain of PE2 or in another prime editor.

[0083] FIG. 29 shows the incorporation of PE2 mutations into retroviral RTs AVIRE, KORV, WMSV and PERV improve average prime editing activity compared to the WT enzyme at 4 different loci in HEK293T cells.

[0084] FIG. 30 shows the incorporation of all 5 mutations into PERV-RT improves activity 6.6-fold compared to the WT enzyme across 9 different edits in HEK293T cells. (21.6 mutations are D199N, T305K, W312F, E329P, L602W).

[0085] FIG. 31A-31D shows the creation and validation of a PE-PACE Circuit of FIG. 10. FIG. 31A shows initial overnight propagation of PE2 RT phage in circuit. FIG. 31B shows overnight propagation screening of pegRNAs. FIG. 31C shows overnight propagation of PE1 and PE2 in a circuit with an optimized pegRNA. FIG. 31D shows PANCE selection of PE1 RT phage. Rounds shaded in green are drifts, in which no selective pressure was applied.

[0086] FIG. 32 provides a summary of the mutations in M-MLV RT introduced by PANCE of PE1.

[0087] FIG. 33A-33B Modified PE-PACE Circuits. FIG. 33A shows phage propagation decreases as the expression of T7 RNAP is decreased, either via RBS or promoter. This increases stringency. FIG. 33B shows pegRNA optimization for a 20-bp insertion PE-PACE circuit. Numbers on the x axis indicate different pegRNAs.

[0088] FIG. 34 bar graphs showing that evolved variants of Tf1 (evolved using PANCE), 5.27, 5.59 and 5.60 show improved editing compared with the WT enzyme Tf1 variant in HEK293T cells. Variants 5.59 and 5.60 have comparable editing to PE2 in the sites tested above.

[0089] FIG. 35 shows the editing activity of seven (7) unique small bacterial RT enzymes exhibit activity in HEK293T cells.

[0090] FIG. 36 Evolved variant 38.14 is on average 23-fold better than the WT enzyme across 4 loci in HEK293T cells.

[0091] FIG. 37 Vc95 variant (L11M+S75A+V97M+N146D+N245T) is on average 7-fold better than the WT enzyme across 4 loci.

[0092] FIG. 38A-38B Evolution of Gs RT. Mammalian prime editing in HEK293T cells for Gs RT mutants derived from (A) PANCE or (B) PACE.

[0093] FIG. 39 PE-PACE Evolution of Cas9. The bar graph compares the editing efficiency of PE2 in HEK293T cells versus three evolved prime editors using the PE-PACE system of FIG. 13. The evolved editors comprise modifications to the Cas9 (H840A) component of PE2.

[0094] FIG. 40 shows structural-guided engineering of Tf1 reverse transcriptase wherein variants I260L, E274R, R288Q and Q293K showed improved editing over WT in HEK293T cells.

[0095] FIG. 41 shows structural-guided engineering of 28 Tf1 reverse transcriptase mutants wherein variants K118R, S188K, I64L, I64W, N316Q, K321R, L133N showed improved editing over WT in HEK293T cells.

[0096] FIG. 42 shows the editing capabilities of rationally designed Tf1 variants comprising mutation combinations (5.19 = wildtype Tf1 + K118R + S297Q; 5.618 = K118R + S297Q + S188K + I64L + I260L + R288Q; 5.59 = E22K + P70T + G72V + M102I + K106R + A139T + L158Q + F269L + A363V + K413E + S492N) wherein variant 5.618 exhibited comparable editing to the best evolved variant 5.59 in HEK293T cells.

[0097] FIG. 43 shows the editing capabilities of Tf1 variants comprising mutation combinations (5.59 = E22K + P70T + G72V + M102I + K106R + A139T + L158Q + F269L + A363V + K413E + S492N; 5.618 = K118R + S297Q + S188K + I64L + I260L + R288Q; 5.612 = 5.59 + K118R + S297Q + S188K + I64L + I260L) derived from rational design and evolution approaches wherein variant 5.59 further improved activity in HEK293T cells and Tf1 variant 5.612 showed improved activity over PE2.

[0098] FIGs. 44A-44B show an exemplary evolution approach that yielded Ec48 reverse transcriptase variants. FIG. 44A shows the genotype of Ec48 after selection using PANCE on a higher stringency strain. FIG. 44B shows the use of a more stringent promoter called ProB which comprises the Syn 4.0 regulatory sequence combined with 20bp deletion that was used instead of ProD which comprises the sd8 regulatory sequence and a 20bp deletion.

[0099] FIG. 45 shows the editing capabilities of Ec48 mutants in HEK293T cells wherein variants 3.500 (E60K + K87E + E165D + D243N + R267I + E279K + K318E + K343N) and 3.501 (E60K + K87E + S151T + E165D + D243N + R267I + E279K + V303M + K318E +

K343N) outperformed previously characterized best evolved variant 3.35 (E54K + K87E + D243N + R267I + E279K + K318E).

[0100] FIG. 46 shows improved editing efficiency of Tf1-based prime editor using five mutations (K118R, S188K, I260L, S297Q, and R288Q) predicted via structure-guided engineering.

[0101] FIG. 47 shows improved editing of Tf1-based prime editor when combining mutations to generate the rat1 (K118R + S188K), rat2 (K118R + S188K + I260L), rat3 (K118R + S188K + I260L + S297Q), and rat4(K118R + S188K + I260L + S297Q + R288Q) variants.

[0102] FIG. 48 shows improved editing of the Tf1-based prime editor using the Tf1evo3.1 and Tf1evo3.2 variants.

[0103] FIG. 49 Combining rational mutations into best evolved variants slightly improves editing on average at particular sites.

[0104] FIGs. 50A-50B show improved editing efficiency of Ec48-based prime editor using five mutations predicted via structure-guided engineering. **FIG. 50A** shows editing efficiency of the T189N EC48 mutant. **FIG. 50B** shows editing efficiency of the R378K, K307R, T385R, L182N, and R315K mutants.

[0105] FIG. 51 shows improved editing efficiency of Ec48-based prime editor when combining mutations to generate the Ec48-v2 (R315K + L182N + T189N) variant.

[0106] FIG. 52 shows the Ec48-evo3 variant exhibits further improvements in editing efficiency.

[0107] FIG. 53 shows the editing efficiency represented as editing percent at the indicated target genes of Tf1 and Ec48 variants in the PEmax architecture**.**

[0108] FIG. 54 shows a summary of improvements on short RTT edits performed in N2A cells by the indicated M-MLV mutants.

[0109] FIGs. 55A-55B show a summary of improvements on long RTT edits by the indicated M-MLV mutants. **FIG.55A** shows improvements relative to full-length PE2max in HEK293T cells. **FIG. 55B** shows improvements relative to truncated PE2max in HEK293T cells.

[0110] FIG. 56 shows additional PACE and PANCE-evolved and engineered Cas9 mutants that improve mammalian prime editing in N2A cells.

[0111] FIGs. 57A-57C show a Tay-Sachs disease circuit. FIG. 57A shows a circuit setup, demonstrating where in T7 RNAP the pathogenic fragment is inserted. FIG. 57B shows the

sequence of the mutation-containing T7 region before prime editing. FIG 57C shows the resulting sequencing after prime editing, in which the correct frame is restored.

[0112] FIGs. 58A-58B show the editing efficiency represented as editing percent of Ec48 and Gs variants. FIG. 58A shows the editing efficiency of the Ec48-3.35, Ec48-3.500, and Ec48-TSD1 variants. FIG. 58B shows the editing efficiency of the Gs811, Gs813, Gs814, Gs815, Gs816, Gs-TSD1, Gs-TSD2, and Gs-TSD3 variants.

[0113] FIG. 59. Shows improved editing capabilities of penta-mutant versions of each retroviral RT enzyme over individual mutants. For the AVIRE RT, KORV RT and WMSV RT, the five mutations that improved editing were combined which resulted in an additive effect in editing efficiency. The final variants PERV_penta, AVIRE_penta, KORV_penta and WMSV_penta demonstrated approximately 4-fold to 7-fold improvements in editing efficiency on average across 5 edits.

DEFINITIONS

[0114] 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 unless specified otherwise.

Cas9

[0115] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 domain, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A "Cas9 domain" as used herein, is a protein fragment comprising an active or inactive cleavage domain of Cas9 and/or the gRNA binding domain of Cas9. A "Cas9 protein" is a full length Cas9 protein. A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements, and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II

CRISPR systems, correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc), and a Cas9 domain. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves a 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 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 are hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease comprises one or more mutations that partially impair or inactivate the DNA cleavage domain.

[0116] A nuclease-inactivated Cas9 domain may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 domain (or a fragment thereof) having an inactive DNA cleavage domain are known (see, *e.g.*, Jinek *et al.*, *Science*.

337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell.* 28;152(5):1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, at least about 99.8% identical, or at least about 99.9% identical to wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 9). In some embodiments, the Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more amino acid changes compared to wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 9). In some embodiments, the Cas9 variant comprises a fragment of SEQ ID NO: 9 Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 9). In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 9).

[0117] The wild type canonical *Streptococcus pyogenes* Cas9 (SpCas9) sequence reference herein has the following amino acid sequence:

Description	Sequence	SEQ ID NO:
SpCas9 *Streptococcus pyogenes* M1 SwissProt Accession No. Q99ZW2 Wild type	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9

CRISPR

[0118] CRISPR is a family of DNA sequences (i.e., CRISPR clusters) in bacteria and archaea that represent snippets of prior infections by a virus that have invaded the prokaryote. The snippets of DNA are used by the prokaryotic cell to detect and destroy DNA from subsequent attacks by similar viruses and effectively compose, along with an array of CRISPR-associated proteins (including Cas9 and homologs thereof) and CRISPR-associated RNA, a prokaryotic immune defense system. In nature, CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In certain types of CRISPR systems (e.g., type II

CRISPR systems), correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves a linear or circular dsDNA target complementary to the RNA. Specifically, 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 to incorporate aspects of both the crRNA and tracrRNA into a single RNA species – the guide RNA. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. CRISPR biology, as well as Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor Rnase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference.

[0119] In certain types of CRISPR systems (e.g., type II CRISPR systems), correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc), and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically

cleaves a linear or circular nucleic acid target complementary to the RNA. Specifically, the target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered to incorporate embodiments of both the crRNA and tracrRNA into a single RNA species—the guide RNA.

[0120] In general, a "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. The tracrRNA of the system is complementary (fully or partially) to the tracr mate sequence present on the guide RNA.

DNA synthesis template

[0121] As used herein, the term "DNA synthesis template" refers to the region or portion of the extension arm of a PEgRNA that is utilized as a template strand by a polymerase of a prime editor to encode a 3′ single-strand DNA flap that contains the desired edit and which then, through the mechanism of prime editing, replaces the corresponding endogenous strand of DNA at the target site. The extension arm, including the DNA synthesis template, may be comprised of DNA or RNA. In the case of RNA, the polymerase of the prime editor can be an RNA-dependent DNA polymerase (e.g., a reverse transcriptase). In the case of DNA, the polymerase of the prime editor can be a DNA-dependent DNA polymerase. In various embodiments the DNA synthesis template may comprise the "edit template" and the "homology arm", and all or a portion of the optional 5′ end modifier region, e2. That is, depending on the nature of the e2 region (e.g., whether it includes a hairpin, toeloop, or stem/loop secondary structure), the polymerase may encode none, some, or all of the e2 region as well. Said another way, in the case of a 3′ extension arm, the DNA synthesis template can include the portion of the extension arm that spans from the 5′ end of the primer binding site (PBS) to 3′ end of the gRNA core that may operate as a template for the synthesis of a single-strand of DNA by a polymerase (e.g., a reverse transcriptase). In the case of a 5′ extension arm, the DNA synthesis template can include the portion of the

extension arm that spans from the 5′ end of the PEgRNA molecule to the 3′ end of the edit template. Preferably, the DNA synthesis template excludes the primer binding site (PBS) of PEgRNAs either having a 3′ extension arm or a 5′ extension arm. Certain embodiments described here refer to an "an RT template," which is inclusive of the edit template and the homology arm, i.e., the sequence of the PEgRNA extension arm which is actually used as a template during DNA synthesis. The term "RT template" is equivalent to the term "DNA synthesis template."

Edit template

[0122] The term "edit template" refers to a portion of the extension arm that encodes the desired edit in the single strand 3′ DNA flap that is synthesized by the polymerase, e.g., a DNA-dependent DNA polymerase, RNA-dependent DNA polymerase (e.g., a reverse transcriptase). Certain embodiments described here refer to "an RT template," which refers to both the edit template and the homology arm together, i.e., the sequence of the PEgRNA extension arm which is actually used as a template during DNA synthesis. The term "RT edit template" is also equivalent to the term "DNA synthesis template," but wherein the RT edit template reflects the use of a prime editor having a polymerase that is a reverse transcriptase, and wherein the DNA synthesis template reflects more broadly the use of a prime editor having any polymerase.

Extension arm

[0123] The term "extension arm" refers to a nucleotide sequence component of a PEgRNA which provides several functions, including a primer binding site and an edit template for reverse transcriptase. In some embodiments, the extension arm is located at the 3′ end of the guide RNA. In other embodiments, the extension arm is located at the 5′ end of the guide RNA. In some embodiments, the extension arm also includes a homology arm. In various embodiments, the extension arm comprises the following components in a 5′ to 3′ direction: the homology arm, the edit template, and the primer binding site. Since polymerization activity of the reverse transcriptase is in the 5′ to 3′ direction, the preferred arrangement of the homology arm, edit template, and primer binding site is in the 5′ to 3′ direction such that the reverse transcriptase, once primed by an annealed primer sequence, polymerizes a single strand of DNA using the edit template as a complementary template strand. Further details, such as the length of the extension arm, are described elsewhere herein.

[0124] The extension arm may also be described as comprising generally two regions: a primer binding site (PBS) and a DNA synthesis template, for instance. The primer binding

site binds to the primer sequence that is formed from the endogenous DNA strand of the target site when it becomes nicked by the prime editor complex, thereby exposing a 3′ end on the endogenous nicked strand. As explained herein, the binding of the primer sequence to the primer binding site on the extension arm of the PEgRNA creates a duplex region with an exposed 3′ end (i.e., the 3′ of the primer sequence), which then provides a substrate for a polymerase to begin polymerizing a single strand of DNA from the exposed 3′ end along the length of the DNA synthesis template. The sequence of the single strand DNA product is the complement of the DNA synthesis template. Polymerization continues towards the 5′ of the DNA synthesis template (or extension arm) until polymerization terminates. Thus, the DNA synthesis template represents the portion of the extension arm that is encoded into a single strand DNA product (i.e., the 3′ single strand DNA flap containing the desired genetic edit information) by the polymerase of the prime editor complex and which ultimately replaces the corresponding endogenous DNA strand of the target site that sits immediately downstream of the PE-induced nick site. Without being bound by theory, polymerization of the DNA synthesis template continues towards the 5′ end of the extension arm until a termination event. Polymerization may terminate in a variety of ways, including, but not limited to (a) reaching a 5′ terminus of the PEgRNA (e.g., in the case of the 5′ extension arm wherein the DNA polymerase simply runs out of template), (b) reaching an impassable RNA secondary structure (e.g., hairpin or stem/loop), or (c) reaching a replication termination signal, e.g., a specific nucleotide sequence that blocks or inhibits the polymerase, or a nucleic acid topological signal, such as, supercoiled DNA or RNA.

Fusion protein

[0125] The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an “amino-terminal fusion protein” or a “carboxy-terminal fusion protein,” respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a nucleic-acid editing protein. Another example includes a Cas9 or equivalent thereof to a reverse transcriptase. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins

comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

Guide RNA ("gRNA")

[0126] As used herein, the term "guide RNA" is a particular type of guide nucleic acid which is mostly commonly associated with a Cas protein of a CRISPR-Cas9 and which associates with Cas9, directing the Cas9 protein to a specific sequence in a DNA molecule that includes complementarity to the protospacer sequence of the guide RNA. However, this term also embraces the equivalent guide nucleic acid molecules that associate with Cas9 equivalents, homologs, orthologs, or paralogs, whether naturally occurring or non-naturally occurring (e.g., engineered or recombinant), and which otherwise program the Cas9 equivalent to localize to a specific target nucleotide sequence. The Cas9 equivalents may include other napDNAbp from any type of CRISPR system (e.g., type II, V, VI), including Cpf1 (a type-V CRISPR-Cas systems), C2c1 (a type V CRISPR-Cas system), C2c2 (a type VI CRISPR-Cas system) and C2c3 (a type V CRISPR-Cas system). Further Cas-equivalents are described in Makarova *et al.*, "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," Science 2016; 353(6299), the contents of which are incorporated herein by reference. Exemplary sequences are and structures of guide RNAs are provided herein. In addition, methods for designing appropriate guide RNA sequences are provided herein. As used herein, the "guide RNA" may also be referred to as a "traditional guide RNA" to contrast it with the modified forms of guide RNA termed "prime editing guide RNAs" (or "PEgRNAs").

[0127] Guide RNAs or PEgRNAs may comprise various structural elements that include, but are not limited to:

[0128] Spacer sequence – the sequence in the guide RNA or PEgRNA (having about 20 nts in length) which has the same sequence as the protospacer in the target DNA.

[0129] gRNA core (or gRNA scaffold or backbone sequence) – refers to the sequence within the gRNA that is responsible for Cas9 binding, it does not include the 20 bp spacer/targeting sequence that is used to guide Cas9 to target DNA.

[0130] Extension arm – a single strand extension at the 3′ end or the 5′ end of the PEgRNA which comprises a primer binding site and a DNA synthesis template sequence that encodes via a polymerase (e.g., a reverse transcriptase) a single stranded DNA flap containing the

genetic change of interest, which then integrates into the endogenous DNA by replacing the corresponding endogenous strand, thereby installing the desired genetic change.

[0131] Transcription terminator – the guide RNA or PEgRNA may comprise a transcriptional termination sequence at the 3′ of the molecule.

Host cell

[0132] The term "host cell," as used herein, refers to a cell that can host, replicate, and express a vector described herein, e.g., a vector comprising a nucleic acid molecule encoding an MLH1 variant and a fusion protein comprising a Cas9 or Cas9 equivalent and a reverse transcriptase.

Linker

[0133] The term "linker," as used herein, refers to a molecule linking two other molecules or moieties. The linker can be an amino acid sequence in the case of a linker joining two fusion proteins. For example, a Cas9 can be fused to a reverse transcriptase by an amino acid linker sequence. The linker can also be a nucleotide sequence in the case of joining two nucleotide sequences together. For example, in the instant case, the traditional guide RNA is linked via a spacer or linker nucleotide sequence to the RNA extension of a prime editing guide RNA which may comprise a RT template sequence and an RT primer binding site. In other embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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. Longer or shorter linkers are also contemplated. In certain embodiments, the linker is a self-hydrolyzing linker (*e.g.*, a 2A self-cleaving peptide as described further herein). Self-hydrolyzing linkers such as 2A self-cleaving peptides are capable of inducing ribosomal skipping during protein translation, resulting in the ribosome failing to make a peptide bond between two genes, or gene fragments.

napDNAbp

[0134] As used herein, the term "nucleic acid programmable DNA binding protein" or "napDNAbp," of which Cas9 is an example, refer to proteins that use RNA:DNA hybridization to target and bind to specific sequences in a DNA molecule. Each napDNAbp is associated with at least one guide nucleic acid (e.g., guide RNA), which localizes the napDNAbp to a DNA sequence that comprises a DNA strand (i.e., a target strand) that is complementary to the guide nucleic acid, or a portion thereof (e.g., the protospacer of a guide

RNA). In other words, the guide nucleic-acid "programs" the napDNAbp (e.g., Cas9 or equivalent) to localize and bind to a complementary sequence.

[0135] Without being bound by theory, the binding mechanism of a napDNAbp – guide RNA complex, in general, includes the step of forming an R-loop whereby the napDNAbp induces the unwinding of a double-strand DNA target, thereby separating the strands in the region bound by the napDNAbp. The guide RNA protospacer then hybridizes to the "target strand." This displaces a "non-target strand" that is complementary to the target strand, which forms the single strand region of the R-loop. In some embodiments, the napDNAbp includes one or more nuclease activities, which then cut the DNA, leaving various types of lesions. For example, the napDNAbp may comprises a nuclease activity that cuts the non-target strand at a first location, and/or cuts the target strand at a second location. Depending on the nuclease activity, the target DNA can be cut to form a "double-stranded break" whereby both strands are cut. In other embodiments, the target DNA can be cut at only a single site, i.e., the DNA is "nicked" on one strand. Exemplary napDNAbp with different nuclease activities include "Cas9 nickase" ("nCas9") and a deactivated Cas9 having no nuclease activities ("dead Cas9" or "dCas9"). Exemplary sequences for these and other napDNAbp are provided herein.

Nickase

[0136] The term "nickase" refers to a Cas9 with one of the two nuclease domains inactivated. This enzyme is capable of cleaving only one strand of a target DNA.

Nucleic acid molecule

[0137] The term "nucleic acid," as used herein, refers to a polymer of nucleotides. The polymer may include natural nucleosides (i.e., 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, C5 bromouridine, C5 fluorouridine, C5 iodouridine, C5 propynyl uridine, C5 propynyl cytidine, C5 methylcytidine, 7 deazaadenosine, 7 deazaguanosine, 8 oxoadenosine, 8 oxoguanosine, O(6) methylguanine, 4-acetylcytidine, 5-(carboxyhydroxymethyl)uridine, dihydrouridine, methylpseudouridine, 1-methyl adenosine, 1-methyl guanosine, N6-methyl adenosine, and 2-thiocytidine), chemically modified bases, biologically modified bases (e.g., methylated bases), intercalated bases, modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, 2´-O-methylcytidine, arabinose, and hexose), or modified phosphate groups (e.g., phosphorothioates and 5′ N phosphoramidite linkages).

PACE

[0138] The term "phage-assisted continuous evolution (PACE)," as used herein, refers to continuous evolution that employs phage as viral vectors. The general concept of PACE technology has been described, for example, in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Application, U.S. Patent No. 9,023,594, issued May 5, 2015, International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015, and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference.

PEgRNA

[0139] As used herein, the terms "prime editing guide RNA" or "PEgRNA" or "extended guide RNA" refer to a specialized form of a guide RNA that has been modified to include one or more additional sequences for implementing the prime editing methods and compositions described herein. As described herein, the prime editing guide RNA comprise one or more "extended regions" of nucleic acid sequence. The extended regions may comprise, but are not limited to, single-stranded RNA or DNA. Further, the extended regions may occur at the 3′ end of a traditional guide RNA. In other arrangements, the extended regions may occur at the 5′ end of a traditional guide RNA. In still other arrangements, the extended region may occur at an intramolecular region of the traditional guide RNA, for example, in the gRNA core region which associates and/or binds to the napDNAbp. The extended region comprises a "DNA synthesis template" which encodes (by the polymerase of the prime editor) a single-stranded DNA which, in turn, has been designed to be (a) homologous with the endogenous target DNA to be edited, and (b) which comprises at least one desired nucleotide change (e.g., a transition, a transversion, a deletion, or an insertion) to be introduced or integrated into the endogenous target DNA. The extended region may also comprise other functional sequence elements, such as, but not limited to, a "primer binding site" and a "spacer or linker" sequence, or other structural elements, such as, but not limited to aptamers, stem loops, hairpins, toe loops (e.g., a 3′ toeloop), or an RNA-protein recruitment domain (e.g., MS2 hairpin). As used herein the "primer binding site" comprises a

sequence that hybridizes to a single-strand DNA sequence having a 3′end generated from the nicked DNA of the R-loop.

[0140] In certain embodiments, the PEgRNAs have a 5′ extension arm, a spacer, and a gRNA core. The 5′ extension further comprises in the 5′ to 3′ direction a reverse transcriptase template, a primer binding site, and a linker. The reverse transcriptase template may also be referred to more broadly as the "DNA synthesis template" where the polymerase of a prime editor described herein is not an RT, but another type of polymerase.

[0141] In certain other embodiments, the PEgRNAs have a 5′ extension arm, a spacer, and a gRNA core. The 5′ extension further comprises in the 5′ to 3′ direction a reverse transcriptase template, a primer binding site, and a linker. The reverse transcriptase template may also be referred to more broadly as the "DNA synthesis template" where the polymerase of a prime editor described herein is not an RT, but another type of polymerase.

[0142] In still other embodiments, the PEgRNAs have in the 5′ to 3′ direction a spacer (1), a gRNA core (2), and an extension arm (3). The extension arm (3) is at the 3′ end of the PEgRNA. The extension arm (3) further comprises in the 5′ to 3′ direction a "primer binding site" (A), an "edit template" (B), and a "homology arm" (C). The extension arm (3) may also comprise an optional modifier region at the 3′ and 5′ ends, which may be the same sequences or different sequences. In addition, the 3′ end of the PEgRNA may comprise a transcriptional terminator sequence. These sequence elements of the PEgRNAs are further described and defined herein.

[0143] In still other embodiments, the PEgRNAs have in the 5′ to 3′ direction an extension arm (3), a spacer (1), and a gRNA core (2). The extension arm (3) is at the 5′ end of the PEgRNA. The extension arm (3) further comprises in the 3′ to 5′ direction a "primer binding site" (A), an "edit template" (B), and a "homology arm" (C). The extension arm (3) may also comprise an optional modifier region at the 3′ and 5′ ends, which may be the same sequences or different sequences. The PEgRNAs may also comprise a transcriptional terminator sequence at the 3′ end. These sequence elements of the PEgRNAs are further described and defined herein.

PE1

[0144] As used herein, "PE1" refers to a PE complex comprising a fusion protein comprising Cas9(H840A) and a wild type MMLV RT having the following structure: [NLS]-[Cas9(H840A)]-[linker]-[MMLV_RT(wt)] + a desired PEgRNA, wherein the PE fusion has the amino acid sequence of SEQ ID NO: 3, which is shown as follows;

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN
TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD
KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA
ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV
ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL
SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV
SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD
VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT
LIHQSITGLYETRIDLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS*TL
*NIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY
*PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVED
*IHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT
*WTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALL
*QTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF
*LGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTK
*PFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG
*KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL
*PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTET
*EVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSE
*GKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTS
*TLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 3)

KEY:
NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96)
CAS9(H840A) (SEQ ID NO: 10)
33-AMINO ACID LINKER (SEQ ID NO: 80)
M-MLV reverse transcriptase (SEQ ID NO: 33).

PE2

[0145] As used herein, "PE2" refers to a PE complex comprising a fusion protein comprising Cas9(H840A) and a variant MMLV RT having the following structure: [NLS]-

[Cas9(H840A)]-[linker]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)] + a desired PEgRNA, wherein the PE fusion has the amino acid sequence of SEQ ID NO: 4, which is shown as follows

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNK VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGL SELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYD VRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS***TL NIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVED IHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALL QTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF LGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTK PFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTET EVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTS EGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT STLLIENSSP**SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 4)

KEY:
NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96)
CAS9(H840A) (SEQ ID NO: 10)
33-AMINO ACID LINKER (SEQ ID NO: 80)
M-MLV reverse transcriptase (SEQ ID NO: 34).

PE3

[0146] As used herein, "PE3" refers to PE2 plus a second-strand nicking guide RNA that complexes with the PE2 and introduces a nick in the non-edited DNA strand in order to induce preferential replacement of the edited strand.

PE3b

[0147] As used herein, "PE3b" refers to PE3 but wherein the second-strand nicking guide RNA is designed for temporal control such that the second strand nick is not introduced until after the installation of the desired edit. This is achieved by designing a gRNA with a spacer sequence that matches only the edited strand, but not the original allele. Using this strategy, referred to hereafter as PE3b, mismatches between the protospacer and the unedited allele should disfavor nicking by the sgRNA until after the editing event on the PAM strand takes place.

PEmax

[0148] As used herein, "PEmax" refers to a PE complex comprising a fusion protein comprising Cas9(R221K N39K H840A) and a variant MMLV RT pentamutant (D200N T306K W313F T330P L603W) having the following structure: [bipartite NLS]-[Cas9(R221K)(N394K)(H840A)]-[linker]-[MMLV_RT(D200N)(T330P)(L603W)]-[bipartite NLS]-[NLS] + a desired PEgRNA, wherein the PE fusion has the amino acid sequence of SEQ ID NO: 2, and the nucleic acid sequence of SEQ ID NO: 1 which are shown as follows:

MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLG**
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA
KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST
DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI
NASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA

NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKD
DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVIL
ADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT
STKEVLDATLIHQSITGLYETRIDLSQLGGD<u>**SGGSSGGSKRTADGSEFESPKKKR**</u>
<u>**KVSGGSSGGS**</u>TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA
PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGT
NDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPT
SQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ
YVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEG
QRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTL
FNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWR
RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALV
KQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILA
EAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGT
SAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKN
KDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
IENSSP*SGGS*<u>KRTADGSEFESPKKKRKV</u>*GSG*<u>*PAAKRVKLD*</u> (SEQ ID NO: 2)

<u>ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAA</u>
<u>AGTC</u>**GACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGG**
CTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGT
GCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCT
GCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGC
CAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGAT
CTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGA
AGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTT
CGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTA

CCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCT
GATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATC
GAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAG
CTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGC
GGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGAAAG
CTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTC
GGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAAC
TTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGAC
GACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTG
TTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTG
AGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAG
AGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG
CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAAC
GGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAG
TTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTG
AAGCTGAAGAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGC
AGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGG
CAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAG
ATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC
AGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGG
AACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAG
CGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAG
CACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTG
AAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAG
AAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTG
AAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTG
GAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCAC
GATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAAC
GAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGA
GAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAA
GTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAG
CCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCT
GGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATC
CACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCC
GGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCC
GCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTG
AAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGA
GAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAA
GCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACA
CCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCT
GCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCT
GTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGA
CTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAG
CGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCG
GCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGAC
CAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAA

GAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCT
GGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGA
AGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGA
TTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGAC
GCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAG
CTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTC
TTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACG
GCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAG
ATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGC
ATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTC
AGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGA
AAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTG
GCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAA
CTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGC
TTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTG
AAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAA
ACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAAC
GAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTG
TGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGT
TCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCG
CCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCA
TCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTT
TGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGA
CGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGA
CCTGTCTCAGCTGGGAGGTGAC<u>**TCCGGCGGAAGCTCTGGTGGCAGCAAGCG**</u>
<u>**GACCGCCGACGGCTCTGAATTCGAGAGCCCTAAGAAGAAAAGAAAGGTGAG**</u>
<u>**CGGAGGCTCTAGCGGCGGAAGC**</u>ACCCTGAACATTGAAGACGAGTATAGACTG
CATGAAACAAGCAAGGAACCCGACGTGTCCCTGGGCTCCACCTGGCTGTCCGAC
TTTCCCCAGGCCTGGGCCGAGACAGGAGGAATGGGCCTGGCCGTGCGGCAGGCA
CCCCTGATCATCCCTCTGAAGGCCACCTCTACACCCGTGAGCATCAAGCAGTACC
CTATGTCTCAGGAGGCCAGACTGGGCATCAAGCCTCACATCCAGAGGCTGCTGG
ACCAGGGCATCCTGGTGCCATGCCAGAGCCCCTGGAACACACCACTGCTGCCCG
TGAAGAAGCCAGGCACCAATGACTATAGACCCGTGCAGGATCTGAGAGAGGTGA
ACAAGAGGGTGGAGGATATCCACCCCACCGTGCCCAACCCTTACAATCTGCTGTC
CGGCCTGCCCCCTTCTCACCAGTGGTATACAGTGCTGGACCTGAAGGATGCCTTC
TTTTGTCTGAGACTGCACCCTACCAGCCAGCCACTGTTCGCCTTTGAGTGGAGGG
ACCCTGAGATGGGCATCTCTGGCCAGCTGACCTGGACACGCCTGCCTCAGGGCTT
CAAGAATAGCCCAACACTGTTTAACGAGGCCCTGCACCGCGACCTGGCAGATTT
CCGGATCCAGCACCCAGATCTGATCCTGCTGCAGTACGTGGACGATCTGCTGCTG
GCCGCCACCAGCGAGCTGGATTGCCAGCAGGGAACACGCGCCCTGCTGCAGACC
CTGGGAAACCTGGGATATAGGGCATCCGCCAAGAAGGCCCAGATCTGTCAGAAG
CAGGTGAAGTACCTGGGCTATCTGCTGAAGGAGGGCCAGAGATGGCTGACAGAG
GCCAGGAAGGAGACAGTGATGGGCCAGCCAACACCCAAGACCCCAAGACAGCT
GAGGGAGTTCCTGGGCAAAGCAGGATTTTGCAGGCTGTTCATCCCAGGATTCGC

AGAGATGGCAGCACCTCTGTACCCACTGACCAAGCCGGGCACCCTGTTTAATTGG
GGCCCTGACCAGCAGAAGGCCTATCAGGAGATCAAGCAGGCCCTGCTGACAGCA
CCAGCCCTGGGCCTGCCAGACCTGACCAAGCCTTTCGAGCTGTTTGTGGATGAGA
AGCAGGGCTACGCCAAGGGCGTGCTGACCCAGAAGCTGGGACCATGGAGACGG
CCCGTGGCCTATCTGTCCAAGAAGCTGGACCCAGTGGCAGCAGGATGGCCACCA
TGCCTGAGGATGGTGGCAGCAATCGCCGTGCTGACAAAGGATGCCGGCAAGCTG
ACCATGGGACAGCCACTGGTCATCCTGGCACCACACGCAGTGGAGGCCCTGGTG
AAGCAGCCTCCAGATCGCTGGCTGTCTAACGCCCGGATGACACACTACCAGGCC
CTGCTGCTGGACACCGATCGCGTGCAGTTTGGCCCTGTGGTGGCCCTGAATCCAG
CCACCCTGCTGCCTCTGCCAGAGGAGGGCCTGCAGCACAACTGTCTGGACATCCT
GGCAGAGGCACACGGAACAAGGCCAGACCTGACCGATCAGCCCCTGCCTGACGC
CGATCACACATGGTATACCGATGGAAGCTCCCTGCTGCAGGAGGGCCAGAGGAA
GGCAGGAGCAGCAGTGACCACAGAGACAGAAGTGATCTGGGCCAAGGCCCTGC
CAGCAGGCACATCCGCCCAGCGGGCCGAGCTGATCGCCCTGACCCAGGCCCTGA
AGATGGCCGAGGGCAAGAAGCTGAACGTGTACACAGACTCCAGATATGCCTTCG
CCACCGCACACATCCACGGAGAGATCTACAGGCGCCGGGGCTGGCTGACCTCTG
AGGGCAAGGAGATCAAGAACAAGGATGAGATCCTGGCCCTGCTGAAGGCCCTGT
TTCTGCCCAAGCGGCTGAGCATCATCCACTGTCCTGGACACCAGAAGGGACACTC
CGCCGAGGCAAGGGGCAATCGGATGGCCGACCAGGCCGCCAGAAAGGCTGCTAT
TACTGAAACTCCCGACACTTCCACTCTGCTGATTGAAAACTCCTCCCCT*TCTGGCG*
*GCTCA*AAAAGAACCGCCGACGGCAGCGAATTCGAGTCTCCCAAGAAGAAGAGGA
AAGTC*GGCTCTGGCCCTGCCGCTAAGAGAGTGAAGCTGGAC* (SEQ ID NO: 1)

KEY:

BIPARTITE SV40 NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP: (SEQ ID NO: 95),

CAS9(R221K N39K H840A) (SEQ ID NO: 11)

SGGSx2-BIPARTITE SV40NLS-SGGSx2 LINKER (SEQ ID NO: 79)

M-MLV reverse transcriptase (D200N T306K W313F T330P L603W) (SEQ ID NO: 34)

Other linker sequence (SEQ ID NOs: 81)

BIPARTITE SV40 NLS (SEQ ID NO: 97)

Other linker sequence (SEQ ID NOs: 82)

c-Myc NLS (SEQ ID NO: 98)

Polymerase

[0149] As used herein, the term "polymerase" refers to an enzyme that synthesizes a nucleotide strand and that may be used in connection with the prime editor systems described herein. The polymerase can be a "template-dependent" polymerase (i.e., a polymerase that synthesizes a nucleotide strand based on the order of nucleotide bases of a template strand).

The polymerase can also be a "template-independent" polymerase (i.e., a polymerase that synthesizes a nucleotide strand without the requirement of a template strand). A polymerase may also be further categorized as a "DNA polymerase" or an "RNA polymerase." In various embodiments, the prime editor system comprises a DNA polymerase. In various embodiments, the DNA polymerase can be a "DNA-dependent DNA polymerase" (i.e., whereby the template molecule is a strand of DNA). In such cases, the DNA template molecule can be a PEgRNA, wherein the extension arm comprises a strand of DNA. In such cases, the PEgRNA may be referred to as a chimeric or hybrid PEgRNA which comprises an RNA portion (i.e., the guide RNA components, including the spacer and the gRNA core) and a DNA portion (i.e., the extension arm). In various other embodiments, the DNA polymerase can be an "RNA-dependent DNA polymerase" (i.e., whereby the template molecule is a strand of RNA). In such cases, the PEgRNA is RNA, i.e., including an RNA extension. The term "polymerase" may also refer to an enzyme that catalyzes the polymerization of nucleotide (i.e., the polymerase activity). Generally, the enzyme will initiate synthesis at the 3'-end of a primer annealed to a polynucleotide template sequence (e.g., such as a primer sequence annealed to the primer binding site of a PEgRNA) and will proceed toward the 5´ end of the template strand. A "DNA polymerase" catalyzes the polymerization of deoxynucleotides. As used herein in reference to a DNA polymerase, the term DNA polymerase includes a "functional fragment thereof". A "functional fragment thereof" refers to any portion of a wild-type or mutant DNA polymerase that encompasses less than the entire amino acid sequence of the polymerase and which retains the ability, under at least one set of conditions, to catalyze the polymerization of a polynucleotide. Such a functional fragment may exist as a separate entity, or it may be a constituent of a larger polypeptide, such as a fusion protein.

Prime editing

[0150] As used herein, the term "prime editing" refers to an approach for gene editing using napDNAbps, a polymerase (e.g., a reverse transcriptase), and specialized guide RNAs that include a DNA synthesis template for encoding desired new genetic information (or deleting genetic information) that is then incorporated into a target DNA sequence. Classical prime editing is described in the inventors publication of Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019), which is incorporated herein by reference in its entirety.

[0151] Prime editing represents a platform for genome editing that is a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site using a nucleic acid programmable DNA binding protein ("napDNAbp") working in association with a polymerase (i.e., in the form of a fusion protein or otherwise provided *in trans* with the napDNAbp), wherein the prime editing system is programmed with a prime editing (PE) guide RNA ("PEgRNA") that both specifies the target site and templates the synthesis of the desired edit in the form of a replacement DNA strand by way of an extension (either DNA or RNA) engineered onto a guide RNA (e.g., at the 5′ or 3′ end, or at an internal portion of a guide RNA). The replacement strand containing the desired edit (e.g., a single nucleobase substitution) shares the same (or is homologous to) sequence as the endogenous strand (immediately downstream of the nick site) of the target site to be edited (with the exception that it includes the desired edit). Through DNA repair and/or replication machinery, the endogenous strand downstream of the nick site is replaced by the newly synthesized replacement strand containing the desired edit. In some cases, prime editing may be thought of as a "search-and-replace" genome editing technology since the prime editors, as described herein, not only search and locate the desired target site to be edited, but at the same time, encode a replacement strand containing a desired edit which is installed in place of the corresponding target site endogenous DNA strand. The prime editors of the present disclosure relate, in part, to the discovery that the mechanism of target-primed reverse transcription (TPRT) or "prime editing" can be leveraged or adapted for conducting precision CRISPR/Cas-based genome editing with high efficiency and genetic flexibility. TPRT is naturally used by mobile DNA elements, such as mammalian non-LTR retrotransposons and bacterial Group II introns. The inventors have herein used Cas protein-reverse transcriptase fusions or related systems *in trans* to target a specific DNA sequence with a guide RNA, generate a single strand nick at the target site, and use the nicked DNA as a primer for reverse transcription of an engineered reverse transcriptase template that is integrated with the guide RNA. However, while the concept begins with prime editors that use reverse transcriptase as the DNA polymerase component, the prime editors described herein are not limited to reverse transcriptases but may include the use of virtually any DNA polymerase. Indeed, while the application throughout may refer to prime editors with "reverse transcriptases," it is set forth here that reverse transcriptases are only one type of DNA polymerase that may work with prime editing. Thus, wherever the specification mentions a "reverse transcriptase," the person having ordinary skill in the art should appreciate that any suitable DNA polymerase may be used in place of the reverse transcriptase. Thus, in one aspect, the prime editors may comprise

Cas9 (or an equivalent napDNAbp) which is programmed to target a DNA sequence by associating it with a specialized guide RNA (i.e., PEgRNA) containing a spacer sequence that anneals to a complementary protospacer in the target DNA. The specialized guide RNA also contains new genetic information in the form of an extension that encodes a replacement strand of DNA containing a desired genetic alteration which is used to replace a corresponding endogenous DNA strand at the target site. To transfer information from the PEgRNA to the target DNA, the mechanism of prime editing involves nicking the target site in one strand of the DNA to expose a 3′-hydroxyl group. The exposed 3′-hydroxyl group can then be used to prime the DNA polymerization of the edit-encoding extension on PEgRNA directly into the target site. In various embodiments, the extension—which provides the template for polymerization of the replacement strand containing the edit—can be formed from RNA or DNA. In the case of an RNA extension, the polymerase of the prime editor can be an RNA-dependent DNA polymerase (such as, a reverse transcriptase). In the case of a DNA extension, the polymerase of the prime editor may be a DNA-dependent DNA polymerase. The newly synthesized strand (i.e., the replacement DNA strand containing the desired edit) that is formed by the herein disclosed prime editors would be homologous to the genomic target sequence (i.e., have the same sequence as) except for the inclusion of a desired nucleotide change (e.g., a single nucleotide change, a deletion, or an insertion, or a combination thereof). The newly synthesized (or replacement) strand of DNA may also be referred to as a single strand DNA flap, which would compete for hybridization with the complementary homologous endogenous DNA strand, thereby displacing the corresponding endogenous strand. In certain embodiments, the system can be combined with the use of an error-prone reverse transcriptase enzyme (e.g., provided as a fusion protein with the Cas9 domain, or provided *in trans* to the Cas9 domain). The error-prone reverse transcriptase enzyme can introduce alterations during synthesis of the single strand DNA flap. Thus, in certain embodiments, error-prone reverse transcriptase can be utilized to introduce nucleotide changes to the target DNA. Depending on the error-prone reverse transcriptase that is used with the system, the changes can be random or non-random. Resolution of the hybridized intermediate (comprising the single strand DNA flap synthesized by the reverse transcriptase hybridized to the endogenous DNA strand) can include removal of the resulting displaced flap of endogenous DNA (e.g., with a 5′ end DNA flap endonuclease, FEN1), ligation of the synthesized single strand DNA flap to the target DNA, and assimilation of the desired nucleotide change as a result of cellular DNA repair and/or replication processes. Because templated DNA synthesis offers single nucleotide precision for the modification of any

nucleotide, including insertions and deletions, the scope of this approach is very broad and could foreseeably be used for myriad applications in basic science and therapeutics.

[0152] In various embodiments, prime editing operates by contacting a target DNA molecule (for which a change in the nucleotide sequence is desired to be introduced) with a nucleic acid programmable DNA binding protein (napDNAbp) complexed with a prime editing guide RNA (PEgRNA). In various embodiments, the prime editing guide RNA (PEgRNA) comprises an extension at the 3′ or 5′ end of the guide RNA, or at an intramolecular location in the guide RNA and encodes the desired nucleotide change (e.g., single nucleotide change, insertion, or deletion). In step (a), the napDNAbp/extended gRNA complex contacts the DNA molecule and the extended gRNA guides the napDNAbp to bind to a target locus. In step (b), a nick in one of the strands of DNA of the target locus is introduced (e.g., by a nuclease or chemical agent), thereby creating an available 3′ end in one of the strands of the target locus. In certain embodiments, the nick is created in the strand of DNA that corresponds to the R-loop strand, i.e., the strand that is not hybridized to the guide RNA sequence, i.e., the "non-target strand." The nick, however, could be introduced in either of the strands. That is, the nick could be introduced into the R-loop "target strand" (i.e., the strand hybridized to the protospacer of the extended gRNA) or the "non-target strand" (i.e., the strand forming the single-stranded portion of the R-loop and which is complementary to the target strand). In step (c), the 3′ end of the DNA strand (formed by the nick) interacts with the extended portion of the guide RNA in order to prime reverse transcription (i.e., "target-primed RT"). In certain embodiments, the 3′ end DNA strand hybridizes to a specific RT priming sequence on the extended portion of the guide RNA, i.e., the "reverse transcriptase priming sequence" or "primer binding site" on the PEgRNA. In step (d), a reverse transcriptase (or other suitable DNA polymerase) is introduced which synthesizes a single strand of DNA from the 3′ end of the primed site towards the 5′ end of the prime editing guide RNA. The DNA polymerase (e.g., reverse transcriptase) can be fused to the napDNAbp or alternatively can be provided *in trans* to the napDNAbp. This forms a single-strand DNA flap comprising the desired nucleotide change (e.g., the single base change, insertion, or deletion, or a combination thereof) and which is otherwise homologous to the endogenous DNA at or adjacent to the nick site. In step (e), the napDNAbp and guide RNA are released. Steps (f) and (g) relate to the resolution of the single strand DNA flap such that the desired nucleotide change becomes incorporated into the target locus. This process can be driven towards the desired product formation by removing the corresponding 5′ endogenous DNA flap that forms once the 3′ single strand DNA flap invades and hybridizes to the endogenous

DNA sequence. Without being bound by theory, the cells endogenous DNA repair and replication processes resolves the mismatched DNA to incorporate the nucleotide change(s) to form the desired altered product. The process can also be driven towards product formation with "second strand nicking." This process may introduce at least one or more of the following genetic changes: transversions, transitions, deletions, and insertions.

[0153] The term "prime editor (PE) system" or "prime editor (PE)" or "PE system" or "PE editing system" refers the compositions involved in the method of genome editing using target-primed reverse transcription (TPRT) describe herein, including, but not limited to the napDNAbps, reverse transcriptases, fusion proteins (e.g., comprising napDNAbps and reverse transcriptases), prime editing guide RNAs, and complexes comprising fusion proteins and prime editing guide RNAs, as well as accessory elements, such as second strand nicking components (e.g., second strand sgRNAs) and 5′ endogenous DNA flap removal endonucleases (e.g., FEN1) for helping to drive the prime editing process towards the edited product formation.

[0154] Although in the embodiments described thus far the PEgRNA constitutes a single molecule comprising a guide RNA (which itself comprises a spacer sequence and a gRNA core or scaffold) and a 5′ or 3′ extension arm comprising the primer binding site and a DNA synthesis template, the PEgRNA may also take the form of two individual molecules comprised of a guide RNA and a *trans* prime editor RNA template (tPERT), which essentially houses the extension arm (including, in particular, the primer binding site and the DNA synthesis domain) and an RNA-protein recruitment domain (e.g., MS2 aptamer or hairpin) in the same molecule which becomes co-localized or recruited to a modified prime editor complex that comprises a tPERT recruiting protein (e.g., MS2cp protein, which binds to the MS2 aptamer).

Prime editor

[0155] The term "prime editor" refers to constructs comprising a napDNAbp (e.g., Cas9 nickase) and a reverse transcriptase that are capable of carrying out prime editing on a target nucleotide sequence in the presence of a PEgRNA (or "extended guide RNA"). The term "prime editor" may refer to a fusion protein or to a fusion protein complexed with a PEgRNA, and/or further complexed with a second-strand nicking sgRNA. In some embodiments, the prime editor may also refer to the complex comprising a fusion protein (reverse transcriptase fused to a napDNAbp), a PEgRNA, and a regular guide RNA capable of directing the second-site nicking step of the non-edited strand as described herein. In some

embodiments, the term prime editor refers to a napDNAbp and a reverse transcriptase that are provided *in trans*, or that are otherwise not fused to one another.

Primer binding site

[0156] The term "primer binding site" or "the PBS" refers to the nucleotide sequence located on a PEgRNA as a component of the extension arm (typically at the 3′ end of the extension arm) and serves to bind to the primer sequence that is formed after Cas9 nicking of the target sequence by the prime editor. As detailed elsewhere, when the Cas9 nickase component of a prime editor nicks one strand of the target DNA sequence, a 3′-ended ssDNA flap is formed, which serves a primer sequence that anneals to the primer binding site on the PEgRNA to prime reverse transcription.

Protospacer

[0157] As used herein, the term "protospacer" refers to the sequence (~20 bp) in DNA adjacent to the PAM (protospacer adjacent motif) sequence. The protospacer shares the same sequence as the spacer sequence of the guide RNA. The guide RNA anneals to the complement of the protospacer sequence on the target DNA (specifically, one strand thereof, i.e., the "target strand" versus the "non-target strand" of the target DNA sequence). In order for Cas9 to function it also requires a specific protospacer adjacent motif (PAM) that varies depending on the bacterial species of the Cas9 gene. The most commonly used Cas9 nuclease, derived from *S. pyogenes*, recognizes a PAM sequence of NGG that is found directly downstream of the target sequence in the genomic DNA, on the non-target strand. The skilled person will appreciate that the literature in the state of the art sometimes refers to the "protospacer" as the ~20-nt target-specific guide sequence on the guide RNA itself, rather than referring to it as a "spacer." Thus, in some cases, the term "protospacer" as used herein may be used interchangeably with the term "spacer." The context of the description surrounding the appearance of either "protospacer" or "spacer" will help inform the reader as to whether the term is in reference to the gRNA or the DNA target.

Protospacer adjacent motif (PAM)

[0158] As used herein, the term "protospacer adjacent sequence" or "PAM" refers to an approximately 2-6 base pair DNA sequence that is an important targeting component of a Cas9 nuclease. Typically, the PAM sequence is on either strand, and is downstream in the 5′ to 3′ direction of the Cas9 cut site. The canonical PAM sequence (i.e., the PAM sequence that is associated with the Cas9 nuclease of *Streptococcus pyogenes* or SpCas9) is 5′-NGG-3′ wherein "N" is any nucleobase followed by two guanine ("G") nucleobases. Different PAM

sequences can be associated with different Cas9 nucleases or equivalent proteins from different organisms. In addition, any given Cas9 nuclease, e.g., SpCas9, may be modified to alter the PAM specificity of the nuclease such that the nuclease recognizes alternative PAM sequence.

[0159] For example, with reference to the canonical SpCas9 amino acid sequence is SEQ ID NO: 9, the PAM sequence can be modified by introducing one or more mutations, including (a) D1135V, R1335Q, and T1337R "the VQR variant", which alters the PAM specificity to NGAN or NGNG, (b) D1135E, R1335Q, and T1337R "the EQR variant", which alters the PAM specificity to NGAG, and (c) D1135V, G1218R, R1335E, and T1337R "the VRER variant", which alters the PAM specificity to NGCG. In addition, the D1135E variant of canonical SpCas9 still recognizes NGG, but it is more selective compared to the wild type SpCas9 protein.

[0160] It will also be appreciated that Cas9 enzymes from different bacterial species (i.e., Cas9 orthologs) can have varying PAM specificities. For example, Cas9 from *Staphylococcus aureus* (SaCas9) recognizes NGRRT or NGRRN. In addition, Cas9 from *Neisseria meningitis* (NmCas) recognizes NNNNGATT. In another example, Cas9 from *Streptococcus thermophilis* (StCas9) recognizes NNAGAAW. In still another example, Cas9 from *Treponema denticola* (TdCas) recognizes NAAAAC. These are examples and are not meant to be limiting. It will be further appreciated that non-SpCas9s bind a variety of PAM sequences, which makes them useful when no suitable SpCas9 PAM sequence is present at the desired target cut site. Furthermore, non-SpCas9s may have other characteristics that make them more useful than SpCas9. For example, Cas9 from *Staphylococcus aureus* (SaCas9) is about 1 kilobase smaller than SpCas9, so it can be packaged into adeno-associated virus (AAV). Further reference may be made to Shah *et al.*, "Protospacer recognition motifs: mixed identities and functional diversity," *RNA Biology*, 10(5): 891-899 (which is incorporated herein by reference).

Reverse transcriptase

[0161] The term "reverse transcriptase" describes a class of polymerases characterized as RNA-dependent DNA polymerases. All known reverse transcriptases require a primer to synthesize a DNA transcript from an RNA template. Historically, reverse transcriptase has been used primarily to transcribe mRNA into cDNA which can then be cloned into a vector for further manipulation. Avian myoblastosis virus (AMV) reverse transcriptase was the first widely used RNA-dependent DNA polymerase (Verma, Biochim. Biophys. Acta 473:1

(1977)). The enzyme has 5′-3′ RNA-directed DNA polymerase activity, 5′-3′ DNA-directed DNA polymerase activity, and RNase H activity. RNase H is a processive 5′ and 3′ ribonuclease specific for the RNA strand for RNA-DNA hybrids (Perbal, A Practical Guide to Molecular Cloning, New York: Wiley & Sons (1984)). Errors in transcription cannot be corrected by reverse transcriptase because known viral reverse transcriptases lack the 3′-5′ exonuclease activity necessary for proofreading (Saunders and Saunders, Microbial Genetics Applied to Biotechnology, London: Croom Helm (1987)). A detailed study of the activity of AMV reverse transcriptase and its associated RNase H activity has been presented by Berger *et al.*, Biochemistry 22:2365-2372 (1983). Another reverse transcriptase which is used extensively in molecular biology is reverse transcriptase originating from Moloney murine leukemia virus (M-MLV or "MMLV"). See, e.g., Gerard, G. R., DNA 5:271-279 (1986) and Kotewicz, M. L., *et al.*, Gene 35:249-258 (1985). M-MLV reverse transcriptase substantially lacking in RNase H activity has also been described. See, e.g., U.S. Pat. No. 5,244,797. The invention contemplates the use of any such reverse transcriptases, or variants or mutants thereof.

[0162] In addition, the invention contemplates the use of reverse transcriptases that are error-prone, i.e., that may be referred to as error-prone reverse transcriptases or reverse transcriptases that do not support high fidelity incorporation of nucleotides during polymerization. During synthesis of the single-strand DNA flap based on the RT template integrated with the guide RNA, the error-prone reverse transcriptase can introduce one or more nucleotides which are mismatched with the RT template sequence, thereby introducing changes to the nucleotide sequence through erroneous polymerization of the single-strand DNA flap. These errors introduced during synthesis of the single strand DNA flap then become integrated into the double strand molecule through hybridization to the corresponding endogenous target strand, removal of the endogenous displaced strand, ligation, and then through one more round of endogenous DNA repair and/or sequencing processes.

[0163] The disclosure provides in some embodiments prime editors comprising MMLV RT.

Reverse transcription

[0164] As used herein, the term "reverse transcription" indicates the capability of an enzyme to synthesize a DNA strand (that is, complementary DNA or cDNA) using RNA as a template. In some embodiments, the reverse transcription can be "error-prone reverse

transcription," which refers to the properties of certain reverse transcriptase enzymes which are error-prone in their DNA polymerization activity.

Protein, peptide, and polypeptide

[0165] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, Molecular Cloning: A Laboratory Manual (^{4}th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

Spacer sequence

[0166] As used herein, the term "spacer sequence" in connection with a guide RNA or a PEgRNA refers to the portion of the guide RNA or PEgRNA of about 20 nucleotides which contains a nucleotide sequence that shares the same sequence as the protospacer sequence in the target DNA sequence. The spacer sequence anneals to the complement of the protospacer sequence to form a ssRNA/ssDNA hybrid structure at the target site and a corresponding R loop ssDNA structure of the endogenous DNA strand.

Target site

[0167] The term "target site" refers to a sequence within a nucleic acid molecule that is edited by a prime editor (PE) disclosed herein. The target site further refers to the sequence within a nucleic acid molecule to which a complex of the prime editor (PE) and gRNA binds.

Variant

[0168] As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature, e.g., a variant Cas9 is a Cas9 comprising one or more changes in amino acid residues as compared to a wild type Cas9 amino acid sequence. The term "variant" encompasses homologous proteins having at least 75%, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 99% percent identity with a reference sequence and having the same or substantially the same functional activity or activities as the reference sequence. The term also encompasses mutants, truncations, or domains of a reference sequence, and which display the same or substantially the same functional activity or activities as the reference sequence.

Vector

[0169] The term "vector," as used herein, refers to a nucleic acid that can be modified to encode a gene of interest and that is able to enter into a host cell, mutate and replicate within the host cell, and then transfer a replicated form of the vector into another host cell. Exemplary suitable vectors include viral vectors, such as retroviral vectors or bacteriophages and filamentous phage, and conjugative plasmids. Additional suitable vectors will be apparent to those of skill in the art based on the instant disclosure.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[0170] The present disclosure provides compositions and methods for prime editing with improved editing efficiency and/or reduced indel formation. In particular, the disclosure provides improved prime editor proteins wherein one or more components, including the napDNAbp domain and/or reverse transcriptase domain are modified (e.g., the amino acid sequence is changed relative to a starting point prime editor, such as PE1 or PE2). As exemplified in the Examples and described herein, various strategies can be used to obtain variant or engineered protein components, such as variant napDNAbp domain and variant RT domains, such as the PACE and PANCE evolution methods, and substitution of domains with replacement homologous domains (e.g., see representation of FIG. 27A).

[0171] The present disclosure describes improved prime editor systems, including prime editor proteins, which comprises an engineered Cas9 domain, an engineered reverse transcriptase domain, or a combination of an engineered Cas9 domain and an engineered reverse transcriptase domain. In the case of a prime editor system, the components of the prime editor (i.e., the Cas9 domain and the RT domain) can be provide as individual elements

(i.e., uncoupled or unfused). In the case of a prime editor fusion protein, the prime editor components (i.e., the Cas9 domain and the RT domain) are provided as a fusion protein.

[0172] In various embodiments, the engineered Cas9 domain of the herein disclosed prime editor system or fusion protein can comprise a variant Cas9 sequence of SEQ ID NO: 178, SEQ ID NO: 179, or SEQ ID NO: 180, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 178, SEQ ID NO: 179, or SEQ ID NO: 180, provided the amino acid sequence comprises at least one substitution selected from the group consisting of D23G, H99Q, H99R, E102K, E102S, E102R, N175K, D177G, K218R, N309D, I312V, E471K, G485S, K562N, D608N, I632V, D645N, D645E, R654C, G687D, G715E, H721Y, R753K, R753G, H754R, K775R, E790K, T804A, K918A, K1003R, M1021Y, E1071K, and E1260D relative to wild type Cas9.

[0173] In various embodiments, the prime editor systems or fusion proteins provided herein may comprise a nucleic acid-programmable DNA-binding protein (napDNAbp) and a mouse mammary tumor virus (MMTV) reverse transcriptase or a variant thereof, an avian sarcoma leukosis virus (ASLV) reverse transcriptase or a variant thereof, a porcine endogenous retrovirus (PERV) reverse transcriptase or a variant thereof, an HIV-MMLV reverse transcriptase or a variant thereof, an AVIRE reverse transcriptase or a variant thereof, a baboon endogenous virus (BAEVM) reverse transcriptase or a variant thereof, a gibbon ape leukemia virus (GALV) reverse transcriptase or a variant thereof, a koala retrovirus (KORV) reverse transcriptase or a variant thereof, a Mason-Pfizer monkey virus (MPMV) reverse transcriptase or a variant thereof, a POK11ERV reverse transcriptase or a variant thereof, a simian retrovirus type 2 (SRV2) reverse transcriptase or a variant thereof, a woolly monkey sarcoma virus (WMSV) reverse transcriptase or a variant thereof, a Vp96 reverse transcriptase or a variant thereof, a Vc95 reverse transcriptase or a variant thereof, an Ec48 reverse transcriptase or a variant thereof, a Gs reverse transcriptase or a variant thereof, an Er reverse transcriptase or a variant thereof, an Ne144 reverse transcriptase or a variant thereof, a Tf1 reverse transcriptase or a variant thereof, or an Rs09415 reverse transcriptase ("CRISPR-RT") or a variant thereof.

[0174] In various other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on MMLV RT wildtype of SEQ ID NO: 33 and can include the variants of SEQ ID NOs: 172-177 or 183-184, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 172-177 or 183-184, wherein the amino acid sequence comprises at least one of residues 13I, 19I, 32T, 38V, 60Y, 111L, 120R, 126Y, 128N, 128F, 128H, 129S, 132S, 138R, 157F, 175Q, 175S, 200S, 200Y, 200N, 200C, 222F, 223A, 223M, 223T, 223W, 223Y, 234I, 246I, 249S, 287A, 292T, 302A, 302K, 306K, 316R, 346K, 373N, 388C, 402A, 445N, 457I, and 462S.

[0175] In still various other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Ec48 RT and can include the variants of SEQ ID NOs: 188-195, 256, and 257 or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 188-195, 256, and 257, wherein the amino acid sequence comprises at least one of residues 36V, 54K, 60K, 87E, 151T, 165D, 182N, 189N, 205K, 214L, 243N, 267I, 277F, 279K, 303M, 307R, 315K, 317S, 318E, 324Q, 326E, 328K, 343N, 372K, 378K, and 385.

[0176] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Tf1 RT and can include the variants of SEQ ID NOs: 196-213 and 251-255, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 196-213 and 251-255, wherein the amino acid sequence comprises at least one of residues 14A, 22K, 64L, 64W, 70T, 72V, 102I, 106R, 118R, 133N, 139T, 158Q, 188K, 260L, 269L, 274R, 288Q, 293K, 297Q, 316Q, 321R, 356E, 363V, 413E, 423V, and 492N.

[0177] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on PERV RT and can include the variants of SEQ ID NOs: 214-215 or 234-238, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 214-215 or 234-238, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 602W.

[0178] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on AVIRE RT wildtype (SEQ ID NO: 216) and can include the variants of SEQ ID NOs: 217-221, or an

amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 217-221, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 604W.

[0179] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on KORV RT wildtype (SEQ ID NO: 222) and can include the variants of SEQ ID NOs: 223-227, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 223-227, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 310F, 327P, and 599W.

[0180] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on WMSV RT wildtype (SEQ ID NO: 228) and can include the variants of SEQ ID NOs: 229-233, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 229-233, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 311F, 327P, and 599W.

[0181] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Ne144 RT wildtype (SEQ ID NO: 239) and can include the variants of SEQ ID NO: 240, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 240, wherein the amino acid sequence comprises at least one of residues 157T, 165T, and 288V.

[0182] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence based on Vc95 RT wildtype (SEQ ID NO: 241) and can include the variant of SEQ ID NO: 242, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NO: 242, wherein the amino acid sequence comprises at least one of residues 11M, 75A, 97M, 146D, and 245T.

[0183] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor systems or fusion proteins can comprise a variant RT sequence based on Gs RT

wildtype (SEQ ID NO: 60), or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 159-171, wherein the amino acid sequence comprises at least one of residues 12D, 16E, 16V, 17P, 20G, 37R, 37P, 38H, 40C, 41N, 41S, 45R, 67T, 67R, 72E, 73V, 78V, 93R, 123V, 126F, 129G, 162N, 190L, 206V, 233K, 234V, 263G, 264S, 267M, 279E, 287I, 291K, 309T, 344S, 358S, 360S, 363G, 374A, and 412H.

[0184] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a pentamutant variant RT sequence based on AVIRE RT, KORV RT, and WMSV RT and can include the variants of SEQ ID NOs: 243-245, or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 243-245, wherein the AVIRE RT comprises the residues 199N, 305K, 312F, 329P, and 604W, the KORV RT comprises the residues 197N, 303K, 310F, 327P, and 599W, and the WMSV RT comprises the residues 197N, 303K, 311F, 327P, and 599W.

[0185] In yet other embodiments, the engineered RT domain of the herein disclosed prime editor system or fusion protein can comprise a variant RT sequence of Tfl-rat4 (SEQ ID NO: 251), Tflevo3.1 (SEQ ID NO: 252), Tflevo+rat-1 (SEQ ID NO: 254), Tf1evo+rat2 (SEQ ID NO: 255), Ec48-v2 (SEQ ID NO: 256), Ec48-evo3 (SEQ ID NO: 257) , or an amino acid sequence having 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%, or up to 100% sequence identity with any of SEQ ID NOs: 251-257, provided the sequences comprise at least one of the amino acid substitutions provided in the present disclosure.

[0186] In other embodiments, the present disclosure describes improved prime editors and prime editor systems, including prime editor fusion proteins, including PEmax of SEQ ID NO: 2, which may be encoded by a nucleic acid sequence of SEQ ID NO: 1, and which may be modified with any one of the herein disclosed variant Cas9 domains or variant RT domains. The present disclosure also provides other improved prime editor variants, including fusion proteins of SEQ ID NOs: 2-8 and fusion proteins comprising evolved nucleic acid programmable DNA binding proteins of SEQ ID NOs: 9-32 and reverse transcriptases of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241. The disclosure also contemplates fusion proteins having an amino acid sequence with a sequence identity of 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 up to 100% with SEQ ID NO: 2 and any one of SEQ ID NOs: 3-8. The disclosure also contemplates evolved nucleic acid programmable DNA binding proteins having an amino acid sequence with a sequence identity of 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 up to 100% with any one of SEQ ID NOs: 9-32. Further, the disclosure contemplates reverse transcriptases having an amino acid sequence with a sequence identity of 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 up to 100% with any one of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241.

[0187] In addition, the instant specification provides for nucleic acid molecules encoding and/or expressing the evolved and/or modified prime editors as described herein, as well as expression vectors or constructs for expressing the evolved and/or modified prime editors described herein, host cells comprising said nucleic acid molecules and expression vectors, and compositions for delivering and/or administering nucleic acid-based embodiments described herein. In addition, the disclosure provides for isolated evolved and/or modified prime editors, as well as compositions comprising said isolated evolved and/or modified prime editors as described herein. Still further, the present disclosure provides for methods of making the evolved and/or modified prime editors, as well as methods of using the evolved and/or modified prime editors or nucleic acid molecules encoding the evolved and/or modified prime editors in applications including editing a nucleic acid molecule, e.g., a genome, with improved efficiency as compared to prime editor that forms the state of the art, preferably in a sequence-context agnostic manner (i.e., wherein the desired editing site does not require a specific sequence-context). In embodiments, the method of making provide herein is an improved phage-assisted continuous evolution (PACE) system which may be utilized to evolve one or more components of a prime editor (e.g., a Cas9 domain or a reverse transcriptase domain). The specification also provides methods for efficiently editing a target nucleic acid molecule, e.g., a single nucleobase of a genome, with a prime editing system described herein (e.g., in the form of an isolated evolved and/or modified prime editor as described herein or a vector or construct encoding same) and conducting prime editing, preferably in a sequence-context agnostic manner. Still further, the specification provides therapeutic methods for treating a genetic disease and/or for altering or changing a genetic trait or condition by contacting a target nucleic acid molecule, e.g., a genome, with a prime editing system (e.g., in the form of an isolated evolved and/or modified prime editor protein

or a vector encoding same) and conducting prime editing to treat the genetic disease and/or change the genetic trait (e.g., eye color).

[0188] Accordingly, the present disclosure provides a method for editing a nucleic acid molecule by prime editing that involves contacting a nucleic acid molecule with a modified prime editor and a pegRNA, thereby installing one or more modifications to the nucleic acid molecule at a target site with increased editing efficiency and/or lower indel formation. The present disclosure further provides polynucleotides for editing a DNA target site by prime editing comprising a nucleic acid sequence encoding a modified prime editor protein comprising a modified napDNAbp and/or polymerase domain, wherein the napDNAbp and polymerase domains are capable in the presence of a pegRNA of installing one or more modifications in the DNA target site with increased editing efficiency and/or lower indel formation. The disclosure further provides, vectors, cells, and kits comprising the compositions and polynucleotides of the disclosure, as well as methods of making such vectors, cells, and kits, as well as methods for delivery of such compositions, polynucleotides, vectors, cells and kits to cells *in vitro, ex vivo* (e.g., during cell-based therapy which modify cells outside of the body), and *in vivo*.

Modified prime editors

[0189] The present disclosure provides modified prime editors and prime editor fusion proteins, such as, but not limited to PEmax, and can further include variants of PEmax where one or both of the napDNAbp and RT domains have been replaced with one of the herein disclosed engineered Cas9 or RT variants.

[0190] In one embodiment, the modified prime editor fusion protein is PEmax (of SEQ ID NO: 2), or an amino acid sequence having 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 up to 100% sequence identify with SEQ ID NO: 2. PEmax has the amino acid sequence of SEQ ID NO: 2, and the nucleic acid sequence of SEQ ID NO: 1.

[0191] PEmax (of SEQ ID NO: 2) includes from the N-terminal to C-terminal ends (a) a bipartite SV40 NLS domain (SEQ ID NO: 95), (b) an SpCas9 based on wildtype SpCas9 of SEQ ID NO: 10 with amino acid substitutions at R221K, N394K, and H840A relative to said sequence, (c) a linker sequence, (d) a Genscript codon optimized MMLV RT pentamutant based on wildtype MMLV RT of SEQ ID NO: 33 with amino acid substitutions at D200N T306K W313F T330P L603W relative to said sequence, (e) a linker, (f) a bipartite SV40

NLS domain, (g) a linker, and (h) a c-Myc NLS domain. These amino acid sequences are provided as follows:

PEmax component sequences of SEQ ID NO: 2:

- Bipartite SV40 NLS
MKRTADGSEFESPKKKRKV (SEQ ID NO: 95)

- SpCas9 R221K N394K H840A
DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG
ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRKL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR
KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA
VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY
LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
11)

- Linker = (SGGSx2–*bipartite SV40 NLS*–SGGSx2)
SGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGS (SEQ ID NO: 79)

- Genscript codon optimized MMLV RT pentamutant (D200N T306K W313F T330P L603W)
TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT
STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRP
VQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQP
LFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ
YVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLK
EGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLT
KPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT
QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL

APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE
GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTE
TEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI
YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA
DQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 34)

- *Other linker sequences*

SGGS (SEQ ID NO: 81)

- Bipartite SV40 NLS

KRTADGSEFESPKKKRKV (SEQ ID NO: 97)

- *Other linker sequences*

GSG (SEQ ID NO: 82)

- c-Myc NLS

PAAKRVKLD (SEQ ID NO: 98)

[0192] The prime editors contemplated herein comprise, in some embodiments, systems wherein the nucleic acid programmable DNA binding protein (napDNAbp) and the reverse transcriptase domain (RT) are provided *in trans* such that they are capable of being separately localized and/or targeted to a DNA edit site of interest to carry of their prime editing function. In other embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) and the reverse transcriptase domain (RT) are provided as a fusion protein.

[0193] In those embodiments where the nucleic acid programmable DNA binding protein (napDNAbp) and the reverse transcriptase domain (RT) are provided in the form of a fusion protein, the modified prime editors disclosed herein may comprise any suitable structural configuration. For example, the fusion protein may comprise from the N-terminus to the C-terminus direction, a napDNAbp fused to a polymerase (e.g., DNA-dependent DNA polymerase or RNA-dependent DNA polymerase, such as, reverse transcriptase). In other embodiments, the fusion protein may comprise from the N-terminus to the C-terminus direction, a polymerase (e.g., a reverse transcriptase) fused to a napDNAbp. The fused domain may optionally be joined by a linker, e.g., an amino acid sequence. In other embodiments, the fusion proteins may comprise the structure NH_2-[napDNAbp]-[polymerase]-COOH; or NH_2-[polymerase]-[napDNAbp]-COOH, wherein each instance of "]-[" indicates the presence of an optional linker sequence. In embodiments wherein the polymerase is a reverse transcriptase, the fusion proteins may comprise the structure NH_2-[napDNAbp]-[RT]-COOH; or NH_2-[RT]-[napDNAbp]-COOH, wherein each instance of "]-[" indicates the presence of an optional linker sequence.

[0194] In various embodiments, the modified prime editors may be based on PE1, wherein one or more components of PE1 are substituted with a variant domain. For example the PE1 SpCas9 domain may be exchanged with a modified SpCas9 domain. Or, the RT domain may be exchanged with a modified RT domain (e.g., a codon-optimized variant).

[0195] PE1 includes a Cas9 variant comprising an H840A mutation (i.e., a Cas9 nickase) and an M-MLV RT wild type, as well as an N-terminal NLS sequence (19 amino acids) and an amino acid linker (32 amino acids) that joins the C-terminus of the Cas9 nickase domain to the N-terminus of the RT domain. The PE1 fusion protein has the following structure: [NLS]-[Cas9(H840A)]-[linker]-[MMLV_RT(wt)]. The amino acid sequence of PE1 and its individual components are as follows:

DESCRIPTION	SEQUENCE
PE1 FUSION PROTEIN CAS9(H840A)-MMLV_RT(WT)	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKV PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINA SGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDK LIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD**SGGSSGGSSGSETPG TSESATPESSGGSSGGSS*TLNIEDEYRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQR*

DESCRIPTION	SEQUENCE
	*LLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPE MGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVD DLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLL KEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEM AAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELF VDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLL DTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAE LIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKN KDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAIT ETPDTSTLLIENSSP*SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 3) KEY: NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96) **CAS9(H840A)** (SEQ ID NO: 10) **33-AMINO ACID LINKER** (SEQ ID NO: 80) *M-MLV REVERSE TRANSCRIPTASE* (SEQ ID NO: 33)
PE1 – N-TERMINAL NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 95)
PE1 – CAS9 (H840A) (MET MINUS)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQL VQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED

DESCRIPTION	SEQUENCE
	NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 10)
PE1 – LINKER BETWEEN CAS9 DOMAIN AND RT DOMAIN (33 AMINO ACIDS)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 80)
PE1 – M-MLV RT	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQ APLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLP PSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT RLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSE LDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQ RWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAP LYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFV DEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVA AIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGT RPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAK ALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSA EARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 33)
PE1 – C-TERMINAL NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 96)

[0196] In various other embodiments, the modified prime editor proteins may be based on PE2, wherein one or more components of PE2 are substituted with a variant domain. For example the PE2 SpCas9 domain may be exchanged with a modified SpCas9 domain. Or, the RT domain of PE2 may be exchanged with a modified RT domain (e.g., a codon-optimized variant).

[0197] PE2 includes a Cas9 variant comprising an H840A mutation (i.e., a Cas9 nickase) and an M-MLV RT comprising mutations D200N, T330P, L603W, T306K, and W313F, as well as an N-terminal NLS sequence (19 amino acids) and an amino acid linker (33 amino acids) that joins the C-terminus of the Cas9 nickase domain to the N-terminus of the RT domain. The PE2 fusion protein has the following structure: [NLS]-[Cas9(H840A)]-[linker]-[MMLV_RT(D200N)(T330P)(L603W)(T306K)(W313F)]. The amino acid sequence of PE2 is as follows:

PE2 FUSION PROTEIN CAS9(H840A)-MMLV_RT	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR**

D200N T330P L603W T306K W313F	**LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA** **KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV** **NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ** **SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR** **KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFR** **IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE** **RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA** **FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE** **DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM** **IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG** **KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE** **HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ** **TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL** **YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR** **SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA** **ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK** **LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG** **TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS** **NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV** **LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY** **GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP** **IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG** **NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE** **IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT** **NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS** **QLGGD**<u>**SGGSSGGSSGSETPGTSESATPESSGGSSGGSS**</u>*TLNIEDEYRLH* *ETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ* *YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD* *LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQ* *PLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP* *DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQV* *KYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPG* *FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE* *LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAV* *LTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD* *RVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH* *TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK* *MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKA* *LFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENS* *SP*<u>SGGSKRTADGSEFEPKKKRKV</u> (SEQ ID NO: 4) KEY: <u>NUCLEAR LOCALIZATION SEQUENCE (NLS)</u> TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96) **CAS9(H840A)** (SEQ ID NO: 10) **<u>33-AMINO ACID LINKER</u>** (SEQ ID NO: 80) *M-MLV REVERSE TRANSCRIPTASE* (SEQ ID NO: 34)
PE2 – N-TERMINAL NLS	MKRTADGSEFESPKKKRKV (SEQ ID NO: 95)

PE2 – CAS9 (H840A) (MET MINUS)	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD (SEQ ID NO: 10)
PE2 – LINKER BETWEEN CAS9 DOMAIN AND RT DOMAIN (33 AMINO ACIDS)	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 80)
PE2 – MMLV_RT D200N T330P L603W T306K W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSP TLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQ PTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLL PLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEG QRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLN

	VYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 34)
PE2 – C-TERMINAL NLS	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 96)

[0198] In still other embodiments, the modified prime editor proteins disclosed herein may be based on other prime editor protein sequences, wherein one or more components of such fusion are substituted with a variant domain. Such starting point prime editor proteins may include:

PE FUSION PROTEIN MMLV_RT(WT)-32AA-CAS9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSD FPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLI LLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQ VKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGW PPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILA EAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVI WAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHG EIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSA EARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGSSGGSSGSETP GTSESATPESSGGSSGGSS**DKKYSIGLDIGTNSVGWAVITDEYK VPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD

	FRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQA ENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGDSGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 5) KEY: NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96) **CAS9(H840A)** (SEQ ID NO: 10) **33-AMINO ACID LINKER** (SEQ ID NO: 80) *M-MLV REVERSE TRANSCRIPTASE* (SEQ ID NO: 33)
PE FUSION PROTEIN MMLV_RT(WT)-60AA-CAS9(H840A)	MKRTADGSEFESPKKKRKV*TLNIEDEYRLHETSKEPDVSLGSTWLSD FPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLI LLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQ VKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGW PPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILA EAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVI WAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHG EIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSA EARGNRMADQAARKAAITETPDTSTLLIENSSP**SGGSSGGSSGSETP GTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSS GGS**DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDR HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQ EIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSE ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKT NRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL

	KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK LYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSID NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GDSGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 6) KEY: NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96) **CAS9(H840A)**(SEQ ID NO: 10) **AMINO ACID LINKER** (SEQ ID NO: 83) *M-MLV REVERSE TRANSCRIPTASE* (SEQ ID NO: 33)
PE FUSION PROTEIN CAS9(H840A)-FEN1-MMLV_RT D200N T330P L603W T306K W313F	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDEY KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF DKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQ LVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMP**

	QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL IHQSITGLYETRIDLSQLGGD<u>SGGSSGGSSGSETPGTSESATPES SGGSSGGSS</u>GIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIKPVY VFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQEVEKFT KRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGK VYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELG LNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLD PNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPNEEELIKF MCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSLSSAK RKEPEPKGSTKKKAKTGAAGKFKRGK**<u>SGGSSGGSSGSETPGTSE SATPESSGGSSGGSS</u>***TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQA WAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRL LDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYL GYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDL TKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNA RMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWA KALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIY RRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA RGNRMADQAARKAAITETPDTSTLLIENSSP**<u>SGGSKRTADGSEFEPKK KRKV</u> (SEQ ID NO: 7) KEY: <u>NUCLEAR LOCALIZATION SEQUENCE (NLS)</u> TOP:(SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 96) **CAS9(H840A)** (SEQ ID NO: 10) **<u>33-AMINO ACID LINKER 1</u>** (SEQ ID NO: 80) *M-MLV REVERSE TRANSCRIPTASE* (SEQ ID NO: 34) **<u>33-AMINO ACID LINKER 2</u>** (SEQ ID NO: 80) FEN1 (SEQ ID NO: 111)

[0199] In still other embodiments, the prime editors used in the present disclosure may comprise PEmax. PEmax is a complex comprising a fusion protein comprising Cas9(R221K N39K H840A) and a variant MMLV RT pentamutant (D200N T306K W313F T330P L603W) having the following structure: [bipartite NLS]-[Cas9(R221K)(N394K)(H840A)]-[linker]-[MMLV_RT(D200N)(T330P)(L603W)]-[bipartite NLS]-[NLS] + a desired

PEgRNA, wherein the PE fusion has the amino acid sequence of SEQ ID NO: 2, which is shown as follows:

PEmax fusion protein [bipartite NLS]-[Cas9(R221K)(N394K)(H840A)]-[linker]-[MMLV_RT(D200N)(T330P)(L603W)]-[bipartite NLS]-[NLS]	MKRTADGSEFESPKKKRKV**DKKYSIGLDIGTNSVGWAVITDE** **YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK** **RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV** **EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK** **ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ** **TYNQLFEENPINASGVDAKAILSARLSKSRKLENLIAQLPGE** **KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD** **DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA** **PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK** **NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKRED** **LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREK** **IEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV** **VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN** **ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK** **QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK** **DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK** **VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD** **GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL** **AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ** **TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE** **KLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDD** **SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK** **LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV** **AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF** **YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY** **KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN** **GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV** **KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS** **PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN** **PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG** **ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV** **EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP** **IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL** **DATLIHQSITGLYETRIDLSQLGGD**<u>**SGGSSGGSKRTADGSEF** **ESPKKKRKVSGGSSGGS**</u>TLNIEDEYRLHETSKEPDVSLGSTWL SDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEA RLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQD LREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFC LRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNE ALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALL QTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKP GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEK QGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVA AIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDIL AEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVT

	TETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDS RYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI ENSSP*SGGS*KRTADGSEFESPKKKRKV*GSGPAAKRVKLD* (SEQ ID NO: 2) KEY: BIPARTITE SV40 NUCLEAR LOCALIZATION SEQUENCE (NLS) TOP: (SEQ ID NO: 95), BOTTOM: (SEQ ID NO: 97) **CAS9(R221K N394K H840A)** (SEQ ID NO: 11) **SGGSx2-BIPARTITE SV40NLS-SGGSx2 LINKER** (SEQ ID NO: 79) M-MLV reverse transcriptase(D200N T306K W313F T330P L603W) (SEQ ID NO: 34) *Other linker sequences* (SEQ ID NOs: 81 and 82) *C-MYC NLS* (SEQ ID NO: 98)

[0200] In various embodiments, the prime editor proteins utilized in the methods and compositions contemplated herein may also include any variants of the above-disclosed sequences having an amino acid sequence that is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any of the herein disclosed prime editor sequences.

napDNAbp domain and modified variants thereof

[0201] In various embodiments, the modified prime editor proteins disclosed herein, including PEmax, comprise a nucleic acid programmable DNA binding protein (napDNAbp).

[0202] In various embodiments, the modified prime editor proteins may include a napDNAbp domain having a wild type Cas9 sequence, including, for example the canonical *Streptococcus pyogenes* Cas9 sequence of SEQ ID NO: 9.

[0203] In other embodiments, the modified prime editor proteins may include a napDNAbp domain having a modified Cas9 sequence, including, for example the nickase variant of *Streptococcus pyogenes* Cas9 of SEQ ID NO: 12 having an H840A substitution relative to the wild type SpCas9 (of SEQ ID NO: 9), shown as follows:

Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with H840**A**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS	SEQ ID NO: 12

	ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGD	

[0204] In an embodiment modified prime editor referred to as “PEmax” the napDNAbp component or domain comprises the following amino acid sequence, which is based on the canonical SpCas9 amino acid sequence of SEQ ID NO: 9 with the following substitutions: R221K, N394K, and H840A.

SpCas9 R221K N394K H840A:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRKLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN

FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 11)

[0205] The modified prime editor proteins may further comprise one or more mutations in the napDNAbp (e.g., Cas9) domain that result in improved editing efficiency. For example, the present disclosure describes the development of improved prime editor proteins using PACE. In some embodiments, a prime editor (*e.g.*, a fusion protein, or a prime editor in which the napDNAbp and reverse transcriptase are provided *in trans*) comprises a Cas9 variant comprising one or more mutations relative to SEQ ID NO: 9 selected from the group consisting of D23G, H99Q, H99R, E102K, E102S, E102R, N175K, D177G, K218R, N309D, I312V, E471K, G485S, K562N, D608N, I632V, D645N, D645E, R654C, G687D, G715E, H721Y, R753K, R753G, H754R, K775R, E790K, T804A, K918A, K1003R, M1021Y, E1071K, and E1260D. In some embodiments, such a Cas9 variant comprises a single mutation, wherein the single mutation is selected from D23G, H99Q, H99R, E102K, E102S, E102R, N175K, D177G, K218R, N309D, I312V, E471K, G485S, K562N, D608N, I632V, D645N, D645E, R654C, G687D, G715E, H721Y, R753K, R753G, H754R, K775R, E790K, T804A, K918A, K1003R, M1021Y, E1071K, and E1260D. In some embodiments, the Cas9 variant comprises an R753G mutation. In certain embodiments, the Cas9 variant comprises an H721Y mutation and an R753G mutation; an E102K mutation and an R753G mutation; or an E102K mutation, an H721Y mutation, and an R753G mutation. In certain embodiments, the Cas9 variant comprises the amino acid sequence of any one of SEQ ID NOs: 178-180.

[0206] In some embodiments, the improved prime editor proteins used in the compositions and methods described herein comprise a mutation at the position R753X, wherein X is any amino acid, relative to the amino acid sequence of wild-type Cas9 from *Streptococcus pyogenes*:

Description	Sequence	SEQ ID NO:

SpCas9 *Streptococcus pyogenes* with R753**X** Wherein "X" is any amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMG**X**HKPENIVIEM ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV VKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	13

[0207] In some embodiments, the R753X mutation is an R753G mutation:

Description	Sequence	SEQ ID NO:
SpCas9 *Streptococcus pyogenes* with R753**G**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH	SEQ ID NO: 14

	AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMG**G**HKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDY DVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP LIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD	

[0208] The improved prime editor proteins utilized in the methods and compositions described herein may include any of the modified Cas9 sequences described above, or any variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto, provided the variant comprises one of the amino acid substitutions provided herein. The proteins described herein may also include any Cas9 protein (e.g., including the ones described below) comprising a mutation corresponding to R753X or R753G at a relevant position in the amino acid sequence.

[0209] The present disclosure contemplates the modification of any Cas9 protein known in the art with one or more of the mutations described herein (*i.e.*, R221K, N394K, R753G, and/or H840A) and the combination of any modified Cas9 protein with one or more of the PEmax architecture features described herein (*e.g.*, the optimized MMLV RT pentamutant, NLS's, linkers, *etc.*).

[0210] In some embodiments, the improved prime editor proteins described herein include any of the following other wild type SpCas9 sequences, which may be modified with one or more of the mutations described herein at corresponding amino acid positions:

Description	Sequence

SpCas9 *Streptococcus pyogenes* MGAS1882 wild type NC_017053.1	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAG CGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAA AAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAA AAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGG AAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGA TGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTT TTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTG GAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTA TCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCG TGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGAT GTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTA TTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGC GATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCT CATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAA TCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAAT TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACT TACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAA TATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTT TACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTC CCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAG ACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAA AGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAG GTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTA TCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGG TGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATG CTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACA ATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATT ATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGA CTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAG TTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGA CAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAAC ATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAA GGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTC AGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAA TCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAA AAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATA GATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTAT TAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTT AGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAA GGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACG TTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGG CAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCG CAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGA AGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACA TGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGG TATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAAT

	GGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAA ATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATG AAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCT TAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCT CTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCA AGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACAT TGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGT ACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCC AAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAAC TTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAA CGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTT TTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATG TGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAA ATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTA AATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTAC GTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATG CCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAAT CGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAAT GATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAAT ATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAA TGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTG CCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATT TTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGAC TGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCT TATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAA GAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGA AAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAA AGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTA AATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCA AGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGT TGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTG GAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGT GAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAA GTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAA CAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGA GCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAAC GATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATC AATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGC TAGGAGGTGACTGA (SEQ ID NO: 15)
SpCas9 *Streptococcus pyogenes* MGAS1882 wild type NC_017053.1	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKN LIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL

	EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPEN IVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSID NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFD NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 16)
SpCas9 *Streptococcus pyogenes* wild type SWBC2D7W014	ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCC GTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAG AAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAG AATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAG GCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCG CAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGAT GGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTC CTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGG AAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGA TTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCG TGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGA TGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTT GTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGG CTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACC TGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGT AACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCG AACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGAC ACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGAT CAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCA ATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAG GCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCAC CAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCT GAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTA CGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACA AGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAG TTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGG ACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAA

	TTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCA AAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATAC CTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCAT GGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTT GAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAG AGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATT GCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGA ACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGC CTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATT CAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACT ACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGG TAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCC TAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAAT GAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAA GATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCT GTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATAC GGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAG ACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACG GCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTT AACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAG GGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAG CCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAG CTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATC GAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAA ACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAA CTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCA ATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAG GGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGA TTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGA TTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGG GAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGA AGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAA AGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCT GAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACC CGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATG AATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAA AGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGA TTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGC GCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAA GAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAA AGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGA TAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGA ATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCA AACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTA TGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCC ATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGG AGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAA GCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTG GCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAA AAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAA

	TTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAAC CCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAA GGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGA AAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAA AGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGT ATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATA ACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCG ACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCC TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGC ACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCAT TTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATT TTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAG GTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATAT GAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCC AAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAA GGCTGCAGGA (SEQ ID NO: 17)
SpCas9 *Streptococcus pyogenes* wild type Encoded product of SWBC2D7W014	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL VDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSP KKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAG (SEQ ID NO: 18)
SpCas9 *Streptococcus pyogenes*	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAG CGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAA AAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAA

M1GAS wild type NC_002737.2	AAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGG AAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTC GGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGA TGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTT TTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTG GAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTA TCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGG ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCG TGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGAT GTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTA TTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGC GATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCT CATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAA TCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAAT TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACT TACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAA TATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTT TACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTC CCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAG ACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAA AGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAG GTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTA TCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGG TGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTG ACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATG CTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACA ATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATT ATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGA CTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAG TTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGA CAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAAC ATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAA GGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTC AGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAA TCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAA AAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATA GATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTAT TAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTT AGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGAT GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAA GGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACG TTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGG CAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCG CAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGA AGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTAC ATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAG GTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAA TGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGT GAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCG TATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGA

	TTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAA AGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGG ACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATC ACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATA AGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACG TTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAAT TTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCT GGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAG CATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGAT GAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAA TCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAG TACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAA ATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTG AATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAA AATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAAT TACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAAC TAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATT TTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTG TCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCA ATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAA GACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTA GCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCG AAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATG GAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCT AAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACC TAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCT GGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGC CAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAA AGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTT GTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATC AGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATA AAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTG AACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTG GAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAA ACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCA TCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCA GCTAGGAGGTGACTGA (SEQ ID NO: 19)
SpCas9 *Streptococcus pyogenes* M1GAS wild type Encoded product of NC_002737.2 (100% identical to the canonical Q99ZW2	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL VDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPIL EKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY

wild type)	NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL SRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 258)

[0211] The improved prime editor proteins utilized in the methods and compositions described herein may include any of the above SpCas9 sequences, or any variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

[0212] In other embodiments, the Cas9 protein can be a wild type Cas9 ortholog from another bacterial species different from the canonical Cas9 from *S. pyogenes*. For example, modified versions of the following Cas9 orthologs can be used in connection with the PEmax constructs utilized in the methods and compositions described in this specification by making mutations at positions corresponding to R221K, N394K, R753G, and/or H840A in wild type SpCas9. In addition, any variant Cas9 orthologs having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any of the below orthologs may also be used with the prime editors.

Description	Sequence
LfCas9 *Lactobacillus fermentum* wild type *GenBank: SNX31424.1* 1	MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERR TFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPTK NQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVMNIYKLRE AMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFN VLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLE VKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSS ETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRY WTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKERGFDLEKGL KKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQ AKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQK

Description	Sequence
	ATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLL NEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKT VKASDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDK VDDNRYQTDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRY KGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFKEIVDQPVFKEQIDQL NQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVG NAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEK APDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDNSL DNRVLTSRKENNKKSDQVPAKLYAAKMKPYWNQLLKQGLITQRKFEN LTKDVDQNIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTEIIETR AGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSF FVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTR DEVAKSFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEI KKNRLVDLYGAYTNGTSAFMTIIKFTGNKPKYKVIGIPTTSAASLKRAGK PGSESYNQELHRIIKSNPKVKKGFEIVVPHVSYGQLIVDGDCKFTLASPTV QHPATQLVLSKKSLETISSGYKILKDKPAIANERLIRVFDEVVGQMNRYF TIFDQRSNRQKVADARDKFLSLPTESKYEGAKKVQVGKTEVITNLLMGL HANATQGDLKVLGLATFGFFQSTTGLSLSEDTMIVYQSPTGLFERRICLK DI (SEQ ID NO: 20)
SaCas9 *Staphylococcus aureus* wild type GenBank: AYD60528.1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLY

Description	Sequence
	LASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 259)
SaCas9 *Staphylococcus aureus*	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRS KRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQ KLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEE KYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELR SVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKP TLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSL KAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP VVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNR QTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP FNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISY ETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYK HHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGN TLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDIT DDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEV NSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIE VNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSK KHPQIIKK (SEQ ID NO: 21)
StCas9 *Streptococcus thermophilus* UniProtKB/ Swiss-Prot: G3ECR1.2 Wild type	MLFNKCIIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYKVPSK KMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRI LYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYH DEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKN NDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLF PGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLL GYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEH KEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLK NLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQA KFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNF EDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVR FIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIE LKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIK QRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLI DDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAI KKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLK RLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTG DDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDFPSLEVV KKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETR

Description	Sequence
	QITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYK VREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERK SATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLA TVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNEN LVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISI LDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILSTN NKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEEL FYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKG LFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRID LAKLGEG (SEQ ID NO: 22)
LcCas9 *Lactobacillus crispatus* NCBI Reference Sequence: WP_133478044.1 Wild type	MKIKNYNLALTPSTSAVGHVEVDDDLNILEPVHHQKAIGVAKFGEGETA EARRLARSARRTTKRRANRINHYFNEIMKPEIDKVDPLMFDRIKQAGLSP LDERKEFRTVIFDRPNIASYYHNQFPTIWHLQKYLMITDEKADIRLIYWA LHSLLKHRGHFFNTTPMSQFKPGKLNLKDDMLALDDYNDLEGLSFAVA NSPEIEKVIKDRSMHKKEKIAELKKLIVNDVPDKDLAKRNNKIITQIVNAI MGNSFHLNFIFDMDLDKLTSKAWSFKLDDPELDTKFDAISGSMTDNQIGI FETLQKIYSAISLLDILNGSSNVVDAKNALYDKHKRDLNLYFKFLNTLPD EIAKTLKAGYTLYIGNRKKDLLAARKLLKVNVAKNFSQDDFYKLINKEL KSIDKQGLQTRFSEKVGELVAQNNFLPVQRSSDNVFIPYQLNAITFNKILE NQGKYYDFLVKPNPAKKDRKNAPYELSQLMQFTIPYYVGPLVTPEEQV KSGIPKTSRFAWMVRKDNGAITPWNFYDKVDIEATADKFIKRSIAKDSY LLSELVLPKHSLLYEKYEVFNELSNVSLDGKKLSGGVKQILFNEVFKKTN KVNTSRILKALAKHNIPGSKITGLSNPEEFTSSLQTYNAWKKYFPNQIDNF AYQQDLEKMIEWSTVFEDHKILAKKLDEIEWLDDDQKKFVANTRLRGW GRLSKRLLTGLKDNYGKSIMQRLETTKANFQQIVYKPEFREQIDKISQAA AKNQSLEDILANSYTSPSNRKAIRKTMSVVDEYIKLNHGKEPDKIFLMFQ RSEQEKGKQTEARSKQLNRILSQLKADKSANKLFSKQLADEFSNAIKKS KYKLNDKQYFYFQQLGRDALTGEVIDYDELYKYTVLHIIPRSKLTDDSQ NNKVLTKYKIVDGSVALKFGNSYSDALGMPIKAFWTELNRLKLIPKGKL LNLTTDFSTLNKYQRDGYIARQLVETQQIVKLLATIMQSRFKHTKIIEVR NSQVANIRYQFDYFRIKNLNEYYRGFDAYLAAVVGTYLYKVYPKARRL FVYGQYLKPKKTNQENQDMHLDSEKKSQGFNFLWNLLYGKQDQIFVN GTDVIAFNRKDLITKMNTVYNYKSQKISLAIDYHNGAMFKATLFPRNDR DTAKTRKLIPKKKDYDTDIYGGYTSNVDGYMLLAEIIKRDGNKQYGFYG VPSRLVSELDTLKKTRYTEYEEKLKEIIKPELGVDLKKIKKIKILKNKVPF NQVIIDKGSKFFITSTSYRWNYRQLILSAESQQTLMDLVVDPDFSNHKAR KDARKNADERLIKVYEEILYQVKNYMPMFVELHRCYEKLVDAQKTFKS LKISDKAMVLNQILILLHSNATSPVLEKLGYHTRFTLGKKHNLISENAVL VTQSITGLKENHVSIKQML (SEQ ID NO: 23)
PdCas9 *Pedicoccus damnosus*	MTNEKYSIGLDIGTSSIGFAVVNDNNRVIRVKGKNAIGVRLFDEGKAAA DRRSFRTTRRSFRTTRRRLSRRRWRLKLLREIFDAYITPVDEAFFIRLKES NLSPKDSKKQYSGDILFNDRSDKDFYEKYPTIYHLRNALMTEHRKFDVR EIYLAIHHIMKFRGHFLNATPANNFKVGRLNLEEKFEELNDIYQRVFPDE SIEFRTDNLEQIKEVLLDNKRSRADRQRTLVSDIYQSSEDKDIEKRNKAV

Description	Sequence
NCBI Reference Sequence: WP_062913273.1 Wild type	ATEILKASLGNKAKLNVITNVEVDKEAAKEWSITFDSESIDDDLAKIEGQ MTDDGHEIIEVLRSLYSGITLSAIVPENHTLSQSMVAKYDLHKDHLKLFK KLINGMTDTKKAKNLRAAYDGYIDGVKGKVLPQEDFYKQVQVNLDDS AEANEIQTYIDQDIFMPKQRTKANGSIPHQLQQQELDQIIENQKAYYPWL AELNPNPDKKRQQLAKYKLDELVTFRVPYYVGPMITAKDQKNQSGAEF AWMIRKEPGNITPWNFDQKVDRMATANQFIKRMTTTDTYLLGEDVLPA QSLLYQKFEVLNELNKIRIDHKPISIEQKQQIFNDLFKQFKNVTIKHLQDY LVSQGQYSKRPLIEGLADEKRFNSSLSTYSDLCGIFGAKLVEENDRQEDL EKIIEWSTIFEDKKIYRAKLNDLTWLTDDQKEKLATKRYQGWGRLSRKL LVGLKNSEHRNIMDILWITNENFMQIQAEPDFAKLVTDANKGMLEKTDS QDVINDLYTSPQNKKAIRQILLVVHDIQNAMHGQAPAKIHVEFARGEER NPRRSVQRQRQVEAAYEKVSNELVSAKVRQEFKEAINNKRDFKDRLFL YFMQGGIDIYTGKQLNIDQLSSYQIDHILPQAFVKDDSLTNRVLTNENQV KADSVPIDIFGKKMLSVWGRMKDQGLISKGKYRNLTMNPENISAHTENG FINRQLVETRQVIKLAVNILADEYGDSTQIISVKADLSHQMREDFELLKN RDVNDYHHAFDAYLAAFIGNYLLKRYPKLESYFVYGDFKKFTQKETKM RRFNFIYDLKHCDQVVNKETGEILWTKDEDIKYIRHLFAYKKILVSHEVR EKRGALYNQTIYKAKDDKGSGQESKKLIRIKDDKETKIYGGYSGKSLAY MTIVQITKKNKVSYRVIGIPTLALARLNKLENDSTENNGELYKIIKPQFTH YKVDKKNGEIIETTDDFKIVVSKVRFQQLIDDAGQFFMLASDTYKNNAQ QLVISNNALKAINNTNITDCPRDDLERLDNLRLDSAFDEIVKKMDKYFSA YDANNFREKIRNSNLIFYQLPVEDQWENNKITELGKRTVLTRILQGLHAN ATTTDMSIFKIKTPFGQLRQRSGISLSENAQLIYQSPTGLFERRVQLNKIK (SEQ ID NO: 24)
FnCas9 *Fusobacterium nucleatum* NCBI Reference Sequence: WP_060798984.1	MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFEE AKTAAERRVQRNSRRRLKRRKWRLNLLEEIFSNEILKIDSNFFRRLKESSL WLEDKSSKEKFTLFNDDNYKDYDFYKQYPTIFHLRNELIKNPEKKDIRLV YLAIHSIFKSRGHFLFEGQNLKEIKNFETLYNNLIAFLEDNGINKIIDKNNI EKLEKIVCDSKKGLKDKEKEFKEIFNSDKQLVAIFKLSVGSSVSLNDLFD TDEYKKGEVEKEKISFREQIYEDDKPIYYSILGEKIELLDIAKTFYDFMVL NNILADSQYISEAKVKLYEEHKKDLKNLKYIIRKYNKGNYDKLFKDKNE NNYSAYIGLNKEKSKKEVIEKSRLKIDDLIKNIKGYLPKVEEIEEKDKAIF NKILNKIELKTILPKQRISDNGTLPYQIHEAELEKILENQSKYYDFLNYEE NGIITKDKLLMTFKFRIPYYVGPLNSYHKDKGGNSWIVRKEEGKILPWNF EQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDTFLYSEYVILNELNKVQ VNDEFLNEENKRKIIDELFKENKKVSEKKFKEYLLVKQIVDGTIELKGVK DSFNSNYISYIRFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKIFEKKIK NEYGDILTKDEIKKINTFKFNNWGRLSEKLLTGIEFINLETGECYSSVMDA LRRTNYNLMELLSSKFTLQESINNENKEMNEASYRDLIEESYVSPSLKRAI FQTLKIYEEIRKITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYD SCGNDIANFSIDIKEMKNSLISYDNNSLRQKKLYLYYLQFGKCMYTGREI DLDRLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPV KKEIQEKMKSFWRFLKEKNFISDEKYKRLTGKDDFELRGFMARQLVNV RQTTKEVGKILQQIEPEIKIVYSKAEIASSFREMFDFIKVRELNDTHHAKD

Description	Sequence
	AYLNIVAGNVYNTKFTEKPYRYLQEIKENYDVKKIYNYDIKNAWDKEN SLEIVKKNMEKNTVNITRFIKEKKGQLFDLNPIKKGETSNEIISIKPKVYN GKDDKLNEKYGYYKSLNPAYFLYVEHKEKNKRIKSFERVNLVDVNNIK DEKSLVKYLIENKKLVEPRVIKKVYKRQVILINDYPYSIVTLDSNKLMDF ENLKPLFLENKYEKILKNVIKFLEDNQGKSEENYKFIYLKKKDRYEKNET LESVKDRYNLEFNEMYDKFLEKLDSKDYKNYMNNKKYQELLDVKEKFI KLNLFDKAFTLKSFLDLFNRKTMADFSKVGLTKYLGKIQKISSNVLSKNE LYLLEESVTGLFVKKIKL (SEQ ID NO: 25)
EcCas9 *Enterococcus cecorum* NCBI Reference Sequence: WP_047338501.1 Wild type	RRKQRIQILQELLGEEVLKTDPGFFHRMKESRYVVEDKRTLDGKQVELP YALFVDKDYTDKEYYKQFPTINHLIVYLMTTSDTPDIRLVYLALHYYMK NRGNFLHSGDINNVKDINDILEQLDNVLETFLDGWNLKLKSYVEDIKNIY NRDLGRGERKKAFVNTLGAKTKAEKAFCSLISGGSTNLAELFDDSSLKEI ETPKIEFASSSLEDKIDGIQEALEDRFAVIEAAKRLYDWKTLTDILGDSSS LAEARVNSYQMHHEQLLELKSLVKEYLDRKVFQEVFVSLNVANNYPAY IGHTKINGKKKELEVKRTKRNDFYSYVKKQVIEPIKKKVSDEAVLTKLSE IESLIEVDKYLPLQVNSDNGVIPYQVKLNELTRIFDNLENRIPVLRENRDK IIKTFKFRIPYYVGSLNGVVKNGKCTNWMVRKEEGKIYPWNFEDKVDLE ASAEQFIRRMTNKCTYLVNEDVLPKYSLLYSKYLVLSELNNLRIDGRPLD VKIKQDIYENVFKKNRKVTLKKIKKYLLKEGIITDDDELSGLADDVKSSL TAYRDFKEKLGHLDLSEAQMENIILNITLFGDDKKLLKKRLAALYPFIDD KSLNRIATLNYRDWGRLSERFLSGITSVDQETGELRTIIQCMYETQANLM QLLAEPYHFVEAIEKENPKVDLESISYRIVNDLYVSPAVKRQIWQTLLVIK DIKQVMKHDPERIFIEMAREKQESKKTKSRKQVLSEVYKKAKEYEHLFE KLNSLTEEQLRSKKIYLYFTQLGKCMYSGEPIDFENLVSANSNYDIDHIYP QSKTIDDSFNNIVLVKKSLNAYKSNHYPIDKNIRDNEKVKTLWNTLVSK GLITKEKYERLIRSTPFSDEELAGFIARQLVETRQSTKAVAEILSNWFPESE IVYSKAKNVSNFRQDFEILKVRELNDCHHAHDAYLNIVVGNAYHTKFTN SPYRFIKNKANQEYNLRKLLQKVNKIESNGVVAWVGQSENNPGTIATVK KVIRRNTVLISRMVKEVDGQLFDLTLMKKGKGQVPIKSSDERLTDISKY GGYNKATGAYFTFVKSKKRGKVVRSFEYVPLHLSKQFENNNELLKEYIE KDRGLTDVEILIPKVLINSLFRYNGSLVRITGRGDTRLLLVHEQPLYVSNS FVQQLKSVSSYKLKKSENDNAKLTKTATEKLSNIDELYDGLLRKLDLPIY SYWFSSIKEYLVESRTKYIKLSIEEKALVIFEILHLFQSDAQVPNLKILGLS TKPSRIRIQKNLKDTDKMSIIHQSPSGIFEHEIELTSL (SEQ ID NO: 26)
AhCas9 *Anaerostipes hadrus* NCBI Reference Sequence: WP_044924278.1	MQNGFLGITVSSEQVGWAVTNPKYELERASRKDLWGVRLFDKAETAED RRMFRTNRRLNQRKKNRIHYLRDIFHEEVNQKDPNFFQQLDESNFCEDD RTVEFNFDTNLYKNQFPTVYHLRKYLMETKDKPDIRLVYLAFSKFMKN RGHFLYKGNLGEVMDFENSMKGFCESLEKFNIDFPTLSDEQVKEVRDIL CDHKIAKTVKKKNIITITKVKSKTAKAWIGLFCGCSVPVKVLFQDIDEEIV TDPEKISFEDASYDDYIANIEKGVGIYYEAIVSAKMLFDWSILNEILGDHQ LLSDAMIAEYNKHHDDLKRLQKIIKGTGSRELYQDIFINDVSGNYVCYV GHAKTMSSADQKQFYTFLKNRLKNVNGISSEDAEWIDTEIKNGTLLPKQ TKRDNSVIPHQLQLREFELILDNMQEMYPFLKENREKLLKIFNFVIPYYV GPLKGVVRKGESTNWMVPKKDGVIHPWNFDEMVDKEASAECFISRMT

Description	Sequence
Wild type	GNCSYLFNEKVLPKNSLLYETFEVLNELNPLKINGEPISVELKQRIYEQLF LTGKKVTKKSLTKYLIKNGYDKDIELSGIDNEFHSNLKSHIDFEDYDNLS DEEVEQIILRITVFEDKQLLKDYLNREFVKLSEDERKQICSLSYKGWGNL SEMLLNGITVTDSNGVEVSVMDMLWNTNLNLMQILSKKYGYKAEIEHY NKEHEKTIYNREDLMDYLNIPPAQRRKVNQLITIVKSLKKTYGVPNKIFF KISREHQDDPKRTSSRKEQLKYLYKSLKSEDEKHLMKELDELNDHELSN DKVYLYFLQKGRCIYSGKKLNLSRLRKSNYQNDIDYIYPLSAVNDRSMN NKVLTGIQENRADKYTYFPVDSEIQKKMKGFWMELVLQGFMTKEKYFR LSRENDFSKSELVSFIEREISDNQQSGRMIASVLQYYFPESKIVFVKEKLIS SFKRDFHLISSYGHNHLQAAKDAYITIVVGNVYHTKFTMDPAIYFKNHK RKDYDLNRLFLENISRDGQIAWESGPYGSIQTVRKEYAQNHIAVTKRVV EVKGGLFKQMPLKKGHGEYPLKTNDPRFGNIAQYGGYTNVTGSYFVLV ESMEKGKKRISLEYVPVYLHERLEDDPGHKLLKEYLVDHRKLNHPKILL AKVRKNSLLKIDGFYYRLNGRSGNALILTNAVELIMDDWQTKTANKISG YMKRRAIDKKARVYQNEFHIQELEQLYDFYLDKLKNGVYKNRKNNQA ELIHNEKEQFMELKTEDQCVLLTEIKKLFVCSPMQADLTLIGGSKHTGMI AMSSNVTKADFAVIAEDPLGLRNKVIYSHKGEK (SEQ ID NO: 27)
KvCas9 *Kandleria vitulina* NCBI Reference Sequence: WP_031589969.1 Wild type	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQ ANTAVERRSSRSTRRRYNKRRERIRLLREIMEDMVLDVDPTFFIRLANVS FLDQEDKKDYLKENYHSNYNLFIDKDFNDKTYYDKYPTIYHLRKHLCES KEKEDPRLIYLALHHIVKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFN EINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKDNKAAYK ELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPL LGDCVEFIDLLHDIYSWVELQNILGSAHTSEPSISAAMIQRYEDHKNDLK LLKDVIRKYLPKKYFEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKK LIEKIDDPDVKTILNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQ SVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKDRQFDWIIKKEGKENERIL PWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVLN EINKLRINDHLIKRDMKDKMLHTLFMDHKSISANAMKKWLVKNQYFSN TDDIKIEGFQKENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFED KKILRRRLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTR TPETVLEVMERTNMNLMQVINDEKLGFKKTIDDANSTSVSGKFSYAEVQ ELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSF VNQMLKLYKDYDFEDETEKEANKHLKGEDAKSKIRSERLKLYYTQMG KCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLD DLVIPSSIRNKMYGFWEKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIV ETRQITKHVAQIIDNHYENTKVVTVRADLSHQFRERYHIYKNRDINDFHH AHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQKNKGKEMKKNND GFILNSMRNIYADKDTGEIVWDPNYIDRIKKCFYYKDCFVTKKLEENNG TFFNVTVLPNDTNSDKDNTLATVPVNKYRSNVNKYGGFSGVNSFIVAIK GKKKKGKKVIEVNKLTGIPLMYKNADEEIKINYLKQAEDLEEVQIGKEIL KNQLIEKDGGLYYIVAPTEIINAKQLILNESQTKLVCEIYKAMKYKNYDN LDSEKIIDLYRLLINKMELYYPEYRKQLVKKFEDRYEQLKVISIEEKCNII

Description	Sequence
	KQILATLHCNSSIGKIMYSDFKISTTIGRLNGRTISLDDISFIAESPTGMYSK KYKL (SEQ ID NO: 28)
EfCas9 *Enterococcus faecalis* NCBI Reference Sequence: WP_016631044.1 Wild type	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFF ARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSE QADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKDQFQQFMVIYNQTFV NGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQF LKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVF LAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIR ENCPDEYDNLFKNEQKDGYAGYIAHAGKVSQLKFYQYVKKIIQDIAGAE YFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQE KIEQLVTFRIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQS ATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKA NFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFN ASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFK GQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGV SKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKK GIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEK AMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLS HYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAY WEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNV AGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQD AYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLL RFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFS KESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIK QEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRL LASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEF QEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFN AMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTGLYETRRKVVD (SEQ ID NO: 29)
Staphylococcus aureus Cas9	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKR GARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKL SEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKY VAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFI DTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSV KYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTL KQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAEL LDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVK RSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTN ERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNY EVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETF KKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATR GLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHA EDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYK

Description	Sequence
	EIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDE KNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYP NSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKC YEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMI DITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQ IIKKG (SEQ ID NO: 30)
Geobacillus thermodenitrificans Cas9	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRL ARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRV EALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQ SILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQ REYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK ATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHD VRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVY GKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLA DKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGY TFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHI ELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIV KFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKV LVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLR LHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVN GRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRR EQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEK LESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQ LDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGE LGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTI DMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIK TAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKY QVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ ID NO: 31)
ScCas9 *S. canis* 1375 AA 159.2 kDa	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLM GALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSF FQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPE KADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEE SPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTP NFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDA ILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAE IFKDDTKNGYAGYVDGGASQEEFYKFIKPILEKMDGAEEL LAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKI EKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQS FIERMTNFDEQLPNEKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGF SNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL

Description	Sequence
	QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIK ELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDS RMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAH DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKAT AKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFAT VRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKY GGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGF LEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQ HLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKV NSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRL RYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD (SEQ ID NO: 32)

[0213] The prime editors utilized in the methods and compositions described herein may include any of the above Cas9 ortholog sequences, or any variants thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

[0214] The napDNAbp used in the PEmax constructs described herein may include any suitable homologs and/or orthologs or naturally occurring enzymes, such as, Cas9. Cas9 homologs and/or orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. The Cas moiety may be configured (*e.g.*, mutagenized, recombinantly engineered, or otherwise obtained from nature) as a nickase, *i.e.*, capable of cleaving only a single strand of the target double-stranded DNA. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain; that is, the Cas9 is a nickase. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the Cas9 orthologs in the above tables.

[0215] The present disclosure also contemplates the inclusion of the following additional napDNAbps in the prime editors provided herein. Any suitable napDNAbp may be used in the prime editors utilized in the methods and compositions described herein. In various embodiments, the napDNAbp may be any Class 2 CRISPR-Cas system, including any type

II, type V, or type VI CRISPR-Cas enzyme. Given the rapid development of CRISPR-Cas as a tool for genome editing, there have been constant developments in the nomenclature used to describe and/or identify CRISPR-Cas enzymes, such as Cas9 and Cas9 orthologs. This application references CRISPR-Cas enzymes with nomenclature that may be old and/or new. The skilled person will be able to identify the specific CRISPR-Cas enzyme being referenced in this Application based on the nomenclature that is used, whether it is old (i.e., "legacy") or new nomenclature. CRISPR-Cas nomenclature is extensively discussed in Makarova *et al.*, "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," *The CRISPR Journal*, Vol. 1. No. 5, 2018, the entire contents of which are incorporated herein by reference. The particular CRISPR-Cas nomenclature used in any given instance in this Application is not limiting in any way and the skilled person will be able to identify which CRISPR-Cas enzyme is being referenced.

[0216] For example, the following type II, type V, and type VI Class 2 CRISPR-Cas enzymes have the following art-recognized old (i.e., legacy) and new names. Each of these enzymes, and/or variants thereof, may be used with the prime editors utilized in the methods and compositions described herein:

Legacy nomenclature	**Current nomenclature***
type II CRISPR-Cas enzymes	
Cas9	same
type V CRISPR-Cas enzymes	
Cpf1	Cas12a
CasX	Cas12e
C2c1	Cas12b1
Cas12b2	same
C2c3	Cas12c
CasY	Cas12d
C2c4	same
C2c8	same
C2c5	same
C2c10	same
C2c9	same
type VI CRISPR-Cas enzymes	
C2c2	Cas13a
Cas13d	same
C2c7	Cas13c
C2c6	Cas13b

* See Makarova *et al.*, *The CRISPR Journal*, Vol. 1, No. 5, 2018

[0217] The below description of various napDNAbps which can be used in connection with the prime editors utilized in the presently disclosed methods and compositions is not meant to be limiting in any way. The prime editors may comprise the canonical SpCas9, or any ortholog Cas9 protein, or any variant Cas9 protein —including any naturally occurring variant, mutant, or otherwise engineered version of Cas9 — that is known or that can be

made or evolved through a directed evolutionary or otherwise mutagenic process. In various embodiments, the Cas9 or Cas9 variants have a nickase activity, i.e., only cleave one strand of the target DNA sequence. In other embodiments, the Cas9 or Cas9 variants have inactive nucleases, i.e., are "dead" Cas9 proteins. Other variant Cas9 proteins that may be used are those having a smaller molecular weight than the canonical SpCas9 (e.g., for easier delivery) or having modified or rearranged primary amino acid structure (e.g., the circular permutant formats).

[0218] The prime editors utilized in the methods and compositions described herein may also comprise Cas9 equivalents, including Cas12a (Cpf1) and Cas12b1 proteins which are the result of convergent evolution. The napDNAbps used herein (e.g., SpCas9, Cas9 variant, or Cas9 equivalents) may also contain various modifications that alter/enhance their PAM specificities. Lastly, the application contemplates any Cas9, Cas9 variant, or Cas9 equivalent which has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.9% sequence identity to a reference Cas9 sequence, such as a reference SpCas9 canonical sequence or a reference Cas9 equivalent (e.g., Cas12a (Cpf1)).

[0219] In some embodiments, the napDNAbp directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the napDNAbp directs 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. In some embodiments, a vector encodes a napDNAbp that is mutated to with respect to a corresponding wild-type enzyme such that the mutated napDNAbp lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A in reference to the canonical SpCas9 sequence, or to equivalent amino acid positions in other Cas9 variants or Cas9 equivalents.

[0220] As used herein, the term "Cas protein" refers to a full-length Cas protein obtained from nature, a recombinant Cas protein having a sequences that differs from a naturally occurring Cas protein, or any fragment of a Cas protein that nevertheless retains all or a significant amount of the requisite basic functions needed for the disclosed methods, i.e., (i) possession of nucleic-acid programmable binding of the Cas protein to a target DNA, and (ii)

ability to nick the target DNA sequence on one strand. The Cas proteins contemplated herein embrace CRISPR Cas 9 proteins, as well as Cas9 equivalents, variants (e.g., Cas9 nickase (nCas9) or nuclease inactive Cas9 (dCas9)) homologs, orthologs, or paralogs, whether naturally occurring or non-naturally occurring (e.g., engineered or recombinant), and may include a Cas9 equivalent from any Class 2 CRISPR system (e.g., type II, V, VI), including Cas12a (Cpf1), Cas12e (CasX), Cas12b1 (C2c1), Cas12b2, Cas12c (C2c3), C2c4, C2c8, C2c5, C2c10, C2c9 Cas13a (C2c2), Cas13d, Cas13c (C2c7), Cas13b (C2c6), and Cas13b. Further Cas-equivalents are described in Makarova *et al.*, "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector," Science 2016; 353(6299) and Makarova *et al.*, "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?," *The CRISPR Journal*, Vol. 1. No. 5, 2018, the contents of which are incorporated herein by reference.

[0221] The terms "Cas9" or "Cas9 nuclease" or "Cas9 moiety" or "Cas9 domain" embrace any naturally occurring Cas9 from any organism, any naturally-occurring Cas9 equivalent or functional fragment thereof, any Cas9 homolog, ortholog, or paralog from any organism, and any mutant or variant of a Cas9, naturally-occurring or engineered. The term Cas9 is not meant to be particularly limiting and may be referred to as a "Cas9 or equivalent." Exemplary Cas9 proteins are further described herein and/or are described in the art and are incorporated herein by reference. The present disclosure is unlimited with regard to the particular Cas9 that is employed in the prime editors utilized in the methods and compositions described herein.

[0222] As noted herein, Cas9 nuclease sequences and structures are well-known to those of skill in the art (see, e.g., "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci.* U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).

[0223] Examples of Cas9 and Cas9 equivalents are provided as follows; however, these specific examples are not meant to be limiting. The prime editors utilized in the methods and compositions of the present disclosure may use any suitable napDNAbp, including any suitable Cas9 or Cas9 equivalent.

A. Wild type canonical SpCas9

[0224] In one embodiment, the prime editor constructs utilized in the methods and compositions described herein may comprise the "canonical SpCas9" nuclease from *S. pyogenes*, which has been widely used as a tool for genome engineering and is categorized as the type II subgroup of enzymes of the Class 2 CRISPR-Cas systems. This Cas9 protein is a large, multi-domain protein containing two distinct nuclease domains. Point mutations can be introduced into Cas9 to abolish one or both nuclease activities, resulting in a nickase Cas9 (nCas9) or dead Cas9 (dCas9), respectively, that still retains its ability to bind DNA in a sgRNA-programmed manner. In principle, when fused to another protein or domain, Cas9, or variant thereof (e.g., nCas9) can target that protein to virtually any DNA sequence simply by co-expression with an appropriate sgRNA. As used herein, the canonical SpCas9 protein refers to the wild type protein from *Streptococcus pyogenes* having the following amino acid sequence:

Description	Sequence	SEQ ID NO:
SpCas9 Reverse translation of SwissProt Accession No. Q99ZW2 *Streptococcus pyogenes*	ATGGATAAAAAATATAGCATTGGCCTGGATATTGGCACCAACAGCGTGGGC TGGGCGGTGATTACCGATGAATATAAAGTGCCGAGCAAAAAATTTAAAGTG CTGGGCAACACCGATCGCCATAGCATTAAAAAAAACCTGATTGGCGCGCTG CTGTTTGATAGCGGCGAAACCGCGGAAGCGACCCGCCTGAAACGCACCGCG CGCCGCCGCTATACCCGCCGCAAAAACCGCATTTGCTATCTGCAGGAAATTT TTAGCAACGAAATGGCGAAAGTGGATGATAGCTTTTTTCATCGCCTGGAAGA AAGCTTTCTGGTGGAAGAAGATAAAAAACATGAACGCCATCCGATTTTTGG CAACATTGTGGATGAAGTGGCGTATCATGAAAAATATCCGACCATTTATCAT CTGCGCAAAAAACTGGTGGATAGCACCGATAAAGCGGATCTGCGCCTGATT TATCTGGCGCTGGCGCATATGATTAAATTTCGCGGCCATTTTCTGATTGAAG GCGATCTGAACCCGGATAACAGCGATGTGGATAAACTGTTTATTCAGCTGGT GCAGACCTATAACCAGCTGTTTGAAGAAAACCCGATTAACGCGAGCGGCGT GGATGCGAAAGCGATTCTGAGCGCGCGCCTGAGCAAAAGCCGCCGCCTGGA AAACCTGATTGCGCAGCTGCCGGGCGAAAAAAAAAACGGCCTGTTTGGCAA CCTGATTGCGCTGAGCCTGGGCCTGACCCCGAACTTTAAAAGCAACTTTGAT CTGGCGGAAGATGCGAAACTGCAGCTGAGCAAAGATACCTATGATGATGAT CTGGATAACCTGCTGGCGCAGATTGGCGATCAGTATGCGGATCTGTTTCTGG CGGCGAAAAACCTGAGCGATGCGATTCTGCTGAGCGATATTCTGCGCGTGA ACACCGAAATTACCAAAGCGCCGCTGAGCGCGAGCATGATTAAACGCTATG ATGAACATCATCAGGATCTGACCCTGCTGAAAGCGCTGGTGCGCCAGCAGC TGCCGGAAAAATATAAAGAAATTTTTTTTGATCAGAGCAAAAACGGCTATG CGGGCTATATTGATGGCGGCGCGAGCCAGGAAGAATTTTATAAATTTATTAA ACCGATTCTGGAAAAAATGGATGGCACCGAAGAACTGCTGGTGAAACTGAA CCGCGAAGATCTGCTGCGCAAACAGCGCACCTTTGATAACGGCAGCATTCC GCATCAGATTCATCTGGGCGAACTGCATGCGATTCTGCGCCGCCAGGAAGAT TTTTATCCGTTTCTGAAAGATAACCGCGAAAAAATTGAAAAAATTCTGACCT TTCGCATTCCGTATTATGTGGGCCCGCTGGCGCGCGGCAACAGCCGCTTTGC	8

	GTGGATGACCCGCAAAAGCGAAGAAACCATTACCCCGTGGAACTTTGAAGA AGTGGTGGATAAAGGCGCGAGCGCGCAGAGCTTTATTGAACGCATGACCAA CTTTGATAAAAACCTGCCGAACGAAAAAGTGCTGCCGAAACATAGCCTGCT GTATGAATATTTTACCGTGTATAACGAACTGACCAAAGTGAAATATGTGACC GAAGGCATGCGCAAACCGGCGTTTCTGAGCGGCGAACAGAAAAAAGCGATT GTGGATCTGCTGTTTAAAACCAACCGCAAAGTGACCGTGAAACAGCTGAAA GAAGATTATTTTAAAAAAATTGAATGCTTTGATAGCGTGGAAATTAGCGGCG TGGAAGATCGCTTTAACGCGAGCCTGGGCACCTATCATGATCTGCTGAAAAT TATTAAAGATAAAGATTTTCTGGATAACGAAGAAAACGAAGATATTCTGGA AGATATTGTGCTGACCCTGACCCTGTTTGAAGATCGCGAAATGATTGAAGAA CGCCTGAAAACCTATGCGCATCTGTTTGATGATAAAGTGATGAAACAGCTGA AACGCCGCCGCTATACCGGCTGGGGCCGCCTGAGCCGCAAACTGATTAACG GCATTCGCGATAAACAGAGCGGCAAAACCATTCTGGATTTTCTGAAAAGCG ATGGCTTTGCGAACCGCAACTTTATGCAGCTGATTCATGATGATAGCCTGAC CTTTAAAGAAGATATTCAGAAAGCGCAGGTGAGCGGCCAGGGCGATAGCCT GCATGAACATATTGCGAACCTGGCGGGCAGCCCGGCGATTAAAAAAGGCAT TCTGCAGACCGTGAAAGTGGTGGATGAACTGGTGAAAGTGATGGGCCGCCA TAAACCGGAAAACATTGTGATTGAAATGGCGCGCGAAAACCAGACCACCCA GAAAGGCCAGAAAAACAGCCGCGAACGCATGAAACGCATTGAAGAAGGCA TTAAAGAACTGGGCAGCCAGATTCTGAAAGAACATCCGGTGGAAAACACCC AGCTGCAGAACGAAAAACTGTATCTGTATTATCTGCAGAACGGCCGCGATA TGTATGTGGATCAGGAACTGGATATTAACCGCCTGAGCGATTATGATGTGGA TCATATTGTGCCGCAGAGCTTTCTGAAAGATGATAGCATTGATAACAAAGTG CTGACCCGCAGCGATAAAAACCGCGGCAAAAGCGATAACGTGCCGAGCGAA GAAGTGGTGAAAAAAATGAAAAACTATTGGCGCCAGCTGCTGAACGCGAAA CTGATTACCCAGCGCAAATTTGATAACCTGACCAAAGCGGAACGCGGCGGC CTGAGCGAACTGGATAAAGCGGGCTTTATTAAACGCCAGCTGGTGGAAACC CGCCAGATTACCAAACATGTGGCGCAGATTCTGGATAGCCGCATGAACACC AAATATGATGAAAACGATAAACTGATTCGCGAAGTGAAAGTGATTACCCTG AAAAGCAAACTGGTGAGCGATTTTCGCAAAGATTTTCAGTTTTATAAAGTGC GCGAAATTAACAACTATCATCATGCGCATGATGCGTATCTGAACGCGGTGGT GGGCACCGCGCTGATTAAAAAATATCCGAAACTGGAAAGCGAATTTGTGTA TGGCGATTATAAAGTGTATGATGTGCGCAAAATGATTGCGAAAAGCGAACA GGAAATTGGCAAAGCGACCGCGAAATATTTTTTTTATAGCAACATTATGAAC TTTTTTAAAACCGAAATTACCCTGGCGAACGGCGAAATTCGCAAACGCCCGC TGATTGAAACCAACGGCGAAACCGGCGAAATTGTGTGGGATAAAGGCCGCG ATTTTGCGACCGTGCGCAAAGTGCTGAGCATGCCGCAGGTGAACATTGTGA AAAAAACCGAAGTGCAGACCGGCGGCTTTAGCAAAGAAAGCATTCTGCCGA AACGCAACAGCGATAAACTGATTGCGCGCAAAAAAGATTGGGATCCGAAAA AATATGGCGGCTTTGATAGCCCGACCGTGGCGTATAGCGTGCTGGTGGTGGC GAAAGTGGAAAAAGGCAAAAGCAAAAAACTGAAAAGCGTGAAAGAACTGC TGGGCATTACCATTATGGAACGCAGCAGCTTTGAAAAAAACCCGATTGATTT TCTGGAAGCGAAAGGCTATAAAGAAGTGAAAAAAGATCTGATTATTAAACT GCCGAAATATAGCCTGTTTGAACTGGAAAACGGCCGCAAACGCATGCTGGC GAGCGCGGGCGAACTGCAGAAAGGCAACGAACTGGCGCTGCCGAGCAAAT ATGTGAACTTTCTGTATCTGGCGAGCCATTATGAAAAACTGAAAGGCAGCCC GGAAGATAACGAACAGAAACAGCTGTTTGTGGAACAGCATAAACATTATCT GGATGAAATTATTGAACAGATTAGCGAATTTAGCAAACGCGTGATTCTGGC GGATGCGAACCTGGATAAAGTGCTGAGCGCGTATAACAAACATCGCGATAA ACCGATTCGCGAACAGGCGGAAAACATTATTCATCTGTTTACCCTGACCAAC CTGGGCGCGCCGGCGGCGTTTAAATATTTTGATACCACCATTGATCGCAAAC GCTATACCAGCACCAAAGAAGTGCTGGATGCGACCCTGATTCATCAGAGCA TTACCGGCCTGTATGAAACCCGCATTGATCTGAGCCAGCTGGGCGGCGAT	

[0225] The prime editors utilized in the methods and compositions described herein may include canonical SpCas9, or any variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with a wild type Cas9 sequence provided above. These variants may include SpCas9 variants containing one or more mutations,

including any known mutation reported with the SwissProt Accession No. Q99ZW2 (SEQ ID NO: 9) entry, which include:

SpCas9 mutation (relative to the amino acid sequence of the canonical SpCas9 sequence, SEQ ID NO: 9)	Function/Characteristic (as reported) (see UniProtKB – Q99ZW2 (CAS9_STRPT1) entry – incorporated herein by reference)
D10A	Nickase mutant which cleaves the protospacer strand (but no cleavage of non-protospacer strand)
S15A	Decreased DNA cleavage activity
R66A	Decreased DNA cleavage activity
R70A	No DNA cleavage
R74A	Decreased DNA cleavage
R78A	Decreased DNA cleavage
97-150 deletion	No nuclease activity
R165A	Decreased DNA cleavage
175-307 deletion	About 50% decreased DNA cleavage
312-409 deletion	No nuclease activity
E762A	Nickase
H840A	Nickase mutant which cleaves the non-protospacer strand but does not cleave the protospacer strand
N854A	Nickase
N863A	Nickase
H982A	Decreased DNA cleavage
D986A	Nickase
1099-1368 deletion	No nuclease activity
R1333A	Reduced DNA binding

B. Wild type Cas9 orthologs

[0226] In other embodiments, the Cas9 protein can be a wild type Cas9 ortholog from another bacterial species different from the canonical Cas9 from *S. pyogenes*. For example, the following Cas9 orthologs can be used in connection with the prime editor constructs utilized in the methods and compositions described in this specification. In addition, any variant Cas9 orthologs having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any of the below orthologs may also be used with the prime editors.

Description	Sequence
LfCas9 *Lactobacillus fermentum* wild type *GenBank: SNX31424.1* 1	MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKR RKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPTKNQAFIGKLLFPDLLKKNERGYP TLIKMRDELPVEQRAHYPVMNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNA SVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKA VAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSED EIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLA TQPASARKEFDQVYNKYIGQAPKERGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQR TSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYY VGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLL NEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKASDVASLV MAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQTDLENIIEWRSVFED GEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQN FKEIVDQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDI VKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLS DRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDNSLDNRVLTSRKENNKKS DQVPAKLYAAKMKPYWNQLLKQGLITQRKFENLTKDVDQNIKYRSLGFVKRQLVETRQ VIKLTANILGSMYQEAGTEIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQ

Description	Sequence
	YLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITT RDEVAKSFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEIKKNRLVDLYG AYTNGTSAFMTIIKFTGNKPKYKVIGIPTTSAASLKRAGKPGSESYNQELHRIIKSNPKVKK GFEIVVPHVSYGQLIVDGDCKFTLASPTVQHPATQLVLSKKSLETISSGYKILKDKPAIANE RLIRVFDEVVGQMNRYFTIFDQRSNRQKVADARDKFLSLPTESKYEGAKKVQVGKTEVIT NLLMGLHANATQGDLKVLGLATFGFFQSTTGLSLSEDTMIVYQSPTGLFERRICLKDI (SEQ ID NO: 20)
SaCas9 *Staphylococcus aureus* wild type GenBank: AYD60528.1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDS IDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 259)
SaCas9 *Staphylococcus aureus*	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVH NVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKE AKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCT YFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQI AKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSE DIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVP KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQK MINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPF NYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAK GKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKV KSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQM FEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKD DKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPY RFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLI KINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDIL GNLYEVKSKKHPQIIKK (SEQ ID NO: 21)
StCas9 *Streptococcus thermophilus* UniProtKB/Swiss-Prot: G3ECR1.2 Wild type	MLFNKCIIISINLDFSNKEKCMTKPYSIGLDIGTNSVGWAVITDNYKVPSKKMKVLGNTSK KYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRL DDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHM IKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKD RILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDD YSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKT YNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFLRKQRTF DNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIR KRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVR FIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSL

Description	Sequence
	STYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRH YTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDE DKGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSN SQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDI DRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDFPSLEVVKKRKTFWYQLLKSKL ISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRAVR TVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGD YPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLA TVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPK KYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIE LIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINEN HRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGS ERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG (SEQ ID NO: 22)
LcCas9 *Lactobacillus crispatus* NCBI Reference Sequence: WP_133478044 .1 Wild type	MKIKNYNLALTPSTSAVGHVEVDDDLNILEPVHHQKAIGVAKFGEGETAEARRLARSARR TTKRRANRINHYFNEIMKPEIDKVDPLMFDRIKQAGLSPLDERKEFRTVIFDRPNIASYYHN QFPTIWHLQKYLMITDEKADIRLIYWALHSLLKHRGHFFNTTPMSQFKPGKLNLKDDMLA LDDYNDLEGLSFAVANSPEIEKVIKDRSMHKKEKIAELKKLIVNDVPDKDLAKRNNKIITQ IVNAIMGNSFHLNFIFDMDLDKLTSKAWSFKLDDPELDTKFDAISGSMTDNQIGIFETLQKI YSAISLLDILNGSSNVVDAKNALYDKHKRDLNLYFKFLNTLPDEIAKTLKAGYTLYIGNRK KDLLAARKLLKVNVAKNFSQDDFYKLINKELKSIDKQGLQTRFSEKVGELVAQNNFLPV QRSSDNVFIPYQLNAITFNKILENQGKYYDFLVKPNPAKKDRKNAPYELSQLMQFTIPYYV GPLVTPEEQVKSGIPKTSRFAWMVRKDNGAITPWNFYDKVDIEATADKFIKRSIAKDSYLL SELVLPKHSLLYEKYEVFNELSNVSLDGKKLSGGVKQILFNEVFKKTNKVNTSRILKALA KHNIPGSKITGLSNPEEFTSSLQTYNAWKKYFPNQIDNFAYQQDLEKMIEWSTVFEDHKIL AKKLDEIEWLDDDQKKFVANTRLRGWGRLSKRLLTGLKDNYGKSIMQRLETTKANFQQI VYKPEFREQIDKISQAAAKNQSLEDILANSYTSPSNRKAIRKTMSVVDEYIKLNHGKEPDK IFLMFQRSEQEKGKQTEARSKQLNRILSQLKADKSANKLFSKQLADEFSNAIKKSKYKLN DKQYFYFQQLGRDALTGEVIDYDELYKYTVLHIIPRSKLTDDSQNNKVLTKYKIVDGSVA LKFGNSYSDALGMPIKAFWTELNRLKLIPKGKLLNLTTDFSTLNKYQRDGYIARQLVETQ QIVKLLATIMQSRFKHTKIIEVRNSQVANIRYQFDYFRIKNLNEYYRGFDAYLAAVVGTYL YKVYPKARRLFVYGQYLKPKKTNQENQDMHLDSEKKSQGFNFLWNLLYGKQDQIFVNG TDVIAFNRKDLITKMNTVYNYKSQKISLAIDYHNGAMFKATLFPRNDRDTAKTRKLIPKK KDYDTDIYGGYTSNVDGYMLLAEIIKRDGNKQYGFYGVPSRLVSELDTLKKTRYTEYEEK LKEIIKPELGVDLKKIKKIKILKNKVPFNQVIIDKGSKFFITSTSYRWNYRQLILSAESQQTL MDLVVDPDFSNHKARKDARKNADERLIKVYEEILYQVKNYMPMFVELHRCYEKLVDAQ KTFKSLKISDKAMVLNQILILLHSNATSPVLEKLGYHTRFTLGKKHNLISENAVLVTQSITG LKENHVSIKQML (SEQ ID NO: 23)
PdCas9 *Pedicoccus damnosus* NCBI Reference Sequence: WP_062913273 .1 Wild type	MTNEKYSIGLDIGTSSIGFAVVNDNNRVIRVKGKNAIGVRLFDEGKAAADRRSFRTTRRSF RTTRRRLSRRRWRLKLLREIFDAYITPVDEAFFIRLKESNLSPKDSKKQYSGDILFNDRSDK DFYEKYPTIYHLRNALMTEHRKFDVREIYLAIHHIMKFRGHFLNATPANNFKVGRLNLEE KFEELNDIYQRVFPDESIEFRTDNLEQIKEVLLDNKRSRADRQRTLVSDIYQSSEDKDIEKR NKAVATEILKASLGNKAKLNVITNVEVDKEAAKEWSITFDSESIDDDLAKIEGQMTDDGH EIIEVLRSLYSGITLSAIVPENHTLSQSMVAKYDLHKDHLKLFKKLINGMTDTKKAKNLRA AYDGYIDGVKGKVLPQEDFYKQVQVNLDDSAEANEIQTYIDQDIFMPKQRTKANGSIPHQ LQQQELDQIIENQKAYYPWLAELNPNPDKKRQQLAKYKLDELVTFRVPYYVGPMITAKD QKNQSGAEFAWMIRKEPGNITPWNFDQKVDRMATANQFIKRMTTTDTYLLGEDVLPAQS LLYQKFEVLNELNKIRIDHKPISIEQKQQIFNDLFKQFKNVTIKHLQDYLVSQGQYSKRPLI EGLADEKRFNSSLSTYSDLCGIFGAKLVEENDRQEDLEKIIEWSTIFEDKKIYRAKLNDLT WLTDDQKEKLATKRYQGWGRLSRKLLVGLKNSEHRNIMDILWITNENFMQIQAEPDFAK LVTDANKGMLEKTDSQDVINDLYTSPQNKKAIRQILLVVHDIQNAMHGQAPAKIHVEFAR GEERNPRRSVQRQRQVEAAYEKVSNELVSAKVRQEFKEAINNKRDFKDRLFLYFMQGGI DIYTGKQLNIDQLSSYQIDHILPQAFVKDDSLTNRVLTNENQVKADSVPIDIFGKKMLSVW GRMKDQGLISKGKYRNLTMNPENISAHTENGFINRQLVETRQVIKLAVNILADEYGDSTQI ISVKADLSHQMREDFELLKNRDVNDYHHAFDAYLAAFIGNYLLKRYPKLESYFVYGDFK KFTQKETKMRRFNFIYDLKHCDQVVNKETGEILWTKDEDIKYIRHLFAYKKILVSHEVRE KRGALYNQTIYKAKDDKGSGQESKKLIRIKDDKETKIYGGYSGKSLAYMTIVQITKKNKV SYRVIGIPTLALARLNKLENDSTENNGELYKIIKPQFTHYKVDKKNGEIIETTDDFKIVVSK VRFQQLIDDAGQFFMLASDTYKNNAQQLVISNNALKAINNTNITDCPRDDLERLDNLRLD

Description	Sequence
	SAFDEIVKKMDKYFSAYDANNFREKIRNSNLIFYQLPVEDQWENNKITELGKRTVLTRILQ GLHANATTTDMSIFKIKTPFGQLRQRSGISLSENAQLIYQSPTGLFERRVQLNKIK (SEQ ID NO: 24)
FnCas9 *Fusobaterium nucleatum* NCBI Reference Sequence: WP_060798984 .1	MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFEEAKTAAERRVQ RNSRRRLKRRKWRLNLLEEIFSNEILKIDSNFFRRLKESSLWLEDKSSKEKFTLFNDDNYK DYDFYKQYPTIFHLRNELIKNPEKKDIRLVYLAIHSIFKSRGHFLFEGQNLKEIKNFETLYN NLIAFLEDNGINKIIDKNNIEKLEKIVCDSKKGLKDKEKEFKEIFNSDKQLVAIFKLSVGSSV SLNDLFDTDEYKKGEVEKEKISFREQIYEDDKPIYYSILGEKIELLDIAKTFYDFMVLNNILA DSQYISEAKVKLYEEHKKDLKNLKYIIRKYNKGNYDKLFKDKNENNYSAYIGLNKEKSK KEVIEKSRLKIDDLIKNIKGYLPKVEEIEEKDKAIFNKILNKIELKTILPKQRISDNGTLPYQI HEAELEKILENQSKYYDFLNYEENGIITKDKLLMTFKFRIPYYVGPLNSYHKDKGGNSWIV RKEEGKILPWNFEQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDTFLYSEYVILNELNKV QVNDEFLNEENKRKIIDELFKENKKVSEKKFKEYLLVKQIVDGTIELKGVKDSFNSNYISYI RFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKIFEKKIKNEYGDILTKDEIKKINTFKFNN WGRLSEKLLTGIEFINLETGECYSSVMDALRRTNYNLMELLSSKFTLQESINNENKEMNEA SYRDLIEESYVSPSLKRAIFQTLKIYEEIRKITGRVPKKVFIEMARGGDESMKNKKIPARQE QLKKLYDSCGNDIANFSIDIKEMKNSLISYDNNSLRQKKLYLYYLQFGKCMYTGREIDLD RLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPVKKEIQEKMKSFWRF LKEKNFISDEKYKRLTGKDDFELRGFMARQLVNVRQTTKEVGKILQQIEPEIKIVYSKAEI ASSFREMFDFIKVRELNDTHHAKDAYLNIVAGNVYNTKFTEKPYRYLQEIKENYDVKKIY NYDIKNAWDKENSLEIVKKNMEKNTVNITRFIKEKKGQLFDLNPIKKGETSNEIISIKPKVY NGKDDKLNEKYGYYKSLNPAYFLYVEHKEKNKRIKSFERVNLVDVNNIKDEKSLVKYLI ENKKLVEPRVIKKVYKRQVILINDYPYSIVTLDSNKLMDFENLKPLFLENKYEKILKNVIKF LEDNQGKSEENYKFIYLKKKDRYEKNETLESVKDRYNLEFNEMYDKFLEKLDSKDYKNY MNNKKYQELLDVKEKFIKLNLFDKAFTLKSFLDLFNRKTMADFSKVGLTKYLGKIQKISS NVLSKNELYLLEESVTGLFVKKIKL (SEQ ID NO: 25)
EcCas9 *Enterococcus cecorum* NCBI Reference Sequence: WP_047338501 .1 Wild type	RRKQRIQILQELLGEEVLKTDPGFFHRMKESRYVVEDKRTLDGKQVELPYALFVDKDYTD KEYYKQFPTINHLIVYLMTTSDTPDIRLVYLALHYYMKNRGNFLHSGDINNVKDINDILEQ LDNVLETFLDGWNLKLKSYVEDIKNIYNRDLGRGERKKAFVNTLGAKTKAEKAFCSLISG GSTNLAELFDDSSLKEIETPKIEFASSSLEDKIDGIQEALEDRFAVIEAAKRLYDWKTLTDIL GDSSSLAEARVNSYQMHHEQLLELKSLVKEYLDRKVFQEVFVSLNVANNYPAYIGHTKI NGKKKELEVKRTKRNDFYSYVKKQVIEPIKKKVSDEAVLTKLSEIESLIEVDKYLPLQVNS DNGVIPYQVKLNELTRIFDNLENRIPVLRENRDKIIKTFKFRIPYYVGSLNGVVKNGKCTN WMVRKEEGKIYPWNFEDKVDLEASAEQFIRRMTNKCTYLVNEDVLPKYSLLYSKYLVLS ELNNLRIDGRPLDVKIKQDIYENVFKKNRKVTLKKIKKYLLKEGIITDDDELSGLADDVKS SLTAYRDFKEKLGHLDLSEAQMENIILNITLFGDDKKLLKKRLAALYPFIDDKSLNRIATLN YRDWGRLSERFLSGITSVDQETGELRTIIQCMYETQANLMQLLAEPYHFVEAIEKENPKVD LESISYRIVNDLYVSPAVKRQIWQTLLVIKDIKQVMKHDPERIFIEMAREKQESKKTKSRK QVLSEVYKKAKEYEHLFEKLNSLTEEQLRSKKIYLYFTQLGKCMYSGEPIDFENLVSANS NYDIDHIYPQSKTIDDSFNNIVLVKKSLNAYKSNHYPIDKNIRDNEKVKTLWNTLVSKGLI TKEKYERLIRSTPFSDEELAGFIARQLVETRQSTKAVAEILSNWFPESEIVYSKAKNVSNFR QDFEILKVRELNDCHHAHDAYLNIVVGNAYHTKFTNSPYRFIKNKANQEYNLRKLLQKV NKIESNGVVAWVGQSENNPGTIATVKKVIRRNTVLISRMVKEVDGQLFDLTLMKKGKGQ VPIKSSDERLTDISKYGGYNKATGAYFTFVKSKKRGKVVRSFEYVPLHLSKQFENNNELL KEYIEKDRGLTDVEILIPKVLINSLFRYNGSLVRITGRGDTRLLLVHEQPLYVSNSFVQQLK SVSSYKLKKSENDNAKLTKTATEKLSNIDELYDGLLRKLDLPIYSYWFSSIKEYLVESRTK YIKLSIEEKALVIFEILHLFQSDAQVPNLKILGLSTKPSRIRIQKNLKDTDKMSIIHQSPSGIFE HEIELTSL (SEQ ID NO: 26)
AhCas9 *Anaerostipes hadrus* NCBI Reference Sequence: WP_044924278 .1 Wild type	MQNGFLGITVSSEQVGWAVTNPKYELERASRKDLWGVRLFDKAETAEDRRMFRTNRRL NQRKKNRIHYLRDIFHEEVNQKDPNFFQQLDESNFCEDDRTVEFNFDTNLYKNQFPTVYH LRKYLMETKDKPDIRLVYLAFSKFMKNRGHFLYKGNLGEVMDFENSMKGFCESLEKFNI DFPTLSDEQVKEVRDILCDHKIAKTVKKKNIITITKVKSKTAKAWIGLFCGCSVPVKVLFQ DIDEEIVTDPEKISFEDASYDDYIANIEKGVGIYYEAIVSAKMLFDWSILNEILGDHQLLSDA MIAEYNKHHDDLKRLQKIIKGTGSRELYQDIFINDVSGNYVCYVGHAKTMSSADQKQFY TFLKNRLKNVNGISSEDAEWIDTEIKNGTLLPKQTKRDNSVIPHQLQLREFELILDNMQEM YPFLKENREKLLKIFNFVIPYYVGPLKGVVRKGESTNWMVPKKDGVIHPWNFDEMVDKE ASAECFISRMTGNCSYLFNEKVLPKNSLLYETFEVLNELNPLKINGEPISVELKQRIYEQLF LTGKKVTKKSLTKYLIKNGYDKDIELSGIDNEFHSNLKSHIDFEDYDNLSDEEVEQIILRITV FEDKQLLKDYLNREFVKLSEDERKQICSLSYKGWGNLSEMLLNGITVTDSNGVEVSVMD

Description	Sequence
	MLWNTNLNLMQILSKKYGYKAEIEHYNKEHEKTIYNREDLMDYLNIPPAQRRKVNQLITI VKSLKKTYGVPNKIFFKISREHQDDPKRTSSRKEQLKYLYKSLKSEDEKHLMKELDELND HELSNDKVYLYFLQKGRCIYSGKKLNLSRLRKSNYQNDIDYIYPLSAVNDRSMNNKVLTG IQENRADKYTYFPVDSEIQKKMKGFWMELVLQGFMTKEKYFRLSRENDFSKSELVSFIER EISDNQQSGRMIASVLQYYFPESKIVFVKEKLISSFKRDFHLISSYGHNHLQAAKDAYITIV VGNVYHTKFTMDPAIYFKNHKRKDYDLNRLFLENISRDGQIAWESGPYGSIQTVRKEYAQ NHIAVTKRVVEVKGGLFKQMPLKKGHGEYPLKTNDPRFGNIAQYGGYTNVTGSYFVLVE SMEKGKKRISLEYVPVYLHERLEDDPGHKLLKEYLVDHRKLNHPKILLAKVRKNSLLKID GFYYRLNGRSGNALILTNAVELIMDDWQTKTANKISGYMKRRAIDKKARVYQNEFHIQE LEQLYDFYLDKLKNGVYKNRKNNQAELIHNEKEQFMELKTEDQCVLLTEIKKLFVCSPM QADLTLIGGSKHTGMIAMSSNVTKADFAVIAEDPLGLRNKVIYSHKGEK (SEQ ID NO: 27)
KvCas9 *Kandleria vitulina* NCBI Reference Sequence: WP_031589969.1 Wild type	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQANTAVERRSSR STRRRYNKRRERIRLLREIMEDMVLDVDPTFFIRLANVSFLDQEDKKDYLKENYHSNYNL FIDKDFNDKTYYDKYPTIYHLRKHLCESKEKEDPRLIYLALHHIVKYRGNFLYEGQKFSM DVSNIEDKMIDVLRQFNEINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKD NKAAYKELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPLLGDC VEFIDLLHDIYSWVELQNILGSAHTSEPSISAAMIQRYEDHKNDLKLLKDVIRKYLPKKYF EVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKKLIEKIDDPDVKTILNKIELESFMLKQNS RTNGAVPYQMQLDELNKILENQSVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKDRQFDW IIKKEGKENERILPWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVL NEINKLRINDHLIKRDMKDKMLHTLFMDHKSISANAMKKWLVKNQYFSNTDDIKIEGFQ KENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFEDKKILRRRLKKEYDLDEEKIKK ILKLKYSGWSRLSKKLLSGIKTKYKDSTRTPETVLEVMERTNMNLMQVINDEKLGFKKTI DDANSTSVSGKFSYAEVQELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDE KERKDSFVNQMLKLYKDYDFEDETEKEANKHLKGEDAKSKIRSERLKLYYTQMGKCMY TGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLDDLVIPSSIRNKMYGFW EKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIVETRQITKHVAQIIDNHYENTKVVTVRA DLSHQFRERYHIYKNRDINDFHHAHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQK NKGKEMKKNNDGFILNSMRNIYADKDTGEIVWDPNYIDRIKKCFYYKDCFVTKKLEENN GTFFNVTVLPNDTNSDKDNTLATVPVNKYRSNVNKYGGFSGVNSFIVAIKGKKKKGKKV IEVNKLTGIPLMYKNADEEIKINYLKQAEDLEEVQIGKEILKNQLIEKDGGLYYIVAPTEIIN AKQLILNESQTKLVCEIYKAMKYKNYDNLDSEKIIDLYRLLINKMELYYPEYRKQLVKKF EDRYEQLKVISIEEKCNIIKQILATLHCNSSIGKIMYSDFKISTTIGRLNGRTISLDDISFIAESP TGMYSKKYKL (SEQ ID NO: 28)
EfCas9 *Enterococcus faecalis* NCBI Reference Sequence: WP_016631044.1 Wild type	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPE DKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLI EGKLSTENTSVKDQFQQFMVIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKV LQQFPQEKANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDE YSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDE YDNLFKNEQKDGYAGYIAHAGKVSQLKFYQYVKKIIQDIAGAEYFLEKIAQENFLRKQRT FDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWL KRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTK ISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNA SFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLE RKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQLSFKNAI QKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTT STGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSL HRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAA GLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNAKSKEKKVQI ITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPK FQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKV EVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEI LGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRLLASAKEAQKGNQM VLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIV KLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTG LYETRRKVVD (SEQ ID NO: 29)
Staphylococcus aureus Cas9	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH RIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVN EVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQ

Description	Sequence
	LLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFP EELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAK EILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQ EELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKK VDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMI NEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFN YEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVK SINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMF EEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDD KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYR FDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIK INGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILG NLYEVKSKKHPQIIKKG (SEQ ID NO: 30)
Geobacillus thermodenitrificans Cas9	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRR KHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKR RGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNT VARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKE KRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLN LPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFG YALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALR NILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAII KKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVK FKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGN RTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYIS RFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVAC TTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGN YDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTG HFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIP LNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTE DYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLR SIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ ID NO: 31)
ScCas9 *S. canis* 1375 AA 159.2 kDa	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAE ATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGN LADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVA KLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALA LGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDIL RSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGI GIKHRKRTTKLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHL KELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNF EEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKP EFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLL KIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTG WGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDS LHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERK KRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI VPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKL VSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFA TVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAY SILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYS LFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFK EIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLD VKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD (SEQ ID NO: 32)

[0227] The prime editors utilized in the methods and compositions described herein may include any of the above Cas9 ortholog sequences, or any variants thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

[0228] The napDNAbp may include any suitable homologs and/or orthologs or naturally occurring enzymes, such as, Cas9. Cas9 homologs and/or orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Preferably, the Cas moiety is configured (e.g., mutagenized, recombinantly engineered, or otherwise obtained from nature) as a nickase, i.e., capable of cleaving only a single strand of the target double-stranded DNA. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain; that is, the Cas9 is a nickase. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 3. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the Cas9 orthologs in the above tables.

C. Dead Cas9 variant

[0229] In certain embodiments, the prime editors utilized in the methods and compositions described herein may include a dead Cas9, e.g., dead SpCas9, which has no nuclease activity due to one or more mutations that inactive both nuclease domains of Cas9, namely the RuvC domain (which cleaves the non-protospacer DNA strand) and HNH domain (which cleaves the protospacer DNA strand). The nuclease inactivation may be due to one or mutations that result in one or more substitutions and/or deletions in the amino acid sequence of the encoded protein, or any variants thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

[0230] As used herein, the term "dCas9" refers to a nuclease-inactive Cas9 or nuclease-dead Cas9, or a functional fragment thereof, and embraces any naturally occurring dCas9 from any organism, any naturally-occurring dCas9 equivalent or functional fragment thereof, any engineered dCas9 variant or functional fragment thereof, any dCas9 homolog, ortholog, or

paralog from any organism, and any mutant or variant of a dCas9, naturally-occurring or engineered. The term dCas9 is not meant to be particularly limiting and may be referred to as a "dCas9 or equivalent." Exemplary dCas9 proteins and method for making dCas9 proteins are further described herein and/or are described in the art and are incorporated herein by reference.

[0231] In other 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 other embodiments, Cas9 variants having mutations other than D10A and H840A are provided which may result in the full or partial inactivation of the endogenous Cas9 nuclease activity (e.g., nCas9 or dCas9, respectively). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain) with reference to a wild type sequence such as Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1). In some embodiments, variants or homologues of Cas9 (e.g., variants of Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1 (SEQ ID NO: 16))) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to NCBI Reference Sequence: NC_017053.1. In some embodiments, variants of dCas9 (e.g., variants of NCBI Reference Sequence: NC_017053.1 (SEQ ID NO: 16)) are provided having amino acid sequences which are shorter, or longer than NC_017053.1 (SEQ ID NO: 16) by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

[0232] In one embodiment, the dead Cas9 may be based on the canonical SpCas9 sequence of Q99ZW2 and may have the following sequence, which comprises a D10X and an H810X, wherein X may be any amino acid, substitutions (underlined and bolded), or a variant be variant of SEQ ID NO: 260 having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

[0233] In one embodiment, the dead Cas9 may be based on the canonical SpCas9 sequence of Q99ZW2 and may have the following sequence, which comprises a D10A and an H810A substitutions (underlined and bolded), or be a variant of SEQ ID NO: 261 having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

Description	Sequence	SEQ ID NO:
dead Cas9 or dCas9 *Streptococcus pyogenes* Q99ZW2 Cas9 with D10**X** and H810**X** Where "X" is any amino acid	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	260
dead Cas9 or dCas9 *Streptococcus pyogenes* Q99ZW2 Cas9 with D10**A** and H810**A**	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	261

D. Cas9 nickase variant

[0234] In one embodiment, the prime editors utilized in the methods and compositions described herein comprise a Cas9 nickase. The term "Cas9 nickase" or "nCas9" refers to a variant of Cas9 which is capable of introducing a single-strand break in a double strand DNA molecule target. In some embodiments, the Cas9 nickase comprises only a single functioning nuclease domain. The wild type Cas9 (e.g., the canonical SpCas9) comprises two separate nuclease domains, namely, the RuvC domain (which cleaves the non-protospacer DNA strand) and HNH domain (which cleaves the protospacer DNA strand). In one embodiment, the Cas9 nickase comprises a mutation in the RuvC domain which inactivates the RuvC nuclease activity. For example, mutations in aspartate (D) 10, histidine (H) 983, aspartate (D) 986, or glutamate (E) 762, have been reported as loss-of-function mutations of the RuvC nuclease domain and the creation of a functional Cas9 nickase (e.g., Nishimasu *et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell* 156(5), 935–949, which is incorporated herein by reference). Thus, nickase mutations in the RuvC domain could include D10X, H983X, D986X, or E762X, wherein X is any amino acid other than the wild type amino acid. In certain embodiments, the nickase could be D10A, of H983A, D986A, or E762A, or a combination thereof.

[0235] In various embodiments, the Cas9 nickase can have a mutation in the RuvC nuclease domain and have one of the following amino acid sequences, or a variant thereof having an amino acid sequence that has at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

Description	Sequence	SEQ ID NO:
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with D10**X**, wherein X is any alternate amino acid	MDKKYSIGL**X**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR	262

Description	Sequence	SEQ ID NO:
	KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with E762X, wherein X is any alternate amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVI**X**MARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	263
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with H983X, wherein X is any alternate amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYH**X**AHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	264

Description	Sequence	SEQ ID NO:
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with D986X, wherein X is any alternate amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAH**X**AYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	265
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with D10**A**	MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	266
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with E762A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS	267

Description	Sequence	SEQ ID NO:
	MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVI**A**MARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with H983A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYH**A**AHDAYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	268
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with D986A	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY	269

Description	Sequence	SEQ ID NO:
	AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILK EHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAH**A**AYLNAVVGTALIKKYPKLE SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	

[0236] In another embodiment, the Cas9 nickase comprises a mutation in the HNH domain which inactivates the HNH nuclease activity. For example, mutations in histidine (H) 840 or asparagine (R) 863 have been reported as loss-of-function mutations of the HNH nuclease domain and the creation of a functional Cas9 nickase (e.g., Nishimasu *et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell* 156(5), 935–949, which is incorporated herein by reference). Thus, nickase mutations in the HNH domain could include H840X and R863X, wherein X is any amino acid other than the wild type amino acid. In certain embodiments, the nickase could be H840A or R863A or a combination thereof.

[0237] In various embodiments, the Cas9 nickase can have a mutation in the HNH nuclease domain and have one of the following amino acid sequences, or a variant thereof having an amino acid sequence that has at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

Description	Sequence	SEQ ID NO:
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with H840**X**, wherein X is any alternate amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS	270

Description	Sequence	SEQ ID NO:
	EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with H840**A**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKN**R**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	12
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with R863X, wherein X is any alternate amino acid	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**X**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN	271

Description	Sequence	SEQ ID NO:
	KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	
Cas9 nickase *Streptococcus pyogenes* Q99ZW2 Cas9 with R863**A**	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILE KMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGE QKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**A**GKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS ITGLYETRIDLSQLGGD	272

[0238] In some embodiments, the N-terminal methionine is removed from a Cas9 nickase, or from any Cas9 variant, ortholog, or equivalent disclosed or contemplated herein. For example, methionine-minus Cas9 nickases include the following sequences, or a variant thereof having an amino acid sequence that has at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto.

Description	Sequence
Cas9 nickase (Met minus) *Streptococcus pyogenes* Q99ZW2 Cas9 with H840**X**, wherein X is any alternate amino acid	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVD**X**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT

	GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 273)
Cas9 nickase (Met minus) *Streptococcus pyogenes* Q99ZW2 Cas9 with H840**A**	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKN**R**GKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 10)
Cas9 nickase (Met minus) *Streptococcus pyogenes* Q99ZW2 Cas9 with R863X, wherein X is any alternate amino acid	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**X**GKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO: 274)
Cas9 nickase (Met minus) *Streptococcus pyogenes* Q99ZW2 Cas9 with R863**A**	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI

	PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN**A**GKSDNVPSEE VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 275)

E. Other Cas9 variants

[0239] Besides dead Cas9 and Cas9 nickase variants, the Cas9 proteins used herein may also include other "Cas9 variants" having at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any reference Cas9 protein, including any wild type Cas9, or mutant Cas9 (e.g., a dead Cas9 or Cas9 nickase), or fragment Cas9, or circular permutant Cas9, or other variant of Cas9 disclosed herein or known in the art. In some embodiments, a Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to a reference Cas9. In some embodiments, the Cas9 variant comprises a fragment of a reference Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9 (*e.g.*, SEQ ID NO: 9).

[0240] In some embodiments, the disclosure also may utilize Cas9 fragments that retain their functionality and that are fragments of any herein disclosed Cas9 protein. In some

embodiments, the Cas9 fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length.

[0241] In various embodiments, the prime editors utilized in the methods and compositions disclosed herein may comprise one of the Cas9 variants described as follows, or a Cas9 variant thereof having at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any reference Cas9 variants.

F. Small-sized Cas9 variants

[0242] In some embodiments, the prime editors utilized in the methods and compositions contemplated herein can include a Cas9 protein that is of smaller molecular weight than the canonical SpCas9 sequence. In some embodiments, the smaller-sized Cas9 variants may facilitate delivery to cells, e.g., by an expression vector, nanoparticle, or other means of delivery. In certain embodiments, the smaller-sized Cas9 variants can include enzymes categorized as type II enzymes of the Class 2 CRISPR-Cas systems. In some embodiments, the smaller-sized Cas9 variants can include enzymes categorized as type V enzymes of the Class 2 CRISPR-Cas systems. In other embodiments, the smaller-sized Cas9 variants can include enzymes categorized as type VI enzymes of the Class 2 CRISPR-Cas systems.

[0243] The canonical SpCas9 protein is 1368 amino acids in length and has a predicted molecular weight of 158 kilodaltons. The term "small-sized Cas9 variant", as used herein, refers to any Cas9 variant—naturally occurring, engineered, or otherwise—that is less than at least 1300 amino acids, or at least less than 1290 amino acids, or than less than 1280 amino acids, or less than 1270 amino acid, or less than 1260 amino acid, or less than 1250 amino acids, or less than 1240 amino acids, or less than 1230 amino acids, or less than 1220 amino acids, or less than 1210 amino acids, or less than 1200 amino acids, or less than 1190 amino acids, or less than 1180 amino acids, or less than 1170 amino acids, or less than 1160 amino acids, or less than 1150 amino acids, or less than 1140 amino acids, or less than 1130 amino acids, or less than 1120 amino acids, or less than 1110 amino acids, or less than 1100 amino acids, or less than 1050 amino acids, or less than 1000 amino acids, or less than 950 amino acids, or less than 900 amino acids, or less than 850 amino acids, or less than 800 amino acids, or less than 750 amino acids, or less than 700 amino acids, or less than 650 amino acids, or less than 600 amino acids, or less than 550 amino acids, or less than 500 amino

acids, but at least larger than about 400 amino acids and retaining the required functions of the Cas9 protein. The Cas9 variants can include those categorized as type II, type V, or type VI enzymes of the Class 2 CRISPR-Cas system.

[0244] In various embodiments, the prime editors utilized in the methods and compositions disclosed herein may comprise one of the small-sized Cas9 variants described as follows, or a Cas9 variant thereof having at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any reference small-sized Cas9 protein.

Description	Sequence	SEQ ID NO:
SaCas9 *Staphylococcus aureus* 1053 AA 123 kDa	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEG PGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLV ITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPE FTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINE MQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYET FKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLM NLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIAN ADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIK DFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGV YKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKING ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYST DILGNLYEVKSKKHPQIIKK	21
NmeCas9 *N. meningitidis* 1083 AA 124.5 kDa	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTG DSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPN TPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKG VAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILL FEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAA KNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARK LLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLS PELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIV PLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARK VINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREY FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDAALPFSRT WDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRS KKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMN AFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVS VLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKY DKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYY LVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKA RMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEI RPCRLKKRPPVR	276

Description	Sequence	SEQ ID NO:
CjCas9 *C. jejuni* 984 AA 114.9 kDa	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSAR KRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRAL NELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVG EYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSF SKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIIN LLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKG TYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQ IDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFL PAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNH SQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGE KIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGN DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVL NYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKD RNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRK FFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVL KALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNK AVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSST VSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVT KAEFRQREDFKK	277
GeoCas9 *G.* *stearothermophi* *lus* 1087 AA 127 kDa	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARS ARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKL NNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRTVGE MIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEY ITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISP SGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQ NENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSY LRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQG EVYSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQARKVVNAIIKK YGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTG HDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVL VLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDE NEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSR WEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKTEPHF PQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRMPKRSVT GAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEA IRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVA YNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDY TFRFSLYPNDLIRIELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDH RFSLRGVGSRTLKRFEKYQVDVLGNIYKVRGEKRVGLASSAHSKPGKTIRPLQS TRD	278
LbaCas12a *L. bacterium* 1228 AA 143.9 kDa	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKL LDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEE AKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFE GEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPL YKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEY SSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDR RKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADF VLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLA YDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATI LRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFS KKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNA YDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIY NKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVV HPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINT EVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIK TDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIA	279

Description	Sequence	SEQ ID NO:
	LEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQ ITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFD RIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFD WEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQM RNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKV LWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH	
BhCas12b *B. hisashii* 1108 AA 130.4 kDa	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDP KNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKK WEEDKKKDPLAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLD KDMFIQALERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYE KERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDY QRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQA TFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTE SGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGAR VQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVN FKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPD IEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLN FLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKD WVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLR WSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYD VRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQ GEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLT LDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHG FYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLK IKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAG VFFGKLERILISKLTNQYSISTIEDDSSKQSM	280

G. Cas9 equivalents

[0245] In some embodiments, the prime editors utilized in the methods and compositions described herein can include any Cas9 equivalent. As used herein, the term "Cas9 equivalent" is a broad term that encompasses any napDNAbp protein that serves the same function as Cas9 in the prime editors despite that its amino acid primary sequence and/or its three-dimensional structure may be different and/or unrelated from an evolutionary standpoint. Thus, while Cas9 equivalents include any Cas9 ortholog, homolog, mutant, or variant described or embraced herein that are evolutionarily related, the Cas9 equivalents also embrace proteins that may have evolved through convergent evolution processes to have the same or similar function as Cas9, but that do not necessarily have any similarity with regard to amino acid sequence and/or three-dimensional structure. The prime editors utilized in the methods and compositions described here embrace any Cas9 equivalent that would provide the same or similar function as Cas9 despite that the Cas9 equivalent may be based on a protein that arose through convergent evolution. For instance, if Cas9 refers to a type II enzyme of the CRISPR-Cas system, a Cas9 equivalent can refer to a type V or type VI enzyme of the CRISPR-Cas system.

[0246] For example, Cas12e (CasX) is a Cas9 equivalent that reportedly has the same function as Cas9 but which evolved through convergent evolution. Thus, the Cas12e (CasX) protein described in Liu *et al.*, "CasX enzymes comprises a distinct family of RNA-guided genome editors," *Nature*, 2019, Vol.566: 218-223, is contemplated to be used with the prime editors utilized in the methods and compositions described herein. In addition, any variant or modification of Cas12e (CasX) is conceivable and within the scope of the present disclosure.

[0247] Cas9 is a bacterial enzyme that evolved in a wide variety of species. However, the Cas9 equivalents contemplated herein may also be obtained from archaea, which constitute a domain and kingdom of single-celled prokaryotic microbes different from bacteria.

In some embodiments, Cas9 equivalents may refer to Cas12e (CasX) or Cas12d (CasY), which have been described in, for example, Burstein *et al.*, "New CRISPR–Cas systems from uncultivated microbes." *Cell Res.* 2017 Feb 21. doi: 10.1038/cr.2017.21, the entire contents of which is hereby incorporated by reference. Using genome-resolved metagenomics, a number of CRISPR–Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in little-studied nanoarchaea as part of an active CRISPR–Cas system. In bacteria, two previously unknown systems were discovered, CRISPR– Cas12e and CRISPR– Cas12d, which are among the most compact systems yet discovered. In some embodiments, Cas9 refers to Cas12e, or a variant of Cas12e. In some embodiments, Cas9 refers to a Cas12d, or a variant of Cas12d. It should be appreciated that other RNA-guided DNA binding proteins may be used as a nucleic acid programmable DNA binding protein (napDNAbp) and are within the scope of this disclosure. Also see Liu *et al.*, "CasX enzymes comprises a distinct family of RNA-guided genome editors," *Nature*, 2019, Vol.566: 218-223. Any of these Cas9 equivalents are contemplated.

[0248] In some embodiments, the Cas9 equivalent 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 Cas12e (CasX) or Cas12d (CasY) protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12e (CasX) or Cas12d (CasY) protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a wild-type Cas moiety or any Cas moiety provided herein.

[0249] In various embodiments, the nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (*e.g.,* dCas9 and nCas9), Cas12e (CasX), Cas12d (CasY),

Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), Cas12c (C2c3), Argonaute, and Cas12b1. One example of a nucleic acid programmable DNA-binding protein that has different PAM specificity than Cas9 is Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1 (i.e., Cas12a (Cpf1)). Similar to Cas9, Cas12a (Cpf1) is also a Class 2 CRISPR effector, but it is a member of type V subgroup of enzymes, rather than the type II subgroup. It has been shown that Cas12a (Cpf1) mediates robust DNA interference with features distinct from Cas9. Cas12a (Cpf1) is a single RNA-guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN, or YTN). Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from *Acidaminococcus* and *Lachnospiraceae* are shown to have efficient genome-editing activity in human cells. Cpf1 proteins are known in the art and have been described previously, for example Yamano *et al.*, "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962; the entire contents of which is hereby incorporated by reference.

[0250] In still other embodiments, the Cas protein may include any CRISPR associated protein, including but not limited to, Cas12a, Cas12b1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof, and preferably comprising a nickase mutation (e.g., a mutation corresponding to the D10A mutation of the wild type Cas9 polypeptide of SEQ ID NO: 9).

[0251] In various other embodiments, the napDNAbp can be any of the following proteins: a Cas9, a Cas12a (Cpf1), a Cas12e (CasX), a Cas12d (CasY), a Cas12b1 (C2c1), a Cas13a (C2c2), a Cas12c (C2c3), a GeoCas9, a CjCas9, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, a circularly permuted Cas9, or an Argonaute (Ago) domain, or a variant thereof.

[0252] Exemplary Cas9 equivalent protein sequences can include the following:

Description	Sequence
AsCas12a (previously known as Cpf1)*Acidaminococcus sp. (strain BV3L6)*UniPr	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDA INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNV FSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSF PFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISH KKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGK ELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDES

otKB U2UMQ6	NEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMI PKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIA EKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAE LFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEAR ALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPII GIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIK DLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLN CLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPF VWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFE KNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPK LLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 281)
AsCas12a nickase (e.g., R1226A)	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTY ADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDA INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNV FSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSF PFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPL FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISH KKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGK ELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDES NEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN KEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMI PKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQ KGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIA EKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAE LFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEAR ALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPII GIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIK DLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLN CLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPF VWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFE KNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPK LLENDDSHAIDTMVALIRSVLQMANSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO: 282)
LbCas12a (previously known as Cpf1)*Lachnospiraceae bacterium GAM79* Ref Seq. WP_119623382.1	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQQELKEIMDD YYRTFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKIQNEKRKEICCYFTSDKRFK DLFNAKLITDILPNFIKDNKEYTEEEKAEKEQTRVLFQRFATAFTNYFNQRRNNFSEDNIS TAISFRIVNENSEIHLQNMRAFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYSVDFYD RELTQPGIEYYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFE SDQEVYDALNEFIKTMKKKEIIRRCVHLGQECDDYDLGKIYISSNKYEQISNALYGSWD TIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIADIDKIISLYGSEMDRTISAKKCI TEICDMAGQISIDPLVCNSDIKLLQNKEKTTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYF YSELEDVLEDFEGITTLYNHVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAI LLMRDQKFYLGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRS GQETYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNYDFHFSDT KDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEKGQIFLFQIYNKDFSVHSTGKDNLH TMYLKNLFSEENLKDIVLKLNGEAELFFRKASIKTPIVHKKGSVLVNRSYTQTVGNKEIR VSIPEEYYTEIYNYLNHIGKGKLSSEAQRYLDEGKIKSFTATKDIVKNYRYCCDHYFLHL PITINFKAKSDVAVNERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVN GYDYQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNAVVAME DLNYGFKTGRFKVERQVYQKFETMLIEKLNHYLVFKDREVCEEGGVLRGYQLTYIPESLK KVGKQCGFIFYVPAGYTSKIDPTTGFVNLFSFKNLTNRESRQDFVGKFDEIRYDRDKKM FEFSFDYNNYIKKGTILASTKWKVYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKMLQ RAGIEYHDGKDLKGQIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLNDKGEF FDTATADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQFPRNKLVQDNKTWFD FMQKKRYL (SEQ ID NO: 47)

PcCas12a – previously known at Cpf1 *Prevotella copri* Ref Seq. WP_119227726.1	MAKNFEDFKRLYSLSKTLRFEAKPIGATLDNIVKSGLLDEDEHRAASYVKVKKLIDEYH KVFIDRVLDDGCLPLENKGNNNSLAEYYESYVSRAQDEDAKKKFKEIQQNLRSVIAKKL TEDKAYANLFGNKLIESYKDKEDKKKIIDSDLIQFINTAESTQLDSMSQDEAKELVKEFW GFVTYFYGFFDNRKNMYTAEEKSTGIAYRLVNENLPKFIDNIEAFNRAITRPEIQENMGV LYSDFSEYLNVESIQEMFQLDYYNMLLTQKQIDVYNAIIGGKTDDEHDVKIKGINEYINL YNQQHKDDKLPKLKALFKQILSDRNAISWLPEEFNSDQEVLNAIKDCYERLAENVLGDK VLKSLLGSLADYSLDGIFIRNDLQLTDISQKMFGNWGVIQNAIMQNIKRVAPARKHKES EEDYEKRIAGIFKKADSFSISYINDCLNEADPNNAYFVENYFATFGAVNTPTMQRENLFA LVQNAYTEVAALLHSDYPTVKHLAQDKANVSKIKALLDAIKSLQHFVKPLLGKGDESD KDERFYGELASLWAELDTVTPLYNMIRNYMTRKPYSQKKIKLNFENPQLLGGWDANKE KDYATIILRRNGLYYLAIMDKDSRKLLGKAMPSDGECYEKMVYKFFKDVTTMIPKCST QLKDVQAYFKVNTDDYVLNSKAFNKPLTITKEVFDLNNVLYGKYKKFQKGYLTATGD NVGYTHAVNVWIKFCMDFLNSYDSTCIYDFSSLKPESYLSLDAFYQDANLLLYKLSFAR ASVSYINQLVEEGKMYLFQIYNKDFSEYSKGTPNMHTLYWKALFDERNLADVVYKLN GQAEMFYRKKSIENTHPTHPANHPILNKNKDNKKKESLFDYDLIKDRRYTVDKFMFHV PITMNFKSVGSENINQDVKAYLRHADDMHIIGIDRGERHLLYLVVIDLQGNIKEQYSLNE IVNEYNGNTYHTNYHDLLDVREEERLKARQSWQTIENIKELKEGYLSQVIHKITQLMVR YHAIVVLEDLSKGFMRSRQKVEKQVYQKFEKMLIDKLNYLVDKKTDVSTPGGLLNAY QLTCKSDSSQKLGKQSGFLFYIPAWNTSKIDPVTGFVNLLDTHSLNSKEKIKAFFSKFDAI RYNKDKKWFEFNLDYDKFGKKAEDTRTKWTLCTRGMRIDTFRNKEKNSQWDNQEVD LTTEMKSLLEHYYIDIHGNLKDAISAQTDKAFFTGLLHILKLTLQMRNSITGTETDYLVSP VADENGIFYDSRSCGNQLPENADANGAYNIARKGLMLIEQIKNAEDLNNVKFDISNKA WLNFAQQKPYKNG (SEQ ID NO: 283)
ErCas12a – previously known at Cpf1 *Eubacterium rectale* Ref Seq. WP_119223642.1	MFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKDYFKNRANCFSANDISSS SCHRIVNDNAEIFFSNALVYRRIVKNLSNDDINKISGDMKDSLKEMSLEEIYSYEKYGEFI TQEGISFYNDICGKVNLFMNLYCQKNKENKNLYKLRKLHKQILCIADTSYEVPYKFESD EEVYQSVNGFLDNISSKHIVERLRKIGENYNGYNLDKIYIVSKFYESVSQKTYRDWETIN TALEIHYNNILPGNGKSKADKVKKAVKNDLQKSITEINELVSNYKLCPDDNIKAETYIHE ISHILNNFEAQELKYNPEIHLVESELKASELKNVLDVIMNAFHWCSVFMTEELVDKDNN FYAELEEIYDEIYPVISLYNLVRNYVTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAII LMRDNLYYLGIFNAKNKPDKKIIEGNTSENKGDYKKMIYNLLPGPNKMIPKVFLSSKTG VETYKPSAYILEGYKQNKHLKSSKDFDITFCHDLIDYFKNCIAIHPEWKNFGFDFSDTST YEDISGFYREVELQGYKIDWTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSSGNDNLHT MYLKNLFSEENLKDIVLKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEAEEKDQFGNIQI VRKTIPENIYQELYKYFNDKSDKELSDEAAKLKNVVGHHEAATNIVKDYRYTYDKYFL HMPITINFKANKTSFINDRILQYIAKEKDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNI VNGYDYQIKLKQQEGARQIARKEWKEIGKIKEIKEGYLSLVIHEISKMVIKYNAIIAMED LSYGFKKGRFKVERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKN VGHQCGCIFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSDKNLFCFT FDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKDMEKTLEMTDIN WRDGHDLRQDIIDYEIVQHIFEIFKLTVQMRNSLSELEDRDYDRLISPVLNENNIFYDSAK AGDALPKDADANGAYCIALKGLYEIKQITENWKEDGKFSRDKLKISNKDWFDFIQNKR YL (SEQ ID NO: 284)
CsCas12a – previously known at Cpf1 *Clostridium sp. AF34-10BH* Ref Seq. WP_118538418.1	MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQQELKEIMDD YYRAFIEEKLGQIQGIQWNSLFQKMEETMEDISVRKDLDKIQNEKRKEICCYFTSDKRFK DLFNAKLITDILPNFIKDNKEYTEEEKAEKEQTRVLFQRFATAFTNYFNQRRNNFSEDNIS TAISFRIVNENSEIHLQNMRAFQRIEQQYPEEVCGMEEEYKDMLQEWQMKHIYLVDFYD RVLTQPGIEYYNGICGKINEHMNQFCQKNRINKNDFRMKKLHKQILCKKSSYYEIPFRFE SDQEVYDALNEFIKTMKEKEIICRCVHLGQKCDDYDLGKIYISSNKYEQISNALYGSWD TIRKCIKEEYMDALPGKGEKKEEKAEAAAKKEEYRSIADIDKIISLYGSEMDRTISAKKCI TEICDMAGQISTDPLVCNSDIKLLQNKEKTTEIKTILDSFLHVYQWGQTFIVSDIIEKDSYF YSELEDVLEDFEGITTLYNHVRSYVTQKPYSTVKFKLHFGSPTLANGWSQSKEYDNNAI LLMRDQKFYLGIFNVRNKPDKQIIKGHEKEEKGDYKKMIYNLLPGPSKMLPKVFITSRS GQETYKPSKHILDGYNEKRHIKSSPKFDLGYCWDLIDYYKECIHKHPDWKNYDFHFSDT KDYEDISGFYREVEMQGYQIKWTYISADEIQKLDEKGQIFLFQIYNKDFSVHSTGKDNLH TMYLKNLFSEENLKDIVLKLNGEAELFFRKASIKTPVVHKKGSVLVNRSYTQTVGDKEI RVSIPEEYYTEIYNYLNHIGRGKLSTEAQRYLEERKIKSFTATKDIVKNYRYCCDHYFLH LPITINFKAKSDIAVNERTLAYIAKKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSFNIVN GYDYQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNAVVAME DLNYGFKTGRFKVERQVYQKFETMLIEKLNYLVFKDREVCEEGGVLRGYQLTYIPESLK

	KVGKQCGFIFYVPAGYTSKIDPTTGFVNLFSFKNLTNRESRQDFVGKFDEIRYDRDKKM FEFSFDYNNYIKKGTMLASTKWKVYTNGTRLKRIVVNGKYTSQSMEVELTDAMEKML QRAGIEYHDGKDLKGQIVEKGIEAEIIDIFRLTVQMRNSRSESEDREYDRLISPVLNDKGE FFDTATADKTLPQDADANGAYCIALKGLYEVKQIKENWKENEQFPRNKLVQDNKTWF DFMQKKRYL (SEQ ID NO: 50)
BhCas12b *Bacillus hisashii* Ref Seq. WP_095142515.1	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKK VSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEANQLSN KFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILG KLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLK VKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRG WREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPE YPYLYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTE KLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDES IKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRD DFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQK PDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRN VLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLH KRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLE PGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSP GIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCV TTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGY FILKDGVYEWVNAGKLKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSG NVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSM (SEQ ID NO: 280)
ThCas12b *Thermomonas hydrothermalis* Ref Seq. WP_072754838	MSEKTTQRAYTLRLNRASGECAVCQNNSCDCWHDALWATHKAVNRGAKAFGDWLLT LRGGLCHTLVEMEVPAKGNNPPQRPTDQERRDRRVLLALSWLSVEDEHGAPKEFIVAT GRDSADDRAKKVEEKLREILEKRDFQEHEIDAWLQDCGPSLKAHIREDAVWVNRRALF DAAVERIKTLTWEEAWDFLEPFFGTQYFAGIGDGKDKDDAEGPARQGEKAKDLVQKA GQWLSARFGIGTGADFMSMAEAYEKIAKWASQAQNGDNGKATIEKLACALRPSEPPTL DTVLKCISGPGHKSATREYLKTLDKKSTVTQEDLNQLRKLADEDARNCRKKVGKKGK KPWADEVLKDVENSCELTYLQDNSPARHREFSVMLDHAARRVSMAHSWIKKAEQRRR QFESDAQKLKNLQERAPSAVEWLDRFCESRSMTTGANTGSGYRIRKRAIEGWSYVVQA WAEASCDTEDKRIAAARKVQADPEIEKFGDIQLFEALAADEAICVWRDQEGTQNPSILID YVTGKTAEHNQKRFKVPAYRHPDELRHPVFCDFGNSRWSIQFAIHKEIRDRDKGAKQD TRQLQNRHGLKMRLWNGRSMTDVNLHWSSKRLTADLALDQNPNPNPTEVTRADRLG RAASSAFDHVKIKNVFNEKEWNGRLQAPRAELDRIAKLEEQGKTEQAEKLRKRLRWYV SFSPCLSPSGPFIVYAGQHNIQPKRSGQYAPHAQANKGRARLAQLILSRLPDLRILSVDLG HRFAAACAVWETLSSDAFRREIQGLNVLAGGSGEGDLFLHVEMTGDDGKRRTVVYRRI GPDQLLDNTPHPAPWARLDRQFLIKLQGEDEGVREASNEELWTVHKLEVEVGRTVPLID RMVRSGFGKTEKQKERLKKLRELGWISAMPNEPSAETDEKEGEIRSISRSVDELMSSAL GTLRLALKRHGNRARIAFAMTADYKPMPGGQKYYFHEAKEASKNDDETKRRDNQIEFL QDALSLWHDLFSSPDWEDNEAKKLWQNHIATLPNYQTPEEISAELKRVERNKKRKENR DKLRTAAKALAENDQLRQHLHDTWKERWESDDQQWKERLRSLKDWIFPRGKAEDNPS IRHVGGLSITRINTISGLYQILKAFKMRPEPDDLRKNIPQKGDDELENFNRRLLEARDRLR EQRVKQLASRIIEAALGVGRIKIPKNGKLPKRPRTTVDTPCHAVVIESLKTYRPDDLRTR RENRQLMQWSSAKVRKYLKEGCELYGLHFLEVPANYTSRQCSRTGLPGIRCDDVPTGD FLKAPWWRRAINTAREKNGGDAKDRFLVDLYDHLNNLQSKGEALPATVRVPRQGGNL FIAGAQLDDTNKERRAIQADLNAAANIGLRALLDPDWRGRWWYVPCKDGTSEPALDRI EGSTAFNDVRSLPTGDNSSRRAPREIENLWRDPSGDSLESGTWSPTRAYWDTVQSRVIE LLRRHAGLPTS (SEQ ID NO: 285)
LsCas12b *Laceyella sacchari* WP_132221894.1	MSIRSFKLKLKTKSGVNAEQLRRGLWRTHQLINDGIAYYMNWLVLLRQEDLFIRNKET NEIEKRSKEEIQAVLLERVHKQQQRNQWSGEVDEQTLLQALRQLYEEIVPSVIGKSGNA SLKARFFLGPLVDPNNKTTKDVSKSGPTPKWKKMKDAGDPNWVQEYEKYMAERQTL VRLEEMGLIPLFPMYTDEVGDIHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVR ERRAQFEKKTHDFASRFSESDVQWMNKLREYEAQQEKSLEENAFAPNEPYALTKKALR GWERVYHSWMRLDSAASEEAYWQEVATCQTAMRGEFGDPAIYQFLAQKENHDIWRG YPERVIDFAELNHLQRELRRAKEDATFTLPDSVDHPLWVRYEAPGGTNIHGYDLVQDT KRNLTLILDKFILPDENGSWHEVKKVPFSLAKSKQFHRQVWLQEEQKQKKREVVFYDY STNLPHLGTLAGAKLQWDRNFLNKRTQQQIEETGEIGKVFFNISVDVRPAVEVKNGRLQ NGLGKALTVLTHPDGTKIVTGWKAEQLEKWVGESGRVSSLGLDSLSEGLRVMSIDLGQ

	RTSATVSVFEITKEAPDNPYKFFYQLEGTEMFAVHQRSFLLALPGENPPQKIKQMREIRW KERNRIKQQVDQLSAILRLHKKVNEDERIQAIDKLLQKVASWQLNEEIATAWNQALSQL YSKAKENDLQWNQAIKNAHHQLEPVVGKQISLWRKDLSTGRQGIAGLSLWSIEELEAT KKLLTRWSKRSREPGVVKRIERFETFAKQIQHHINQVKENRLKQLANLIVMTALGYKYD QEQKKWIEVYPACQVVLFENLRSYRFSFERSRRENKKLMEWSHRSIPKLVQMQGELFG LQVADVYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEHRPYLQQG DLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRFWHPSMWFRVNCESVMEGE IVTYVPKNKTVHKKQGKTFRFVKVEGSDVYEWAKWSKNRNKNTFSSITERKPPSSMILF RDPSGTFFKEQEWVEQKTFWGKVQSMIQAYMKKTIVQRMEE (SEQ ID NO: 286)
DtCas12b *Dsulfonatronum thiodismutans WP_031386437*	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSYWTLDRRGDPVHVPES QVAEDALAMAREAQRRNGWPVVGEDEEILLALRYLYEQIVPSCLLDDLGKPLKGDAQK IGTNYAGPLFDSDTCRRDEGKDVACCGPFHEVAGKYLGALPEWATPISKQEFDGKDAS HLRFKATGGDDAFFRVSIEKANAWYEDPANQDALKNKAYNKDDWKKEKDKGISSWA VKYIQKQLQLGQDPRTEVRRKLWLELGLLPLFIPVFDKTMVGNLWNRLAVRLALAHLL SWESWNHRAVQDQALARAKRDELAALFLGMEDGFAGLREYELRRNESIKQHAFEPVD RPYVVSGRALRSWTRVREEWLRHGDTQESRKNICNRLQDRLRGKFGDPDVFHWLAED GQEALWKERDCVTSFSLLNDADGLLEKRKGYALMTFADARLHPRWAMYEAPGGSNLR TYQIRKTENGLWADVVLLSPRNESAAVEEKTFNVRLAPSGQLSNVSFDQIQKGSKMVG RCRYQSANQQFEGLLGGAEILFDRKRIANEQHGATDLASKPGHVWFKLTLDVRPQAPQ GWLDGKGRPALPPEAKHFKTALSNKSKFADQVRPGLRVLSVDLGVRSFAACSVFELVR GGPDQGTYFPAADGRTVDDPEKLWAKHERSFKITLPGENPSRKEEIARRAAMEELRSLN GDIRRLKAILRLSVLQEDDPRTEHLRLFMEAIVDDPAKSALNAELFKGFGDDRFRSTPDL WKQHCHFFHDKAEKVVAERFSRWRTETRPKSSSWQDWRERRGYAGGKSYWAVTYLE AVRGLILRWNMRGRTYGEVNRQDKKQFGTVASALLHHINQLKEDRIKTGADMIIQAAR GFVPRKNGAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWSHREIVNEVGMQ GELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEEDFHDGLPGMHLVGELDWLLPKD KDRTANEARRLLGGMVRPGMLVPWDGGELFATLNAASQLHVIHADINAAQNLQRRFW GRCGEAIRIVCNQLSVDGSTRYEMAKAPKARLLGALQQLKNGDAPFHLTSIPNSQKPEN SYVMTPTNAGKKYRAGPGEKSSGEEDELALDIVEQAEELAQGRKTFFRDPSGVFFAPDR WLPSEIYWSRIRRRIWQVTLERNSSGRQERAEMDEMPY (SEQ ID NO: 287)

[0253]

[0254] The prime editors utilized in the methods and compositions described herein may also comprise Cas12a (Cpf1) (dCpf1) variants that may be used as a guide nucleotide sequence-programmable DNA-binding protein domain. The Cas12a (Cpf1) protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cas12a (Cpf1) does not have the alfa-helical recognition lobe of Cas9. It was shown in Zetsche *et al.*, *Cell*, 163, 759–771, 2015 (which is incorporated herein by reference) that, the RuvC-like domain of Cas12a (Cpf1) is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cas12a (Cpf1) nuclease activity. In some embodiments, the napDNAbp is a single effector of a microbial CRISPR-Cas system. Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cas12a (Cpf1), Cas12b1 (C2c1), Cas13a (C2c2), and Cas12c (C2c3). Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multi-subunit effector complexes, while Class 2 systems have a single protein effector. For example, Cas9 and Cas12a (Cpf1) are Class 2 effectors. In addition to Cas9 and Cas12a (Cpf1), three distinct Class 2 CRISPR-Cas systems (Cas12b1, Cas13a, and Cas12c) 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 are hereby incorporated by reference.

[0255] Effectors of two of the systems, Cas12b1 and Cas12c, contain RuvC-like endonuclease domains related to Cas12a. A third system, Cas13a contains an effector with two predicted HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by Cas12b1. Cas12b1 depends on both CRISPR RNA and tracrRNA for DNA cleavage. Bacterial Cas13a has been shown to possess a unique RNase activity for CRISPR RNA maturation distinct from its RNA-activated single-stranded RNA degradation activity. These RNase functions are different from each other and from the CRISPR RNA-processing behavior of Cas12a. See, e.g., East-Seletsky, *et al.*, "Two distinct RNase activities of CRISPR-Cas13a enable guide-RNA processing and RNA detection", *Nature,* 2016 Oct 13;538(7624):270-273, the entire contents of which are hereby incorporated by reference. *In vitro* biochemical analysis of Cas13a in *Leptotrichia shahii* has shown that Cas13a is guided by a single CRISPR RNA and can be programed to cleave ssRNA targets carrying complementary protospacers. Catalytic residues in the two conserved HEPN domains mediate cleavage. Mutations in the catalytic residues generate catalytically inactive RNA-binding proteins. See e.g., Abudayyeh *et al.*, "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", *Science,* 2016 Aug 5; 353(6299), the entire contents of which are hereby incorporated by reference.

[0256] The crystal structure of *Alicyclobaccillus acidoterrastris* Cas12b1 (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 C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems. In some embodiments, the napDNAbp may be a C2c1, a C2c2, or a C2c3 protein. In some embodiments, the napDNAbp is a C2c1 protein. In some embodiments, the napDNAbp is a

Cas13a protein. In some embodiments, the napDNAbp is a Cas12c protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12b1 (C2c1), Cas13a (C2c2), or Cas12c (C2c3) protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b1 (C2c1), Cas13a (C2c2), or Cas12c (C2c3) protein.

H. Cas9 circular permutants

[0257] In various embodiments, the prime editors utilized in the methods and compositions disclosed herein may comprise a circular permutant of Cas9.

[0258] The term "circularly permuted Cas9" or "circular permutant" of Cas9 or "CP-Cas9") refers to any Cas9 protein, or variant thereof, that occurs or has been modified or engineered as a circular permutant variant, which means the N-terminus and the C-terminus of a Cas9 protein (e.g., a wild type Cas9 protein) have been topically rearranged. Such circularly permuted Cas9 proteins, or variants thereof, retain the ability to bind DNA when complexed with a guide RNA (gRNA). See, Oakes *et al.*, "Protein Engineering of Cas9 for enhanced function," *Methods Enzymol*, 2014, 546: 491–511 and Oakes *et al.*, "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell*, January 10, 2019, 176: 254-267, each of are incorporated herein by reference. The instant disclosure contemplates any previously known CP-Cas9 or use a new CP-Cas9 so long as the resulting circularly permuted protein retains the ability to bind DNA when complexed with a guide RNA (gRNA).

Any of the Cas9 proteins described herein, including any variant, ortholog, or any engineered or naturally occurring Cas9 or equivalent thereof, may be reconfigured as a circular permutant variant.

In various embodiments, the circular permutants of Cas9 may have the following structure: N-terminus-[original C-terminus] – [optional linker] – [original N-terminus]-C-terminus.

[0259] As an example, the present disclosure contemplates the following circular permutants of canonical *S. pyogenes* Cas9 (1368 amino acids of UniProtKB - Q99ZW2 (CAS9_STRP1) (numbering is based on the amino acid position in SEQ ID NO: 9)):

N-terminus-[1268-1368]-[optional linker]-[1-1267]-C-terminus;

N-terminus-[1168-1368]-[optional linker]-[1-1167]-C-terminus;

N-terminus-[1068-1368]-[optional linker]-[1-1067]-C-terminus;

N-terminus-[968-1368]-[optional linker]-[1-967]-C-terminus;

N-terminus-[868-1368]-[optional linker]-[1-867]-C-terminus;

N-terminus-[768-1368]-[optional linker]-[1-767]-C-terminus;
N-terminus-[668-1368]-[optional linker]-[1-667]-C-terminus;
N-terminus-[568-1368]-[optional linker]-[1-567]-C-terminus;
N-terminus-[468-1368]-[optional linker]-[1-467]-C-terminus;
N-terminus-[368-1368]-[optional linker]-[1-367]-C-terminus;
N-terminus-[268-1368]-[optional linker]-[1-267]-C-terminus;
N-terminus-[168-1368]-[optional linker]-[1-167]-C-terminus;
N-terminus-[68-1368]-[optional linker]-[1-67]-C-terminus; or
N-terminus-[10-1368]-[optional linker]-[1-9]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc.).

[0260] In particular embodiments, the circular permutant Cas9 has the following structure (based on *S. pyogenes* Cas9 (1368 amino acids of UniProtKB - Q99ZW2 (CAS9_STRP1) (numbering is based on the amino acid position in SEQ ID NO: 9):
N-terminus-[102-1368]-[optional linker]-[1-101]-C-terminus;
N-terminus-[1028-1368]-[optional linker]-[1-1027]-C-terminus;
N-terminus-[1041-1368]-[optional linker]-[1-1043]-C-terminus;
N-terminus-[1249-1368]-[optional linker]-[1-1248]-C-terminus; or
N-terminus-[1300-1368]-[optional linker]-[1-1299]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc.).

[0261] In still other embodiments, the circular permeant Cas9 has the following structure (based on *S. pyogenes* Cas9 (1368 amino acids of UniProtKB - Q99ZW2 (CAS9_STRP1) (numbering is based on the amino acid position in SEQ ID NO: 9):
N-terminus-[103-1368]-[optional linker]-[1-102]-C-terminus;
N-terminus-[1029-1368]-[optional linker]-[1-1028]-C-terminus;
N-terminus-[1042-1368]-[optional linker]-[1-1041]-C-terminus;
N-terminus-[1250-1368]-[optional linker]-[1-1249]-C-terminus; or
N-terminus-[1301-1368]-[optional linker]-[1-1300]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc.).

[0262] In some embodiments, the circular permutant can be formed by linking a C-terminal fragment of a Cas9 to an N-terminal fragment of a Cas9, either directly or by using a linker, such as an amino acid linker. In some embodiments, The C-terminal fragment may correspond to the C-terminal 95% or more of the amino acids of a Cas9 (e.g., amino acids about 1300-1368), or the C-terminal 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% or more of a Cas9 (e.g., any one of SEQ

ID NOs: 54-63). The N-terminal portion may correspond to the N-terminal 95% or more of the amino acids of a Cas9 (e.g., amino acids about 1-1300), or the N-terminal 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% or more of a Cas9 (e.g., of SEQ ID NO: 9).

[0263] In some embodiments, the circular permutant can be formed by linking a C-terminal fragment of a Cas9 to an N-terminal fragment of a Cas9, either directly or by using a linker, such as an amino acid linker. In some embodiments, the C-terminal fragment that is rearranged to the N-terminus includes or corresponds to the C-terminal 30% or less of the amino acids of a Cas9 (*e.g.,* amino acids 1012-1368 of SEQ ID NO: 9). In some embodiments, the C-terminal fragment that is rearranged to the N-terminus includes or corresponds to the C-terminal 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the amino acids of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 9). In some embodiments, the C-terminal fragment that is rearranged to the N-terminus includes or corresponds to the C-terminal 410 residues or less of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 9). In some embodiments, the C-terminal portion that is rearranged to the N-terminus includes or corresponds to the C-terminal 410, 400, 390, 380, 370, 360, 350, 340, 330, 320, 310, 300, 290, 280, 270, 260, 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20, or 10 residues of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 9). In some embodiments, the C-terminal portion that is rearranged to the N-terminus includes or corresponds to the C-terminal 357, 341, 328, 120, or 69 residues of a Cas9 (e.g., the Cas9 of SEQ ID NO: 9).

[0264] In other embodiments, circular permutant Cas9 variants may be defined as a topological rearrangement of a Cas9 primary structure based on the following method, which is based on *S. pyogenes* Cas9 of SEQ ID NO: 9: (a) selecting a circular permutant (CP) site corresponding to an internal amino acid residue of the Cas9 primary structure, which dissects the original protein into two halves: an N-terminal region and a C-terminal region; (b) modifying the Cas9 protein sequence (e.g., by genetic engineering techniques) by moving the original C-terminal region (comprising the CP site amino acid) to precede the original N-terminal region, thereby forming a new N-terminus of the Cas9 protein that now begins with the CP site amino acid residue. The CP site can be located in any domain of the Cas9 protein, including, for example, the helical-II domain, the RuvCIII domain, or the CTD domain. For example, the CP site may be located (relative the *S. pyogenes* Cas9 of SEQ ID NO: 9) at original amino acid residue 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247,

1249, or 1282. Thus, once relocated to the N-terminus, original amino acid 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, or 1282 would become the new N-terminal amino acid. Nomenclature of these CP-Cas9 proteins may be referred to as Cas9-CP^{181}, Cas9-CP^{199}, Cas9-CP^{230}, Cas9-CP^{270}, Cas9-CP^{310}, Cas9-CP^{1010}, Cas9-CP^{1016}, Cas9-CP^{1023}, Cas9-CP^{1029}, Cas9-CP^{1041}, Cas9-CP^{1247}, Cas9-CP^{1249}, and Cas9-CP^{1282}, respectively. This description is not meant to be limited to making CP variants from SEQ ID NO: 9, but may be implemented to make CP variants in any Cas9 sequence, either at CP sites that correspond to these positions, or at other CP sites entirely. This description is not meant to limit the specific CP sites in any way. Virtually any CP site may be used to form a CP-Cas9 variant.

[0265] Exemplary CP-Cas9 amino acid sequences, based on the Cas9 of SEQ ID NO: 9, are provided below in which linker sequences are indicated by underlining and optional methionine (M) residues are indicated in bold. It should be appreciated that the disclosure provides CP-Cas9 sequences that do not include a linker sequence or that include different linker sequences. It should be appreciated that CP-Cas9 sequences may be based on Cas9 sequences other than that of SEQ ID NO: 9 and any examples provided herein are not meant to be limiting. Exemplary CP-Cas9 sequences are as follows:

CP name	Sequence	SEQ ID NO:
CP1012	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GD<u>GGSGGSGGSGGSGGSGGSGG</u>DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY LALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD GFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG	288
CP1028	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL	289

CP name	Sequence	SEQ ID NO:
	FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGG SGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLI AQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ	
CP1041	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKY SIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDI QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGT ALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS	290
CP1249	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG DGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT	291

CP name	Sequence	SEQ ID NO:
	LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS	
CP1300	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGV DAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIE KILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD	292

[0266] The Cas9 circular permutants may be useful in the prime editing constructs utilized in the methods and compositions described herein. Exemplary C-terminal fragments of Cas9, based on the Cas9 of SEQ ID NO: 2, which may be rearranged to an N-terminus of Cas9, are provided below. It should be appreciated that such C-terminal fragments of Cas9 are exemplary and are not meant to be limiting. These exemplary CP-Cas9 fragments have the following sequences:

CP name	Sequence	SEQ ID NO:
CP1012 C-terminal fragment	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGD	54
CP1028 C-	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL	58

CP name	Sequence	SEQ ID NO:
terminal fragment	FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
CP1041 C-terminal fragment	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	61
CP1249 C-terminal fragment	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG D	62
CP1300 C-terminal fragment	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGD	293

I. Cas9 variants with modified PAM specificities

[0267] The prime editors utilized in the methods and compositions of the present disclosure may also comprise Cas9 variants with modified PAM specificities. Some aspects of this disclosure provide Cas9 proteins that exhibit activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′, where N is A, C, G, or T) at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGG-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNG-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNT-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGT-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAT-3′ PAM sequence at its 3′-end. In still other embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAG-3′ PAM sequence at its 3′-end.

[0268] It should be appreciated that any of the amino acid mutations described herein, (e.g., A262T) from a first amino acid residue (e.g., A) to a second amino acid residue (e.g., T) may also include mutations from the first amino acid residue to an amino acid residue that is similar to (e.g., conserved) the second amino acid residue. For example, mutation of an amino acid with a hydrophobic side chain (e.g., alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, or tryptophan) may be a mutation to a second amino acid with a different hydrophobic side chain (e.g., alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, or tryptophan). For example, a mutation of an alanine to a threonine (e.g., a A262T mutation) may also be a mutation from an alanine to an amino acid that is similar in size and chemical properties to a threonine, for example, serine. As another example, mutation of an amino acid with a positively charged side chain (e.g., arginine, histidine, or lysine) may be a mutation to a second amino acid with a different positively charged side chain (e.g., arginine, histidine, or lysine). As another example, mutation of an amino acid with a polar side chain (e.g., serine, threonine, asparagine, or glutamine) may be a mutation to a second amino acid with a different polar side chain (e.g., serine, threonine, asparagine, or glutamine). Additional similar amino acid pairs include, but are not limited to, the following: phenylalanine and tyrosine; asparagine and glutamine; methionine and cysteine; aspartic acid and glutamic acid; and arginine and lysine. The skilled artisan would recognize that such conservative amino acid substitutions will likely have minor effects on protein structure and are likely to be well tolerated without compromising function. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a threonine may be an amino acid mutation to a serine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an arginine may be an amino acid mutation to a lysine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an isoleucine, may be an amino acid mutation to an alanine, valine, methionine, or leucine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a lysine may be an amino acid mutation to an arginine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an aspartic acid may be an amino acid mutation to a glutamic acid or asparagine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a valine may be an amino acid mutation to an alanine, isoleucine, methionine, or leucine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a glycine may be an amino acid mutation to an alanine. It should be appreciated, however, that additional conserved amino

acid residues would be recognized by the skilled artisan and any of the amino acid mutations to other conserved amino acid residues are also within the scope of this disclosure.

[0269] In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAA-3′ PAM sequence at its 3′-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 1. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 1. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 1.

[0270] Table 1: NAA PAM Clones

Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 9)
D177N, K218R, D614N, D1135N, P1137S, E1219V, A1320V, A1323D, R1333K
D177N, K218R, D614N, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, G715C, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A367T, K710E, R1114G, D1135N, P1137S, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D861N, D1135N, K1188R, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, V743I, R753G, E762G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, S1274R, A1320V, R1333K
A10T, I322V, S409I, E427G, A589S, R753G, D1135N, E1219V, Q1221H, H1264H, A1320V, R1333K
A10T, I322V, S409I, E427G, R753G, E757K, G865G, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, R753G, E757K, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, K599R, M631A, R654L, K673E, V743I, R753G, N758H, E762G, D1135N, D1180G, E1219V, Q1221H, Q1256R, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N869S, N1054D, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, L727I, V743I, R753G, E762G, R859S, N946D, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, N1317T, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, K1151E, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1080S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
A10T, I322V, S409I, E427G, E630K, R654L, K673E, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, G1223S, H1264Y, L1318S, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1801S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K

A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, M673I, N803S, N869S, G1077D, R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K
A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, R1114G, D1135N, E1219V, Q1221H, A1320V, R1333K

[0271] In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 1. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 1.

[0272] In some embodiments, the Cas9 protein exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 9. In some embodiments, the Cas9 protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 9 on the same target sequence. In some embodiments, the Cas9 protein exhibits an activity on a target sequence that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* as provided by SEQ ID NO: 9 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to an AAA, GAA, CAA, or TAA sequence. In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAC-3′ PAM sequence at its 3′-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 2. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 2. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 2.

[0273] Table 2: NAC PAM Clones

Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 9)
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N
T472I, R753G, D1332N, R1335Q, T1337N
D1135N, E1219V, D1332N, R1335Q, T1337N
T472I, R753G, K890E, D1332N, R1335Q, T1337N
I1057S, D1135N, P1301S, R1335Q, T1337N

T472I, R753G, D1332N, R1335Q, T1337N
T472I, R753G, Q771H, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, K1156E, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, V647I, R753G, N803S, K959N, G1030R, I1055E, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, E630G, T638P, V647A, G687R, N767D, N803S, K959N, R1114G, D1135N, E1219V, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, I1057T, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
E627K, M631I, T638P, R753G, N803S, K959N, Y1036H, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N
E627K, T638P, R753G, N803S, V875I, K959N, Y1016C, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N, I1348V
K608R, E627K, T638P, V647I, R654L, R753G, N803S, T804A, K848N, V922A, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R753G, N803S, V922A, K959N, K1014N, V1015A, R1114G, D1135N, K1156N, E1219V, N1252D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, A711T, R753G, K775R, K789E, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
I670S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, K797N, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, T740A, G752R, R753G, K797N, N803S, K948E, K959N, V1015A, Y1016S, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
I570T, A589V, K608R, E627K, T638P, V647I, R654L, Q716R, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1207G, E1219V, N1234D, D1332N, R1335Q, T1337N
K608R, E627K, R629G, T638P, V647I, R654L, Q740R, R753G, N803S, K959N, N990S, T995S, V1015A, Y1036D, R1114G, D1135N, E1207G, E1219V, N1234D, N1266H, D1332N, R1335Q, T1337N
I562F, V565D, I570T, K608R, L625S, E627K, T638P, V647I, R654I, G752R, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, N1177S, N1234D, D1332N, R1335Q, T1337N
I562F, I570T, K608R, E627K, T638P, V647I, R753G, E790A, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, D1180E, A1184T, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1015A, R1114G, D1127A, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, L625S, E627K, T638P, V647I, R654I, T703P, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1016A, R1114G, D1135N, E1219V, K1246E, D1332N, R1335Q, T1337N
K608R, E627K, T638P, V647I, R654L, K673E, R753G, E790A, N803S, K948E, K959N, R1114G, D1127G, D1135N, D1180E, E1219V, N1286H, D1332N, R1335Q, T1337N
K608R, L625S, E627K, T638P, V647I, R654I, I670T, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
E627K, M631V, T638P, V647I, K710E, R753G, N803S, N808D, K948E, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, H1349R

[0274] In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 2. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least

97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 2.

[0275] In some embodiments, the Cas9 protein exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 9. In some embodiments, the Cas9 protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 9 on the same target sequence. In some embodiments, the Cas9 protein exhibits an activity on a target sequence that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* as provided by SEQ ID NO: 9 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to an AAC, GAC, CAC, or TAC sequence.

[0276] In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAT-3′ PAM sequence at its 3′-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 3. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 3. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 3.

[0277] Table 3: NAT PAM Clones

Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 9)
K961E, H985Y, D1135N, K1191N, E1219V, Q1221H, A1320A, P1321S, R1335L
D1135N, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
V743I, R753G, E790A, D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, V748I, V743I, R753G, D853E, V922A, R1114G D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1286K, P1321S, D1322G, R1335L
M631L, R654L, R753G, K797E, D853E, V922A, D1012A, R1114G D1135N, G1218S, E1219V, Q1221H, P1249S, N1317K, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, D596Y, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, Q1256R, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, K710E, V750A, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L

F575S, M631L, K649R, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, K1156E, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
F575S, M631L, R654L, R664K, R753G, D853E, V922A, I1057G, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1308D, P1321S, D1322G, R1335L
M631L, R654L, R753G, D853E, V922A, R1114G, Y1131C, D1135N, E1150V, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, D853E, I1057V, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
M631L, R654L, R664K, R753G, I1057V, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L

[0278] The above description of various napDNAbps which can be used in connection with the prime editors is not meant to be limiting in any way. The prime editors may comprise the canonical SpCas9, or any ortholog Cas9 protein, or any variant Cas9 protein—including any naturally occurring variant, mutant, or otherwise engineered version of Cas9—that is known or which can be made or evolved through a directed evolutionary or otherwise mutagenic process. In various embodiments, the Cas9 or Cas9 variants have a nickase activity, i.e., only cleave of strand of the target DNA sequence. In other embodiments, the Cas9 or Cas9 variants have inactive nucleases, i.e., are "dead" Cas9 proteins. Other variant Cas9 proteins that may be used are those having a smaller molecular weight than the canonical SpCas9 (e.g., for easier delivery) or having modified or rearranged primary amino acid structure (e.g., the circular permutant formats). The prime editors utilized in the methods and compositions described herein may also comprise Cas9 equivalents, including Cas12a/Cpf1 and Cas12b proteins which are the result of convergent evolution. The napDNAbps used herein (e.g., SpCas9, Cas9 variant, or Cas9 equivalents) may also contain various modifications that alter/enhance their PAM specificities. Lastly, the application contemplates any Cas9, Cas9 variant, or Cas9 equivalent which has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.9% sequence identity to a reference Cas9 sequence, such as a references SpCas9 canonical sequences or a reference Cas9 equivalent (e.g., Cas12a/Cpf1).

In a particular embodiment, the Cas9 variant having expanded PAM capabilities is SpCas9 (H840A) VRQR (SEQ ID NO: 294), which has the following amino acid sequence (with the V, R, Q, R substitutions relative to the SpCas9 (H840A) being show in bold underline. In addition, the methionine residue in SpCas9 (H840) was removed for SpCas9 (H840A) VRQR):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY

HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD
NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY
DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATL
IHQSITGLYETRIDLSQLGGD (SEQ ID NO: 294)

[0279] In another particular embodiment, the Cas9 variant having expanded PAM capabilities is SpCas9 (H840A) VRER, which has the following amino acid sequence (with the V, R, E, R substitutions relative to the SpCas9 (H840A) of SEQ ID NO: 12 being shown in bold underline . In addition, the methionine residue in SpCas9 (H840) was removed for SpCas9 (H840A) VRER):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD
NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY
DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR
VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATL
IHQSITGLYETRIDLSQLGGD (SEQ ID NO: 295)

[0280] In some embodiments, the napDNAbp that functions with a non-canonical PAM sequence is an Argonaute protein. One example of such a nucleic acid programmable DNA binding protein is an Argonaute protein from *Natronobacterium gregoryi* (NgAgo). NgAgo is a ssDNA-guided endonuclease. NgAgo binds 5′ phosphorylated ssDNA of ~24 nucleotides (gDNA) to guide it to its target site and will make DNA double-strand breaks at the gDNA

site. In contrast to Cas9, the NgAgo–gDNA system does not require a protospacer-adjacent motif (PAM). Using a nuclease inactive NgAgo (dNgAgo) can greatly expand the bases that may be targeted. The characterization and use of NgAgo have been described in Gao *et al.*, *Nat Biotechnol.*, 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts *et al.*, *Nature*. 507(7491) (2014):258-61; and Swarts *et al.*, *Nucleic Acids Res.* 43(10) (2015):5120-9, each of which is incorporated herein by reference.

[0281] In some embodiments, the napDNAbp is a prokaryotic homolog of an Argonaute protein. Prokaryotic homologs of Argonaute proteins are known and have been described, for example, in Makarova K., *et al.*, "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", *Biol Direct.* 2009 Aug 25;4:29. doi: 10.1186/1745-6150-4-29, the entire contents of which is hereby incorporated by reference. In some embodiments, the napDNAbp is a Marinitoga piezophila Argonaute (MpAgo) protein. The CRISPR-associated Marinitoga piezophila Argonaute (MpAgo) protein cleaves single-stranded target sequences using 5′-phosphorylated guides. The 5′ guides are used by all known Argonautes. The crystal structure of an MpAgo-RNA complex shows a guide strand binding site comprising residues that block 5′ phosphate interactions. This data suggests the evolution of an Argonaute subclass with noncanonical specificity for a 5′-hydroxylated guide. See, e.g., Kaya *et al.*, "A bacterial Argonaute with noncanonical guide RNA specificity", *Proc Natl Acad Sci U S A*. 2016 Apr 12;113(15):4057-62, the entire contents of which are hereby incorporated by reference). It should be appreciated that other argonaute proteins may be used, and are within the scope of this disclosure.

[0282] Some aspects of the disclosure provide Cas9 domains that have different PAM specificities. Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region. 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 where a target base is placed within a 4 base region (*e.g.,* a "editing window"), which is 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), the entire contents of which are hereby incorporated by reference. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.,* NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM

sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference.

[0283] For example, a napDNAbp domain with altered PAM specificity, such as a domain with at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with wild type *Francisella novicida* Cpf1 (D917, E1006, and D1255) (SEQ ID NO: 296), which has the following amino acid sequence:

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSV
CISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQES
DLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDD
NLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLN
QSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLE
DDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEIL
ANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQ
SEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKE
PDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPIT
INFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVE
KQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICP
VTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINF
RNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSK
TGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIK
NEEYFEFVQNRNN (SEQ ID NO: 296)

[0284] An additional napDNAbp domain with altered PAM specificity, such as a domain having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with wild type *Geobacillus thermodenitrificans* Cas9 (SEQ ID NO: 31), which has the following amino acid sequence:

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERI
RRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTN
KENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREY
GNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVS
PGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAY
HKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEE
LIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVV
MRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTL
NPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREK
GNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLA
NFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAF

YQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMP
KRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHN
NDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIY
TIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYY
QTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRP
L (SEQ ID NO: 31)

[0285] In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) is a nucleic acid programmable DNA binding protein that does not require a canonical (NGG) PAM sequence. In some embodiments, the napDNAbp is an argonaute protein. One example of such a nucleic acid programmable DNA binding protein is an Argonaute protein from *Natronobacterium gregoryi* (NgAgo). NgAgo is a ssDNA-guided endonuclease. NgAgo binds 5′ phosphorylated ssDNA of ~24 nucleotides (gDNA) to guide it to its target site and will make DNA double-strand breaks at the gDNA site. In contrast to Cas9, the NgAgo–gDNA system does not require a protospacer-adjacent motif (PAM). Using a nuclease inactive NgAgo (dNgAgo) can greatly expand the bases that may be targeted. The characterization and use of NgAgo have been described in Gao *et al.*, *Nat Biotechnol.*, 34(7): 768-73 (2016), PubMed PMID: 27136078; Swarts *et al.*, *Nature*, 507(7491): 258-61 (2014); and Swarts *et al.*, *Nucleic Acids Res.* 43(10) (2015): 5120-9, each of which is incorporated herein by reference. The sequence of *Natronobacterium gregoryi* Argonaute is provided in SEQ ID NO: 297.

[0286] The disclosed fusion proteins may comprise a napDNAbp domain having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with wild type *Natronobacterium gregoryi* Argonaute (SEQ ID NO: 297), which has the following amino acid sequence:

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPKDVF
TYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHLDDALN
ETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQELYTDHDA
APVATDGLMLLTPEPLGETPLDLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLV
RGIDEVLSKEPVLTCDEFDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKT
TYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRRQGHGD
DAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDPVRLNGSTVEFSS
EFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQW
DTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETY
DELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVS
RSYPEDGASGQINIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHI
VIHRDGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVA
TFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHA
TKGYLVQTGAFESNVGFL (SEQ ID NO: 297)

[0287] In addition, any available methods may be utilized to obtain or construct a variant or mutant Cas9 protein. 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)). Mutations can include a variety of categories, such as single base polymorphisms, microduplication regions, indel, and inversions, and is not meant to be limiting in any way. Mutations can include "loss-of-function" mutations which is the normal result of a mutation that reduces or abolishes a protein activity. Most loss-of-function mutations are recessive, because in a heterozygote the second chromosome copy carries an unmutated version of the gene coding for a fully functional protein whose presence compensates for the effect of the mutation. Mutations also embrace "gain-of-function" mutations, which is one which confers an abnormal activity on a protein or cell that is otherwise not present in a normal condition. Many gain-of-function mutations are in regulatory sequences rather than in coding regions, and can therefore have a number of consequences. For example, a mutation might lead to one or more genes being expressed in the wrong tissues, these tissues gaining functions that they normally lack. Because of their nature, gain-of-function mutations are usually dominant.

[0288] Mutations can be introduced into a reference Cas9 protein using site-directed mutagenesis. Older methods of site-directed mutagenesis known in the art rely on sub-cloning of the sequence to be mutated into a vector, such as an M13 bacteriophage vector, that allows the isolation of single-stranded DNA template. In these methods, one anneals a mutagenic primer (i.e., a primer capable of annealing to the site to be mutated but bearing one or more mismatched nucleotides at the site to be mutated) to the single-stranded template and then polymerizes the complement of the template starting from the 3′ end of the mutagenic primer. The resulting duplexes are then transformed into host bacteria and plaques are screened for the desired mutation. More recently, site-directed mutagenesis has employed PCR methodologies, which have the advantage of not requiring a single-stranded template. In addition, methods have been developed that do not require sub-cloning. Several issues must be considered when PCR-based site-directed mutagenesis is performed. First, in these methods it is desirable to reduce the number of PCR cycles to prevent expansion of undesired mutations introduced by the polymerase. Second, a selection must be employed in order to reduce the number of non-mutated parental molecules persisting in the reaction. Third, an extended-length PCR method is preferred in order to allow the use of a single PCR primer

set. And fourth, because of the non-template-dependent terminal extension activity of some thermostable polymerases it is often necessary to incorporate an end-polishing step into the procedure prior to blunt-end ligation of the PCR-generated mutant product.

[0289] Mutations may also be introduced by directed evolution processes, such as phage-assisted continuous evolution (PACE) or phage-assisted noncontinuous evolution (PANCE). The term "phage-assisted continuous evolution (PACE)," as used herein, refers to continuous evolution that employs phage as viral vectors. The general concept of PACE technology has been described, for example, in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Application, U.S. Patent No. 9,023,594, issued May 5, 2015, International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015, and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference. Variant Cas9s may also be obtain by phage-assisted non-continuous evolution (PANCE)," which as used herein, refers to non-continuous evolution that employs phage as viral vectors. PANCE is a simplified technique for rapid in vivo directed evolution using serial flask transfers of evolving 'selection phage' (SP), which contain a gene of interest to be evolved, across fresh *E. coli* host cells, thereby allowing genes inside the host *E. coli* to be held constant while genes contained in the SP continuously evolve. Serial flask transfers have long served as a widely-accessible approach for laboratory evolution of microbes, and, more recently, analogous approaches have been developed for bacteriophage evolution. The PANCE system features lower stringency than the PACE system.

[0290] Any of the references noted above which relate to Cas9 or Cas9 equivalents are hereby incorporated by reference in their entireties, if not already stated so.

Reverse transcriptase domain and modified variants thereof

[0291] In various embodiments, the improved prime editors disclosed herein include a polymerase (e.g., DNA-dependent DNA polymerase or RNA-dependent DNA polymerase, such as reverse transcriptase), or a variant thereof, which can be provided as a fusion protein with a napDNAbp or other programmable nuclease, or provided *in trans*. In various

embodiments, the improved prime editors disclosed herein include optimized, evolved reverse transcriptases as described further below.

[0292] In some embodiments, the improved prime editor proteins comprise an MMLV reverse transcriptase comprising one or more amino acid substitutions. The wild-type MMLV reverse transcriptase is provided by the following sequence:

DESCRIPTION	SEQUENCE
REVERSE TRANSCRIPTASE (M-MLV RT) WILD TYPE MOLONEY MURINE LEUKEMIA VIRUS	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 33)

[0293] The reverse transcriptases used in the improved prime editors described herein may comprise one or more mutations relative to the wild-type amino acid sequence. In some embodiments, the reverse transcriptase is the MMLV pentamutant described above (*i.e.*, comprising amino acid substitutions D200N, T306K, W313F, T330P, and L603W).

[0294] In some embodiments, the present disclosure provides MMLV reverse transcriptase variants, and prime editors (*e.g.*, fusion proteins and prime editors in which the napDNAbp and reverse transcriptase are provided *in trans*) comprising MMLV reverse transcriptase variants, wherein the variants comprise one or more mutations relative to SEQ ID NO: 33 selected from the group consisting of T13I, V19I, A32T, G38V, S60Y, P111L, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, G138R, C157F, P175Q, P175S, D200S, D200Y, D200N, D200C, Y222F, V223A, V223M, V223T, V223W, V223Y, L234I, T246I, N249S, T287A, P292T, E302A, E302K, T306K, G316R, E346K, K373N, W388C, V402A, K445N, M457I, and A462S. In some embodiments, an MMLV reverse transcriptase variant comprises two or more of these mutations, three or more of these mutations, four or more of these mutations, or five or more of these mutations.

[0295] In some embodiments, the MMLV reverse transcriptase variants used in the prime editors provided herein comprise a single mutation relative to SEQ ID NO: 33. In some embodiments, the single mutations is selected from the group consisting of T13I, G38V, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, P175Q, P175S, D200C, D200Y, V223M, V223T, V223W, V223Y, L234I, P292T, G316R, K373N, M457I, and V402A.

[0296] In certain embodiments, the MMLV reverse transcriptase variants used in the prime editors provided herein comprise any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 33: D200Y and E302A; D200Y, V223A, and M457I; V223M, T306K, and A462S; D200N and E302K; D200Y and E302K; T128N and V223A; V19I, A32T, and D200Y; D200S, V223A, E346K, and W388C; S60Y, V223A, and N249S; P111L, V223A, T287A, and G316R; S60Y, G138R, and V223A; S60Y, Y222F, V223A, and K445N; or S60Y, C157F, V223A, and T246I. In certain embodiments, the MMLV reverse transcriptase variant used in the prime editors provided herein comprises the amino acid sequence of any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, wherein the amino acid sequence comprises at least one of residues 13I, 19I, 32T, 38V, 60Y, 111L, 120R, 126Y, 128N, 128F, 128H, 129S, 132S, 138R, 157F, 175Q, 175S, 200S, 200Y, 200N, 200C, 222F, 223A, 223M, 223T, 223W, 223Y, 234I, 246I, 249S, 287A, 292T, 302A, 302K, 306K, 316R, 346K, 373N, 388C, 402A, 445N, 457I, and 462S.

[0297] In other examples, the proteins described herein may comprise an MMLV reverse transcriptase comprising one or more substitutions at amino acid positions V19, A32, S60, P111, T128, G138R, C157F, D200, Y222, V223, T246, N249, T287, G316, E346, W388, and/or K445. In some embodiments, the proteins described herein comprise an MMLV reverse transcriptase comprising one or more substitutions selected from the group consisting of V19I, A32T, S60Y, P111L, T128N, G138R, C157F, D200S, D200Y, Y222F, V223A, T246I, N249S, T287A, G316R, E346K, W388C, and K445N. In certain embodiments, the proteins described herein comprise an MMLV reverse transcriptase comprising any one of

[0298] the following groups of amino acid substitutions:

[0299]

T128N and V223A;

V19I, A32T, and D200Y;

D200S, V223A, E346K, and W388C;

S60Y, V223A, and N249S;

P111L, V223A, T287A, and G316R;

S60Y, G138R, and V223A;

S60Y, Y222F, V223A, and K445N; or

S60Y, C157F, V223A, and T246I.

[0300] Exemplary evolved reverse transcriptase enzymes are as follows:

DESCRIPTION	SEQUENCE
REVERSE TRANSCRIPTASE (M-MLV RT) T128N and V223A	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP**N**V PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 35)
REVERSE TRANSCRIPTASE (M-MLV RT) V19I, A32T, and D200Y	TLNIEDEYRLHETSKEPD**I**SLGSTWLSDFPQ**T**WAETGGMGLAV RQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLF**Y**EALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 36)

REVERSE TRANSCRIPTASE (M-MLV RT) D200S, V223A, E346K, and W388C	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLF**S**EALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQ**K**IKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGP**C**RRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 37)
REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, V223A, and N249S	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLG**S**LGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 38)

REVERSE TRANSCRIPTASE (M-MLV RT) P111L, V223A, T287A, and G316R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYR**L**VQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKE**A**VMGQPTPKTPR QLREFLGTAGFCRLWIP**R**FAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 39)
REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, G138R, and V223A	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLS**R**LPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 40)

REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, Y222F, V223A, and K445N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQ**FA**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALV**N**QPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 41)
REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, C157F, V223A, and T246I	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFF**F**LRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQ**I**LGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGK LTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIH GEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 42)
REVERSE TRANSCRIPTASE (M-MLV RT) D200S, V223A, E346K, and W388C	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLF**S**EALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQ**K**IKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPW**C**RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 172)

REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, V223A, and N249S	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLG**S**LGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 173)
REVERSE TRANSCRIPTASE (M-MLV RT) P111L, V223A, T287A, and G316R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYR**L**VQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKE**A**VMGQPTPKTPR QLREFLGKAGFCRLFIP**R**FAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 174)
REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, G138R, and V223A	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLS**R**LPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 175)

REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, Y222F, V223A, and K445N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQ**FA**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALV**N**QPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 176)
REVERSE TRANSCRIPTASE (M-MLV RT) S60Y, C157F, V223A, and T246I	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPV**Y**IKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFF**F**LRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQY**A**DDLLLAATSELDCQQGTRALLQ**I**LGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 177)
REVERSE TRANSCRIPTASE (M-MLV RT) V223M, T306K, A462S	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQY**M**DDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQ**S**LL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 183)

REVERSE TRANSCRIPTASE (M-MLV RT) D200N and E302K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHP DLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPR QLRKFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALL LDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 184)

[0301] The use of reverse transcriptase enzymes comprising an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any of the evolved variants described herein in the improved prime editors disclosed herein is also contemplated by the present disclosure, provided the RT sequence comprises one of the amino acid substitutions disclosed herein.

[0302] The disclosure also contemplates the use of any wild-type reverse transcriptase in the improved prime editors described herein. Exemplary wild-type reverse transcriptases which may be used include, but are not limited to, the following sequences, or any variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity thereto:

DESCRIPTION	SEQUENCE	SEQ ID NO:
MOUSE MAMMARY TUMOR VIRUS (MMTV) REVERSE TRANSCRIPTASE	VFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQIS WKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSP WNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSP VAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQ RFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHY MDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPF LKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLST ARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLL QEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTP LEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVI TAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHL QRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT	43
AVIAN SARCOMA LEUKOSIS VIRUS (ASLV) REVERSE TRANSCRIPTASE	TVALHLAIPLKWKPDHTPVWIDQWPLPEGKLVALTQLVEKEL QLGHIEPSLSCWNTPVFVIRKASGSYRLLHDLRAVNAKLVPFG AVQQGAPVLSALPRGWPLMVLDLKDCFFSIPLAEQDREAFAF TLPSVNNQAPARRFQWKVLPQGMTCSPTICQLVVGQVLEPLR LKHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTI SPDKIQREPGVQYLGYKLGSTYVAPVGLVAEPRIATLWDVQK LVGSLQWLRPALGIPPRLMGPFYEQLRGSDPNEAREWNLDMK MAWREIVQLSTTAALERWDPALPLEGAVARCEQGAIGVLGQ GLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTF GKEVDILLLPACFREDLPLPEGILLALKGFAGKIRSSDTPSIFDIA RPLHVSLKVRVTDHPVPGPTVFTDASSSTHKGVVVWREGPRW EIKEIADSGASVQQLEARAVAMALLLWPTTPTNVVTDSAFVA KMLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSE VPGFFTEGNDVADSQATFQAY	44
PORCINE ENDOGENOUS RETROVIRUS (PERV) REVERSE TRANSCRIPTASE	TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLA KQVPPQVIQLKASATPVSVRQYPLSREAREGIWPHVQRLIQQG ILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPT VPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQ HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRAS AKKAQICRREVTYLGYSLRGGQRWLTEARKKTVVQIPAPTTA KQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEH QKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLT QTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDA DKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLL TERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKD LTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWAS SLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHV HGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPG HQKAKDLISRGNQMADRVAKQAAQAVNLLPI	45

HIV-MMLV REVERSE TRANSCRIPTASE	PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFKKQ NPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTP DKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQ KLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAEL ELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQE PFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKT PKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQR AELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRG WLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA RGNRMADQAARKAAITETPDTSTLLIEN	46
AVIRE REVERSE TRANSCRIPTASE	APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQ APIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRP VHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPN PYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADA EEGESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVS LLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREF LGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAF QSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGP WKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFG QDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFK QTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQ AEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQ KAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRER GLLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDD APTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYS NVEEALG	216
BABOON ENDOGENOUS VIRUS (BAEVM) REVERSE TRANSCRIPTASE	VSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQ APIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPC RSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPY NLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDP ERGISGQLTWTRLPQGFKNSPTLFDEALHRDLTDFRTQHPEVT LLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKA QICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPREVREF LGTAGFCRLWIPGFAELAAPLYALTKESTPFTWQTEHQLAFEA LKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWK RPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQ PLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFG PPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQELP DADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGT SAQKAELIALTKALELSKGKKANIYTDSRYAFATAHTHGSIYE RRGLLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQD PVAVGNRQADRVARQAAMAEVLTLATEPDNTSHITIEHTYTS EDQEEA	48

GIBBON APE LEUKEMIA VIRUS (GALV) REVERSE TRANSCRIPTASE	LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQV PPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLV PCRSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPN PYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKD PEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQ VVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAK KAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQ VREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQQA FDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLG PWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLT LGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERV SFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLEDQP LPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTS AQKAELVALTQALRLAEGKNINIYTDSRYAFATAHIHGAIYKQ RGLLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNP VATGNRRADEAAKQAALSTRVLAGTTKPQEPIEPAQEK	49
KOALA RETROVIRUS (KORV) REVERSE TRANSCRIPTASE	MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQ VPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGIL VPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTV PNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEW RDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALN PQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVS AKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTP RQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAH QEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQ TLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDAD KLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLN ERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRD QPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPE GTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAI YKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRG TDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK	222
MASON-PFIZER MONKEY VIRUS (MPMV) REVERSE TRANSCRIPTASE	MWGRDLLSQMKIMMCSPNDIVTAQMLAQGYSPGKGLGKKE NGILHPIPNQGQSNKKGFGNFLTAAIDILAPQQCAEPITWKSDE PVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIF VIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAIPQG YLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWK VLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDILI AGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGF ELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTT GDLKPLFDTLKGDSDPNSHRSLSKEALASLEKVETAIAEQFVT HINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPY YDAIADLIILGRDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEM WPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNAL LVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLS AFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQ LIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLATKIVA	51

POK11ERV REVERSE TRANSCRIPTASE	ATVEPPKPIPLTWKTEKPVWVNQWPLPKQKLEALHLLANEQL EKGHIEPSFSPWNSPVFVIQKKSGKWRMLTDLRAVNAVIQPM GPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIP AINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPVREKFS DCYIIHYIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKI QTSTPFHYLGMQIENRKIKPQKIEIRKDTLKTLNDFQKLLGDIN WIRPTLGIPTYAMSNLFSILRGDSDLNSKRILTPEATKEIKLVEE KIQSAQINRIDPLAPLQLLIFATAHSPTGIIIQNTDLVEWSFLPHS TVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDKIVVPLTKEQVR QAFINSGAWQIGLANFVGIIDNHYPKTKIFQFLKMTTWILPKIT RREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRAEL VAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQL NQLFNLLQQTVRKRNFPFYITHIRAHTNLPGPLTKANEEADLL VS	52
SIMIAN RETROVIRUS TYPE 2 (SRV2) REVERSE TRANSCRIPTASE	MWGRDLLSQMKIMMCSPNDIVTAQMLAQGYSPGKGLGKRE DGILQPIPNSGQLDRKGFGNFLATAVDILAPQRYADPITWKSD EPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPI FVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAIPQG YFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQW KVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMDDIL IAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGF QINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTT GDLKPLFDILKGDSNPNSPRSLSEAALASLQKVETAIAEQFVTQ IDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLP YYDAIADLIILGRDNSKKYFGLEPSTIIQPYSKSQIHWLMQNTE TWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDN ALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVL SAFPHRALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQ QLIYNRSIPFYLGHIRAHSGLPGPLSQGNHITDLATKVVA	53
WOOLLY MONKEY SARCOMA VIRUS (WMSV) REVERSE TRANSCRIPTASE	LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQV PPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLV PCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPN PYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRD PEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQ VVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAK KAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPPTTPRQ VREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKA FDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLG PWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLT LGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERV SFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQP LPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGT SAQKAELVALTQALRLAEGKDINIYTDSRYAFATAHIHGAIYK QRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGND PVATGNRRADEAAKQAALSTRVLAETTKPQELI	228

TF1 REVERSE TRANSCRIPTASE	ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPK PIKGLEFEVELTQENYRLPIRNYPLPPGKMQAMNDEINQGLKS GIIRESKAINACPVMFVPKKEGTLRMVVDYKPLNKYVKPNIYP LPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCP RGVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDI LIHSKSESEHVKHVKDVLQKLKNANLIINQAKCEFHQSQVKFI GYHISEKGFTPCQENIDKVLQWKQPKNRKELRQFLGSVNYLR KFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSP PVLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYY SAKMSKAQLNYSVSDKEMLAIIKSLKHWRHYLESTIEPFKILT DHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHI ADALSRIVDETEPIPKDSEDNSINFVNQISI	55
CRISPR REVERSE TRANSCRIPTASE	NSQAQSACCAGANQIVEGATLEKVVAPACLQQAWTRVRKNK GGPGGDGVTIEIFAQNAEVELEKLRAETLAGIYRPRKVRHAIV PKPKGGERKLTIPSVVDRILQTATMLSLGQTVDHHFSSASWAY REGRGVDDALADLRRLRNSGLFWTFDADIMQYFDRILHKRLI DDLFIWVDDLRIVRLIQLWLRSFSYWGRGIAQGAPISPLLANLF LHPMDRLLELEGLASVRYADDFVVLCRSKALAQKAQLIVASH LAARGLKLNMSKTRILAPSEAFIFLGQTVEPVWDTQP	56
VP96 REVERSE TRANSCRIPTASE	NLVKRLAHHLGKSEPEVIHFLADAPNKYRVYKIPKRSYGHRVI AQPTRELKLYQKAFLELYSFPVHSSATAYCKGKSIKDNALSHV KNHYLLKTDLENFFNSITPNIFWKSIENDSIATPKFSTSEIALVE RLIFWRPSKLQGGKLVLSVGAPSSPTISNFCLYQFDEYLSIICKE QNISYTRYADDLTFSTCDKDVLHTVIPLIQSLLDYFFASELKLN HSKTVFSSKAHNRHVTGITLNNEGKLSLGRERKRYIKHLVHSF KYGKLDNTEIRHLQGMLSFAKHIEPIFIDRLKEKYTDELIKIIYE AGHE	57
VC95 REVERSE TRANSCRIPTASE	NILTTLREQLLTNNVIMPQEFERLEVRGSHAYKVYSIPKRKAG RRTIAHPSSKLKICQRHLNAILNPLLKVHDSSYAYVKGRSIKDN ALVHSHSAYVLKMDFQNFFNSITPTILRQCLIQNDILLSVNELE KLEQLIFWNPSKKRNGKLILSVGSPISPLISNAIMYPFDKIINDIC TKHGINYTRYADDITFSTNIKNTLNKLPEIVEQLIIQTYAGRIIIN KRKTVFSSKKHNRHVTGITLTNDSKISIGRSRKRYISSLVFKYIN KNLDIDEINHMKGMLAFAYNIEPIYIHRLSHKYKVNIVEKILRG SN	241
EC48 REVERSE TRANSCRIPTASE	GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEE LKAIAELSLDEKYTLKEIPKIDGSKRIVYSLHPKMRLLQSRINK RIFKELVVFPSFLFGSVPSKNDVLNSNVKRDYVSCAKAHCGA KTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTK DDFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVD DITVSSKISNYDFSQMQSHIERMLSEHDLPINKHKTKIFHCSSEP IKVHGLRVDYDSPRLPSDEVKRIRASIHNLKLLAAKNNTKTSV AYRKEFNRCMGRVNKLGRVGHEKYESFKKQLQAIKPMPSKR DVAVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTR SESFKEKLECFKSRLASLKPL	59

GS REVERSE TRANSCRIPTASE	ALLERILARDNLITALKRVEANQGAPGIDGVSTDQLRDYIRAH WSTIHAQLLAGTYRPAPVRRVEIPKPGGGTRQLGIPTVVDRLI QQAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGY RYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYL QAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKR GLKFCRYADDCNIYVKSLRAGQRVKQSIQRFLEKTLKLKVNE EKSAVDRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLT NPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIR RRLRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGA WRTTKTPQLHQALGKTYWTAQGLKSLTQRYFELRQG	60
ER REVERSE TRANSCRIPTASE	DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKE HLAKNGETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPT VTDRFIQQAIAQVLTPIYEEQFHDHSYGFRPNRCAQQAILTALN IMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIV RKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEM EKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKV NMTKSKVDRPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKF KKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTL CKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARR VAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKC VTC	185
NE144 REVERSE TRANSCRIPTASE	AGQPTSREALYERIRSTSKEEVILEEMIRLGFWPAQGAVPHDP AEEIRRRGELERQLSELREKSRKLYNEKALIAEQRKQRLAESR RKQKETKARRERERQERAQKWAQRKAGEILFLGEDVSGGMS HKTCDAELIKREGVPAIASAEELARAMGIALKELRFLAYNRKV SRVTHYRRFLLPKKTGGLRLISAPMPRLKRAQAWALEHIFNKL SFEPAAHGFVAGRSIVSNARPHVGADVVVNLDLKDFFPTVSFP RVKGALRHLGYSESVATALALVCTEPEVDEVGLDGTTWYVA RGERFLPQGSPCSPAITNLLCRRLDRRLHGLAQALGFVYTRYA DDLTFSGRGEAAESKRVGKLLRGAADIVAHEGFVVHPDKTRV MRRGRRQEVTGVVVNDKTSVPRDELRKFRATLYQIEKDGPA DKRWGNGGDVLAAVHGYACFVAMVDPSRGQPLLARARALL AKHGGPSKPPGGSGPRAPTPVQPTANAPEAPKPVAPATPAAPA KKGWKLF	239

[0303] The use of reverse transcriptase enzymes comprising an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to any of the enzymes above in the improved prime editor proteins disclosed herein is also contemplated by the present disclosure.

[0304] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is an AVIRE reverse transcriptase of SEQ ID NO: 216, or an AVIRE reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 216, wherein the AVIRE reverse transcriptase variant comprises one or more mutations selected from the

group consisting of D199N, T305K, W312F, G329P, and L604W. In some embodiments, the AVIRE reverse transcriptase variant comprises two or more, three or more, four or more, or all five of these mutations. In some embodiments, the AVIRE reverse transcriptase variant comprises the mutation D199N. In some embodiments, the AVIRE reverse transcriptase variant comprises the mutation T305K. In some embodiments, the AVIRE reverse transcriptase variant comprises the mutation W312F. In some embodiments, the AVIRE reverse transcriptase variant comprises the mutation G329P. In some embodiments, the AVIRE reverse transcriptase variant comprises the mutation L604W.

[0305] In certain embodiments, the AVIRE reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 217-221, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 217-221, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 604W:

AVIRE-RT (D199N):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 217)

AVIRE-RT (T305K):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGKIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSL
KLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVA
AGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQ
VLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAE
ATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKD
KSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV

MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 218)

AVIRE-RT (W312F):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLFIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 219)

AVIRE-RT (G329P):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 220)

AVIRE-RT (L604W):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 221)

[0306] In certain embodiments, the AVIRE reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 243, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 243, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 604W:

AVIRE_penta:

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 243)

[0307] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a KORV reverse transcriptase of SEQ ID NO: 222, or a KORV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 222, wherein the KORV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W310F, E327P, and L599W. In some embodiments, the KORV reverse transcriptase variant comprises two or more, three or more, four or more, or all five of these mutations. In some embodiments, the KORV reverse transcriptase variant comprises the mutation D197N. In some embodiments, the KORV reverse transcriptase variant comprises the mutation T303K. In some embodiments, the KORV reverse transcriptase variant comprises the mutation W310F. In some embodiments, the KORV reverse transcriptase variant comprises the mutation E327P. In some embodiments, the KORV reverse transcriptase variant comprises the mutation L599W.

[0308] In certain embodiments, the KORV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 223-227, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 223-227, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 310F, 327P, and 599W:

KORV-RT D197N:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 223)

KORV-RT T303K:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGKAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQ
EAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTN
ARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPL
PGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALR
LAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKR
VAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 224)

KORV-RT W310F:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLFIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 225)

KORV-RT E327P:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 226)

KORV-RT L599W:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 227)

[0309] In certain embodiments, the KORV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 244, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 244, wherein the amino acid sequence comprises the residues 197N, 303K, 310F, 327P, and 599W:

KORV_penta:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP

GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ ID NO: 244)

[0310] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a WMSV reverse transcriptase of SEQ ID NO: 228, or a WMSV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 228, wherein the WMSV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W311F, E327P, and L599W. In some embodiments, the WMSV reverse transcriptase variant comprises two or more, three or more, four or more, or all five of these mutations. In some embodiments, the WMSV reverse transcriptase variant comprises the mutation D197N. In some embodiments, the WMSV reverse transcriptase variant comprises the mutation T303K. In some embodiments, the WMSV reverse transcriptase variant comprises the mutation W311F. In some embodiments, the WMSV reverse transcriptase variant comprises the mutation E327P. In some embodiments, the WMSV reverse transcriptase variant comprises the mutation L599W.

[0311] In certain embodiments, the WMSV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 229-233, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 229-233, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 311F, 327P, and 599W:

WMSV-RT D197N:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAP TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 229)

WMSV-RT T303K:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGKAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 230)

WMSV-RT W311F:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLFIPGFASLAAPLYPLTKESIPFIWTEEHQKAFDR
IKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPVA
SGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMT
HYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLPG
VPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLA
EGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAII
HCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 231)

WMSV-RT E327P:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 232)

WMSV-RT L599W:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP

TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO:
233)

[0312] In certain embodiments, the WMSV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 245, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 245, wherein the amino acid sequence comprises the residues 197N, 303K, 311F, 327P, and 599W:

WMSV_penta:

LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO:
245)

[0313] In some embodiments, the domain comprising an RNA-dependent DNA polymerase activity comprises a PERV reverse transcriptase. For example, the improved prime editor proteins described herein may comprise a PERV reverse transcriptase comprising one or more mutations relative to the amino acid sequence of SEQ ID NO: 45. In some embodiments, the PERV reverse transcriptase comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, E329P, and L602W relative to the amino acid sequence of SEQ ID NO: 45. In certain embodiments, the PERV reverse transcriptase comprises the mutations D199N, T305K, W312F, E329P, and L602W relative to the amino acid sequence of SEQ ID NO: 45. In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a PERV reverse transcriptase of SEQ ID NO: 45, or a PERV reverse

transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 45, wherein the PERV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, E329P, and L602W. In some embodiments, the PERV reverse transcriptase variant comprises two or more, three or more, four or more, or all five of these mutations. In some embodiments, the PERV reverse transcriptase variant comprises the mutation D199N. In some embodiments, the PERV reverse transcriptase variant comprises the mutation T305K. In some embodiments, the PERV reverse transcriptase variant comprises the mutation W312F. In some embodiments, the PERV reverse transcriptase variant comprises the mutation E329P. In some embodiments, the PERV reverse transcriptase variant comprises the mutation L602W.

[0314] In certain embodiments, the PERV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 214 and 234-238, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 214 and 234-238, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 602W:

PERV variant 21:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 214)

PERV-RT D199N:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK

KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 234)

PERV-RT T305K:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGKAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 235)

PERV-RT W313F:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLFIPGFATLAAPLYPLTKEKGEFSWAPEHQK
AFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNA
RMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALTQ
ALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLP
KRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 236)

PERV-RT E329P:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 237)

PERV-RT L602W:

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 238)

[0315] In certain embodiments, the PERV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 215, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 215, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 602W:

PERV variant 21.6 (pentamutant comprising D199N, T305K, W312F, E329P, and L602W substitutions):

TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQK
AFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNA
RMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALTQ
ALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHL
PKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 215)

[0316] In some embodiments, the domain comprising an RNA-dependent DNA polymerase activity comprises a Tf1 reverse transcriptase. For example, the improved prime editor proteins described herein may comprise a Tf1 reverse transcriptase comprising one or more mutations relative to the amino acid sequence of SEQ ID NO: 55. In some embodiments, the Tf1 reverse transcriptase comprises one or more mutations selected from the group consisting of V14A, E22K, P70T, G72V, M102I, K106R, K118R, A139T, L158Q, F269L, S297Q, K356E, A363V, K413E, I423V, and S492N relative to the amino acid sequence of SEQ ID NO: 55. In certain embodiments, the Tf1 reverse transcriptase comprises any one of the

following groups of amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 55:

K118R and S297Q;

V14A, L158Q, F269L, and K356E;

K106R, L158Q, F269L, A363V, and I423V;

E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, and S492N; or

P70T, G72V, M102I, K106R, L158Q, F269L, A363V, K413E, and S492N.

[0317] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a Tf1 reverse transcriptase of SEQ ID NO: 171, or a Tf1 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 171, wherein the Tf1 reverse transcriptase variant comprises one or more mutations selected from the group consisting of V14A, E22K, I64L, I64W, P70T, G72V, M102I, K106R, K118R, L133N, A139T, L158Q, S188K, I260L, F269L, E274R, R288Q, Q293K, S297Q, N316Q, K321R, K356E, A363V, K413E, I423V, and S492N relative to SEQ ID NO: 171. In some embodiments, the Tf1 reverse transcriptase variant comprises a single mutation, wherein the single mutation is an I64L mutation, an I64W mutation, a K118R mutation, an L133N mutation, an S188K mutation, an I260L mutation, an E274R mutation, an R288Q mutation, a Q293K mutation, an S297Q mutation, an N316Q mutation, or a K321R mutation.

[0318] In some embodiments, the Tf1 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 171: K118R and S297Q; V14A, L158Q, F269L, and K356E; E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, and S492N; P70T, G72V, M102I, K106R, L158Q, F269L, A363V, K413E, and S492N; K106R, L158Q, F269L, A363V, and I423V; K118R, S297Q, S188K, I64L, I260L, and R288Q; E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, S492N, K118R, S297Q, S188K, I64L, and I260L; K118R and S188K; K118R, S188K, and I260L; K118R, S188K, I260L, and S297Q; or K118R, S188K, I260L, R288K, and S297Q.

[0319] In certain embodiments, the Tf1 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 196-213 and 251-255, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 196-213 and 251-255, wherein the amino acid sequence comprises at least one of residues 14A, 22K, 64L, 64W, 70T, 72V, 102I, 106R, 118R, 133N, 139T, 158Q, 188K, 260L, 269L, 274R, 288Q, 293K, 297Q, 316Q, 321R, 356E, 363V, 413E, 423V, 492N:

Tf1 variant 5.131:
ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 196)

Tf1 variant 5.27:

ISSSKHTLSQMNKASNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKEILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 197)

Tf1 variant 5.47:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPRKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTVEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 198)

Tf1 variant 5.59:
ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 199)

Tf1 variant 5.60:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFEPKK
KRKV (SEQ ID NO: 200)

Tf1 variant 5.612:

ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPLRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNRK
ELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 201)

Tf1 variant 5.618:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 202)

Tf1 variant S188K:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRK
ELRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 203)

Tf1 variant I260L:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 204)

Tf1 variant R288Q:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNQKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 205)

Tf1 variant Q293K:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRKFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 206)

Tf1 variant I64L:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPLRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 207)

Tf1 variant I64W:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPWRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVV
DYKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCP
RGVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRK

ELRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 208)

Tf1 variant N316Q:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLQKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 209)

Tf1 variant K321R:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKRDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 210)

Tf1 variant L133N:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPNIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 211)

Tf1 variant K118R:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 212)

Tf1 variant K118R: Tf1 variant S297Q:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 213)

Tf1-rat4:
MISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQE
NYRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVV
DYRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCP
RGVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGFTPCQENIDKVLQWKQPKNQK
ELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 251)

Tf1evo3.1:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 252)

Tf1evo3.2:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNVYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 253)

Tf1evo+rat-1:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNVYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQ

KELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSP
PVLRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 254)

Tf1evo+rat2:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGIKTAPAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQKE
LRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 255)

[0320] In some embodiments, the domain comprising an RNA-dependent DNA polymerase activity comprises an Ec48 reverse transcriptase. For example, the improved prime editor proteins described herein may comprise an Ec48 reverse transcriptase comprising one or more mutations relative to the amino acid sequence of SEQ ID NO: 59. In some embodiments, the Ec48 reverse transcriptase comprises one or more mutations selected from the group consisting of A36V, E54K, K87E, R205K, V214L, D243N, R267I, S277F, E279K, N317S, K318E, H324Q, K326E, E328K, and R372K relative to the amino acid sequence of SEQ ID NO: 59. In certain embodiments, the Ec48 reverse transcriptase comprises any one of the following groups of amino acid substitutions relative to the amino acid sequence of SEQ ID NO: 59:

R267I, K318E, K326E, E328K, and R372K;

K87E, R205K, V214L, D243N, R267I, N317S, K318E, H324Q, and K326E;

E54K, K87E, D243N, R267I, E279K, and K318E;

A36V, K87E, R205K, D243N, R267I, E279K, and K318E;

E54K, K87E, D243N, R267I, E279K, and K318E; or

E54K, K87E, D243N, R267I, S277F, E279K, and K318E.

[0321] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is an Ec48 reverse transcriptase of SEQ ID NO: 59, or an Ec48 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 59, wherein the Ec48 reverse transcriptase variant comprises one or more mutations selected from the group

consisting of A36V, E54K, E60K, K87E, S151T, E165D, L182N, T189N, R205K, V214L, D243N, R267I, S277F, E279K, V303M, K307R, R315K, N317S, K318E, H324Q, K326E, E328K, K343N, R372K, R378K, and T385R relative to SEQ ID NO: 59. In some embodiments, the Ec48 reverse transcriptase variant comprises a single mutation, wherein the single mutation is an L182N mutation, a T189N mutation, a K307R mutation, an R315K mutation, an R378K mutation, or a T385R mutation.

[0322] In some embodiments, the Ec48 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: : R267I, K318E, K326E, E328K, and R372K; K87E, R205K, V214L, D243N, R267I, N317S, K318E, H324Q, and K326E; E54K, K87E, D243N, R267I, E279K, and K318E; A36V, K87E, R205K, D243N, R267I, E279K, and K318E; E54K, K87E, D243N, R267I, E279K, and K318E; E54K, K87E, D243N, R267I, S277F, E279K, and K318E; E60K, K87E, E165D, D243N, R267I, E279K, K318E, and K343N; E60K, K87E, S151T, E165D, D243N, R267I, E279K, V303M, K318E, and K343N; or R315K, L182N, and T189N.

[0323] In certain embodiments, the Ec48 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 188-195, 256, and 257, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 188-195, 256, and 257, wherein the amino acid sequence comprises at least one of residues 36V, 54K, 60K, 87E, 151T, 165D, 182N, 189N, 205K, 214L, 243N, 267I, 277F, 279K, 303M, 307R, 315K, 317S, 318E, 324Q, 326E, 328K, 343N, 372K, 378K, and 385R:

Ec48 variant 3.23:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQKKGLVYTRLLDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDEVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVSELGRVGQEEYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 188)

Ec48 variant 3.35 (or Ec48-evo2):

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA

VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 189)

Ec48 variant 3.36:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKVLSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQKKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 190)

Ec48 variant 3.37:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 191)

Ec48 variant 3.38:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPFDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 192)

Ec48 variant 3.500:
GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALDYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 193)

Ec48 variant 3.501:
GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRTVFEEILHIKDEALDYLVDICTKD
DFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQM
QSHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLK

LLAAKNNTKTSMAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDV
AVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASL
KPL (SEQ ID NO: 194)

Ec48 variant 3.8 (or Ec48-evo1):

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINKRIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHDLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDEVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEEYKSFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKKKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 195)

Ec48-v2:
GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINKRIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGANTSSYIANLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHDLPINKHKTKIFHCSSEPIKVHGLRVDYDSPRLPSDEVKRIRASIHNLK
LLAAKNNTKTSVAYRKEFNRCMGKVNKLGRVGHEKYESFKKQLQAIKPMPSKRDV
AVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASL
KPL (SEQ ID NO: 256)

Ec48-evo3:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALDYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 257)

[0324] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is an Ne144 reverse transcriptase of SEQ ID NO: 239, or an Ne144 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 239, wherein the Ne144 reverse transcriptase variant comprises one or more mutations selected from the group consisting of A157T, A165T, and G288V relative to SEQ ID NO: 239. In some embodiments, the Ne144 reverse transcriptase variant comprises the mutations A157T, A165T, and G288V.

[0325] In certain embodiments, the Ne144 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 240, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 240, wherein the amino acid sequence comprises at least one of residues 157T, 165T, and 288V:

Ne144 RT 38.14:

AGQPTSREALYERIRSTSKEEVILEEMIRLGFWPAQGAVPHDPAEEIRRRGELERQLSE
LREKSRKLYNEKALIAEQRKQRLAESRRKQKETKARRERERQERAQKWAQRKAGEI
LFLGEDVSGGMSHKTCDAELIKREGVPAIASAEELARAMGITLKELRFLTYNRKVSR
VTHYRRFLLPKKTGGLRLISAPMPRLKRAQAWALEHIFNKLSFEPAAHGFVAGRSIVS
NARPHVGADVVVNLDLKDFFPTVSFPRVKGALRHLGYSESVATALALVCTEPEVDE
VVLDGTTWYVARGERFLPQGSPCSPAITNLLCRRLDRRLHGLAQALGFVYTRYADD
LTFSGRGEAAESKRVGKLLRGAADIVAHEGFVVHPDKTRVMRRGRRQEVTGVVVN
DKTSVPRDELRKFRATLYQIEKDGPADKRWGNGGDVLAAVHGYACFVAMVDPSRG
QPLLARARALLAKHGGPSKPPGGSGPRAPTPVQPTANAPEAPKPVAPATPAAPAKKG
WKLF (SEQ ID NO: 240)

[0326] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a Vc95 reverse transcriptase of SEQ ID NO: 241, or a Vc95 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 241, wherein the Vc95 reverse transcriptase variant comprises one or more mutations selected from the group consisting of L11M, S75A, V97M, N146D, and N245T relative to SEQ ID NO: 241. In some embodiments, the Vc95 reverse transcriptase variant comprises the mutations L11M, S75A, V97M, N146D, and N245T.

[0327] In certain embodiments, the Vc95 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 242, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 242, wherein the amino acid sequence comprises at least one of residues 11M, 75A, 97M, 146D, and 245T:

Vc95 RT variant - 25.8:

NILTTLREQLMTNNVIMPQEFERLEVRGSHAYKVYSIPKRKAGRRTIAHPSSKLKICQ
RHLNAILNPLLKVHDASYAYVKGRSIKDNALVHSHSAYMLKMDFQNFFNSITPTILR
QCLIQNDILLSVNELEKLEQLIFWNPSKKRDGKLILSVGSPISPLISNAIMYPFDKIINDI
CTKHGINYTRYADDITFSTNIKNTLNKLPEIVEQLIIQTYAGRIIINKRKTVFSSKKHNR
HVTGITLTTDSKISIGRSRKRYISSLVFKYINKNLDIDEINHMKGMLAFAYNIEPIYIHR
LSHKYKVNIVEKILRGSN (SEQ ID NO: 242)

[0328] In some embodiments, the present disclosure provides reverse transcriptases, and prime editors (*e.g.* fusion proteins or prime editors in which each component is provided *in trans*) comprising reverse transcriptases, wherein the reverse transcriptase is a Gs reverse transcriptase of SEQ ID NO: 60, or a Gs reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 60, wherein the Gs reverse transcriptase variant comprises one or more mutations selected from the group consisting of N12D, A16E, A16V, L17P, V20G, L37R, L37P, R38H, Y40C, I41N, I41S, W45R, I67T, I67R, G72E, G73V, G78V, Q93R, A123V, Y126F, E129G, K162N, P190L, D206V, R233K, A234V, R263G, P264S, R267M, K279E, R287I, R291K, P309T, R344S, R358S, R360S, E363G, V374A, and Q412H relative to SEQ ID NO: 60. In some embodiments, the Gs reverse transcriptase variant comprises two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, or ten or more of these mutations.

[0329] In some embodiments, the Gs reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 60: L17P and D206V; N12D, L37R, and G78V; A16E, L37P, and A123V; A16V, R38H, W45R, Y126F, and Q412H; A16V, R38H, W45R, and R291K; N12D, L37R, G72E, E129G, P264S, R344S, and R360S; N12D, Y40C, I67T, G73V, Q93R, R287I, and R358S; N12D, Y40C, I67T, G73V, Q93R, and R358S; N12D, I41N, P190L, A234V, and K279E; N12D, L37R, R267M, P309T, R358S, and E363G; A16V, V20G, I41S, R233K, and P264S; L17P, V20G, I41S, I67R, R263G, P264S, and V374A; or L17P, V20G, I41S, I67R, K162N, R263G, and P264S.

[0330] In certain embodiments, the Gs reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 159-171, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 159-171, wherein the amino acid sequence comprises at least one of residues 12D, 16E, 16V, 17P, 20G, 37R, 37P, 38H, 40C, 41N, 41S, 45R, 67T, 67R, 72E, 73V, 78V, 93R, 123V, 126F, 129G, 162N, 190L, 206V, 233K, 234V, 263G, 264S, 267M, 279E, 287I, 291K, 309T, 344S, 358S, 360S, 363G, 374A, and 412H:

Gs variants comprising: L17P + D206V

EANQGAPGIDGVSTDQLRDYIRAHWSTIHAQLLAGTYRPAPVRRVEIPKPGGGTRQL
GIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGYRYVV
DMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQAGVMIEGVKVQTEEGTP
QGGPLSPLLANILLD**V**LDKELEKRGLKFCRYADDCNIYVKSLRAGQRVKQSIQRFLE
KTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLTNPNWS
ISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRLRLCQWLQWKRVRTRIR
ELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTYWTAQGLKSLTQRYFEL
RQG (SEQ ID NO: 159)

Gs variant N12D + L37R + G78V

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQL**V**IPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 160)

Gs A16E + L37P + A123V

ALLERILARDNLIT**E**LKRVEANQGAPGIDGVSTDQ**P**RDYIRAHWSTIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQ**V**QGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 161)

Gs variant A16V + R38H + W45R + Y126F + Q412H

ALLERILARDNLIT**V**LKRVEANQGAPGIDGVSTDQL**H**DYIRAH**R**STIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQAQG**F**IQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLT**H**RYFELRQG (SEQ ID NO: 162)

Gs A16V + R38H + W45R + R291K

ALLERILARDNLIT**V**LKRVEANQGAPGIDGVSTDQL**H**DYIRAH**R**STIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
Q**K**LKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL

RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 163)

Gs variant 814 N12D + L37R + G72E + E129G + P264S + R344S + R360S

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPG**E**GTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQ**G**GYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDR**S**WKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIR**S**RL
RLCQWLQWKRVRT**S**IRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 164)

Gs variant 815 N12D + Y40C + I67T + G73V + Q93R + R287I + R358S

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRD**C**IRAHWSTIHAQLLAGTYRP
APVRRVE**T**PKPGG**V**TRQLGIPTVVDRLIQQAIL**R**ELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAP**I**SI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRV**S**TRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 165)

Gs variant 816 N12D + Y40C + I67T + G73V + Q93R + R358S

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRD**C**IRAHWSTIHAQLLAGTYRP
APVRRVE**T**PKPGG**V**TRQLGIPTVVDRLIQQAIL**R**ELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRV**S**TRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 166)

Gs variant 817 N12D + I41N + P190L + A234V + K279E

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRDY**N**RAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGT**L**QGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLR**V**GQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPER**E**ARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 167)

Gs variant 818 N12D + L37R + R267M + P309T + R358S + E363G

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ

AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWK**M**AFLGFSFTPERKARIRLAPR
SIQRLKQRIRQLTNPNWSISM**T**ERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRR
LRLCQWLQWKRV**S**TRIR**G**LRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKT
YWTAQGLKSLTQRYFELRQG (SEQ ID NO: 168)

Gs variant 819 A16V + V20G + I41S + R233K + P264S

ALLERILARDNLIT**V**LKR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SL**K**AGQRVKQSIQRFLEKTLKLKVNEEKSAVDR**S**WKRAFLGFSFTPERKARIRLAPR
SIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRR
LRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKT
YWTAQGLKSLTQRYFELRQG (SEQ ID NO: 169)

Gs variant 820 L17P + V20G + I41S + I67R + R263G + P264S + V374A

ALLERILARDNLITA**P**KR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVE**R**PKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAH
DAVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYL
QAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYV
KSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVD**GS**WKRAFLGFSFTPERKARIRLAP
RSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRR
RLRLCQWLQWKRVRTRIRELRALGLKETA**A**MEIANTRKGAWRTTKTPQLHQALGK
TYWTAQGLKSLTQRYFELRQG (SEQ ID NO: 170)

Gs variant 821 L17P + V20G + I41S + I67R + K162N + R263G + P264S

ALLERILARDNLITA**P**KR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVE**R**PKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAH
DAVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDK**N**VLKLIRAYL
QAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYV
KSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVD**GS**WKRAFLGFSFTPERKARIRLAP
RSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRR
RLRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGK
TYWTAQGLKSLTQRYFELRQG (SEQ ID NO: 171)

[0331] As illustrated in FIG. 27A, this disclosure in part provides engineered and PACE^2-evolved RT variants for prime editing. Thus far, the only RT enzyme that has been utilized for prime editing in mammalian cells is M-MLV RT. M-MLV RT is a large enzyme (2.2 kB), which poses barriers for many *in vivo* delivery methods such as Adeno-associated Viruses (AAVs). Since RT enzymes vary widely in their size and enzymatic activity, the alternate enzymes disclosed here provide unique advantages for prime editing (e.g., smaller size or improved editing). These improvements lead to prime editors that are more efficient and more easily delivered for therapeutic applications.

[0332] In various embodiments, the modified prime editor proteins, including PEmax, comprise a reverse transcriptase domain. In some embodiments, the reverse transcriptase domain is a variant of wild type MMLV reverse transcriptase having the amino acid sequence of SEQ ID NO: 34.

[0333] For example, PEmax of SEQ ID NO: 2 comprises a variant reverse transcriptase domain of SEQ ID NO: 34, which is based on the wild type MMLV reverse transcriptase domain of SEQ ID NO: 33 (and, in particular, a Genscript codon optimized MMLV reverse transcriptase having the nucleotide sequence of SEQ ID NO: 33) and which comprises amino acid substitutions D200N T306K W313F T330P L603W relative to the wild type MMLV RT of SEQ ID NO: 34. The amino acid sequence of the variant RT of PEmax is SEQ ID NO: 34.

[0334] The modified prime editors may also comprise other variant RTs as well. In various embodiments, the modified prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising one or more of the following mutations: P51L, S67K, E69K, L139P, T197A, D200N, H204R, F209N, E302K, E302R, T306K, F309N, W313F, T330P, L345G, L435G, N454K, D524G, E562Q, D583N, H594Q, L603W, E607K, or D653N in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence.

[0335] Some exemplary reverse transcriptases that can be fused to napDNAbp proteins or provided as individual proteins according to various embodiments of this disclosure are provided below. Exemplary reverse transcriptases include variants with at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity to the following wild-type enzymes or partial enzymes:

Description	Sequence (variant substitutions relative to wild type)
Reverse transcriptase (M-MLV RT) wild type moloney murine leukemia virus	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 33)

Description	Sequence (variant substitutions relative to wild type)
Used in PE1 (prime editor 1 fusion protein disclosed herein)	
M-MLV RT D200N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 63)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL IENSSP (SEQ ID NO: 64)
M-MLV RT D200N T330P L603W	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP

Description	Sequence (variant substitutions relative to wild type)
	LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 65)
M-MLV RT D200N T330P L603W E69K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQKARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA LLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT STLLIENSSP (SEQ ID NO: 66)
M-MLV RT D200N T330P L603W E302R	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLRRFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP(SEQ ID NO: 67)
M-MLV RT D200N T330P L603W E607K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA

Description	Sequence (variant substitutions relative to wild type)
	EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 68)
M-MLV RT D200N T330P L603W L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 69)
M-MLV RT D200N T330P L603W L435G	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVI**G**APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 70)
M-MLV RT D200N T330P L603W N454K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 71)

Description	Sequence (variant substitutions relative to wild type)
M-MLV RT D200N T330P L603W T306K	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 72)
M-MLV RT D200N T330P L603W W313F	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWGP DQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 73)
M-MLV RT D200N T330P L603W D524G E562Q D583N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT**G**GS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRA**Q**LIALTQALKMA EGKKLNVYT**N**SRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 74)
M-MLV RT D200N T330P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY

Description	Sequence (variant substitutions relative to wild type)
L603W E302R W313F	TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLR**R**FLGTAGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 75)
M-MLV RT D200N T330P L603W E607K L139P	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG**P**PPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTS**K**GKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 76)
M-MLV RT P51L S67K T197A H204R E302K F309N W313F T330P L435G N454K D524G D583N H594Q D653N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LFDEAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 77)
M-MLV RT D200N P51L S67K T197A H204R E302K F309N	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LII**L**LKATSTPVSIKQYPM**K**QEARLGIKPHIQRLLDQGILVPCQSPWNTP LLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQW YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSP**A**LF**N**EAL**R**RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET

Description	Sequence (variant substitutions relative to wild type)
W313F T330P L345G N454K D524G D583N H594Q D653N	VMGQPTPKTPRQLR**K**FLGTAG**N**CRL**F**IPGFAEMAAPLYPLTK**P**GTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVI**G**APHAVEALVKQPPDRWLS**K**ARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYT **G**GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALK MAEGKKLNVYT**N**SRYAFATAHI**Q**GEIYRRRGLLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMA**N**QAARKAAITETPDT STLLIENSSP (SEQ ID NO: 78)
M-MLV RT D200N T330P L603W T306K W313F in PE2 and PEmax	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN SPTLF**N**EALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRAL LQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLG**K**AGFCRL**F**IPGFAEMAAPLYPLTK**P**GTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKL GPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA EGKKLNVYTDSRYAFATAHIHGEIYRRRG**W**LTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTST LLIENSSP (SEQ ID NO: 34)

[0336] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising one or more of the following mutations: P51X, S67X, E69X, L139X, T197X, D200X, H204X, F209X, E302X, T306X, F309X, W313X, T330X, L345X, L435X, N454X, D524X, E562X, D583X, H594X, L603X, E607X, or D653X in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid.

[0337] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a P51X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is L.

[0338] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a S67X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in

another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0339] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a E69X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0340] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a L139X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is P.

[0341] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a T197X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is A.

[0342] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a D200X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is N.

[0343] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a H204X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is R.

[0344] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a F209X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is N.

[0345] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a E302X mutation

in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0346] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a E302X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is R.

[0347] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a T306X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0348] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a F309X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is N.

[0349] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a W313X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is F.

[0350] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a T330X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is P.

[0351] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a L345X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is G.

[0352] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a L435X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is G.

[0353] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a N454X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0354] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a D524X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is G.

[0355] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a E562X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is Q.

[0356] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a D583X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is N.

[0357] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a H594X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is Q.

[0358] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a L603X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in

another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is W.

[0359] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a E607X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is K.

[0360] In various other embodiments, the prime editors described herein (with RT provided as either a fusion partner or *in trans*) can include a variant RT comprising a D653X mutation in the wild type M-MLV RT of SEQ ID NO: 33 or at a corresponding amino acid position in another wild type RT polypeptide sequence, wherein "X" can be any amino acid. In certain embodiments, X is N.

[0361] Some exemplary reverse transcriptases that can be fused to napDNAbp proteins or provided as individual proteins according to various embodiments of this disclosure are provided below. Exemplary reverse transcriptases include variants with at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity to the wild-type enzymes or partial enzymes described in SEQ ID NOs: 33-34 and 63-78.

[0362] The prime editor (PE) system described here contemplates any publicly-available reverse transcriptase described or disclosed in any of the following U.S. patents (each of which are incorporated by reference in their entireties): U.S. Patent Nos: 10,202,658; 10,189,831; 10,150,955; 9,932,567; 9,783,791; 9,580,698; 9,534,201; and 9,458,484, and any variant thereof that can be made using known methods for installing mutations, or known methods for evolving proteins. The following references describe reverse transcriptases in art. Each of their disclosures are incorporated herein by reference in their entireties.

[0363] Herzig, E., Voronin, N., Kucherenko, N. & Hizi, A. A Novel Leu92 Mutant of HIV-1 Reverse Transcriptase with a Selective Deficiency in Strand Transfer Causes a Loss of Viral Replication. *J. Virol.* **89**, 8119–8129 (2015).

[0364] Mohr, G. *et al.* A Reverse Transcriptase-Cas1 Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition. *Mol. Cell* **72**, 700-714.e8 (2018).

[0365] Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183–195 (2018).

[0366] Zimmerly, S. & Wu, L. An Unexplored Diversity of Reverse Transcriptases in Bacteria. *Microbiol Spectr* **3**, MDNA3-0058–2014 (2015).

[0367] Ostertag, E. M. & Kazazian Jr, H. H. Biology of Mammalian L1 Retrotransposons. *Annual Review of Genetics* **35**, 501–538 (2001).

[0368] Perach, M. & Hizi, A. Catalytic Features of the Recombinant Reverse Transcriptase of Bovine Leukemia Virus Expressed in Bacteria. *Virology* **259**, 176–189 (1999).

[0369] Lim, D. *et al.* Crystal structure of the moloney murine leukemia virus RNase H domain. *J. Virol.* **80**, 8379–8389 (2006).

[0370] Zhao, C. & Pyle, A. M. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nature Structural & Molecular Biology* **23**, 558–565 (2016).

[0371] Griffiths, D. J. Endogenous retroviruses in the human genome sequence. *Genome Biol.* **2**, REVIEWS1017 (2001).

[0372] Baranauskas, A. *et al.* Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng Des Sel* **25**, 657–668 (2012).

[0373] Zimmerly, S., Guo, H., Perlman, P. S. & Lambowltz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554 (1995).

[0374] Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).

[0375] Berkhout, B., Jebbink, M. & Zsíros, J. Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. *Journal of Virology* **73**, 2365–2375 (1999).

[0376] Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M. & Gerard, G. F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res* **16**, 265–277 (1988).

[0377] Arezi, B. & Hogrefe, H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res* **37**, 473–481 (2009).

[0378] Blain, S. W. & Goff, S. P. Nuclease activities of Moloney murine leukemia virus reverse transcriptase. Mutants with altered substrate specificities. *J. Biol. Chem.* **268**, 23585–23592 (1993).

[0379] Xiong, Y. & Eickbush, T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* **9**, 3353–3362 (1990).

[0380] Herschhorn, A. & Hizi, A. Retroviral reverse transcriptases. *Cell. Mol. Life Sci.* **67**, 2717–2747 (2010).

[0381] Taube, R., Loya, S., Avidan, O., Perach, M. & Hizi, A. Reverse transcriptase of mouse mammary tumour virus: expression in bacteria, purification and biochemical characterization. *Biochem. J.* **329 (Pt 3)**, 579–587 (1998).

[0382] Liu, M. *et al.* Reverse Transcriptase-Mediated Tropism Switching in Bordetella Bacteriophage. *Science* **295**, 2091–2094 (2002).

[0383] Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).

[0384] Nottingham, R. M. *et al.* RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* **22**, 597–613 (2016).

[0385] Telesnitsky, A. & Goff, S. P. RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer-template. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1276–1280 (1993).

[0386] Halvas, E. K., Svarovskaia, E. S. & Pathak, V. K. Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity. *Journal of Virology* **74**, 10349–10358 (2000).

[0387] Nowak, E. *et al.* Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res* **41**, 3874–3887 (2013).

[0388] Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Primer and Its Functional and Evolutionary Implications. *Molecular Cell* **68**, 926-939.e4 (2017).

[0389] Das, D. & Georgiadis, M. M. The Crystal Structure of the Monomeric Reverse Transcriptase from Moloney Murine Leukemia Virus. *Structure* **12**, 819–829 (2004).

[0390] Avidan, O., Meer, M. E., Oz, I. & Hizi, A. The processivity and fidelity of DNA synthesis exhibited by the reverse transcriptase of bovine leukemia virus. *European Journal of Biochemistry* **269**, 859–867 (2002).

[0391] Gerard, G. F. *et al.* The role of template-primer in protection of reverse transcriptase from thermal inactivation. *Nucleic Acids Res* **30**, 3118–3129 (2002).

[0392] Monot, C. *et al.* The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts. *PLOS Genetics* **9**, e1003499 (2013).

[0393] Mohr, S. *et al.* Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).

[0394] Any of the references noted above which relate to reverse transcriptases are hereby incorporated by reference in their entireties, if not already stated so.

Additional domains

A. Linkers

[0395] The modified PE fusion proteins described herein may include one or more linkers.

[0396] As defined above, the term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, e.g., a binding domain and a cleavage domain of a nuclease. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease and the catalytic domain of a polymerase (e.g., a reverse transcriptase). In some embodiments, a linker joins a dCas9 and reverse transcriptase. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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. Longer or shorter linkers are also contemplated.

[0397] 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.

[0398] In some other embodiments, the linker comprises the amino acid sequence $(GGGGS)_n$ (SEQ ID NO: 84), $(G)_n$ (SEQ ID NO: 85), $(EAAAK)_n$ (SEQ ID NO: 86), $(GGS)_n$ (SEQ ID NO: 87), $(SGGS)_n$ (SEQ ID NO: 81), $(XP)_n$ (SEQ ID NO: 88), or any combination thereof, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, the linker comprises the amino acid sequence (GGS)n (SEQ ID NO: 87), wherein n is 1, 3, or 7. In some embodiments, the linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 89). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 90). In some embodiments, the linker comprises the amino acid sequence SGGSGGSGGS (SEQ ID NO: 91). In some embodiments, the linker comprises the amino acid sequence SGGS (SEQ ID NO: 81). In other embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSS GGS (SEQ ID NO: 83, 60AA).

[0399] In certain embodiments, linkers may be used to link any of the peptides or peptide domains or moieties of the invention (e.g., a napDNAbp linked or fused to a reverse transcriptase).

[0400] As defined above, the term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, e.g., a binding domain and a cleavage domain of a nuclease. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease and the catalytic domain of a recombinase. In some embodiments, a linker joins a dCas9 and reverse transcriptase. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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. Longer or shorter linkers are also contemplated.

[0401] 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 amino*HEXA*noic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (e.g., cyclopentane, cyclo*HEXA*ne). 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.

[0402] In some other embodiments, the linker comprises the amino acid sequence (GGGGS)n (SEQ ID NO: 84), (G)n (SEQ ID NO: 85), $(EAAAK)_n$ (SEQ ID NO: 86), $(GGS)_n$ (SEQ ID NO: 87), (SGGS)n (SEQ ID NO: 81), (XP)n (SEQ ID NO: 88), or any combination thereof, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, the linker comprises the amino acid sequence (GGS)n (SEQ ID NO: 87), wherein n is 1, 3, or 7. In some embodiments, the linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 89). In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 90). In some embodiments, the linker comprises the amino acid sequence SGGSGGSGGS (SEQ ID NO: 91). In some embodiments, the linker comprises the amino acid sequence SGGS (SEQ ID NO: 81).

[0403] In particular, the following linkers can be used in various embodiments to join prime editor domains with one another:

GGS (SEQ ID NO: 87);

GGSGGS (SEQ ID NO: 92);

GGSGGSGGS (SEQ ID NO: 93);

SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 80);

SGSETPGTSESATPES (SEQ ID NO: 89);

SGGSSGGSSGSETPGTSESATPESAGSYPYDVPDYAGSAAPAAKKKKLDGSGSGGSS GG S (SEQ ID NO: 83).

[0404] The PE fusion proteins may also comprise various other domains besides the napDNAbp (e.g., Cas9 domain) and the polymerase domain (e.g., RT domain). For example, in the case where the napDNAbp is a Cas9 and the polymerase is a RT, the PE fusion proteins may comprise one or more linkers that join the Cas9 domain with the RT domain. The linkers may also join other functional domains, such as nuclear localization sequences (NLS) or a FEN1 (or other flap endonuclease) to the PE fusion proteins or a domain thereof.

B. Nuclear localization sequence (NLS)

[0405] In various embodiments, the modified PE fusion proteins may comprise one or more nuclear localization sequences (NLS), which help promote translocation of a protein into the cell nucleus. Such sequences are well-known in the art and can include the following examples:

DESCRIPTION	SEQUENCE	SEQ ID NO:
NLS OF SV40 LARGE T-AG	PKKKRKV	94
NLS	MKRTADGSEFESPKKKRKV	95
NLS	MDSLLMNRRKFLYQFKNVRWAKG RRETYLC	99
NLS OF NUCLEOPLASMIN	AVKRPAATKKAGQAKKKKLD	100
NLS OF EGL-13	MSRRRKANPTKLSENAKKLAKEV EN	101
NLS OF C-MYC	PAAKRVKLD	98
NLS OF TUS-PROTEIN	KLKIKRPVK	102
NLS OF POLYOMA LARGE T-AG	VSRKRPRP	103
NLS OF HEPATITIS D VIRUS ANTIGEN	EGAPPAKRAR	104
NLS OF MURINE P53	PPQPKKKPLDGE	105
NLS OF PE1 AND PE2	SGGSKRTADGSEFEPKKKRKV	96
BIPARTITE SV40 NLS	KRTADGSEFESPKKKRKV	97

[0406] The NLS examples above are non-limiting. The modified PE fusion proteins may comprise any known NLS sequence, including any of those described in Cokol *et al.*, "Finding nuclear localization signals," *EMBO Rep.,* 2000, 1(5): 411-415 and Freitas *et al.*, "Mechanisms and Signals for the Nuclear Import of Proteins," *Current Genomics*, 2009, 10(8): 550-7, each of which are incorporated herein by reference.

[0407] In various embodiments, the prime editors and constructs encoding the prime editors utilized in the methods and compositions disclosed herein further comprise one or more, preferably, at least two nuclear localization signals. In certain embodiments, the prime editors comprise at least two NLSs. In embodiments with at least two NLSs, the NLSs can be the same NLSs or they can be different NLSs. In addition, the NLSs may be expressed as part of a fusion protein with the remaining portions of the prime editors. In some embodiments, one or more of the NLSs are bipartite NLSs ("bpNLS"). In certain embodiments, the disclosed fusion proteins comprise two bipartite NLSs. In some embodiments, the disclosed fusion proteins comprise more than two bipartite NLSs.

[0408] The location of the NLS fusion can be at the N-terminus, the C-terminus, or within a sequence of a prime editor (e.g., inserted between the encoded napDNAbp component (e.g., Cas9) and a polymerase domain (e.g., a reverse transcriptase domain).

[0409] The NLSs may be any known NLS sequence in the art. The NLSs may also be any future-discovered NLSs for nuclear localization. The NLSs also may be any naturally-occurring NLS, or any non-naturally occurring NLS (e.g., an NLS with one or more desired mutations).

The term "nuclear localization sequence" or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus, for example, by nuclear transport. 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.*, 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. In some embodiments, an NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 94), MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 99), KRTADGSEFESPKKKRKV (SEQ ID NO: 97), or KRTADGSEFEPKKKRKV (SEQ ID NO: 106). In other embodiments, NLS comprises the amino acid sequences NLSKRPAAIKKAGQAKKKK (SEQ ID NO: 107), PAAKRVKLD (SEQ ID NO: 98), RQRRNELKRSF (SEQ ID NO: 108), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 109).

[0410] In one aspect of the disclosure, a prime editor may be modified with one or more nuclear localization signals (NLS), preferably at least two NLSs. In certain embodiments, the prime editors are modified with two or more NLSs. The disclosure contemplates the use of any nuclear localization signal known in the art at the time of the disclosure, or any nuclear localization signal that is identified or otherwise made available in the state of the art after the time of the instant filing. A representative nuclear localization signal is a peptide sequence that directs the protein to the nucleus of the cell in which the sequence is expressed. A nuclear localization signal is predominantly basic, can be positioned almost anywhere in a protein's amino acid sequence, generally comprises a short sequence of four amino acids (Autieri & Agrawal, (1998) *J. Biol. Chem.* 273: 14731-37, incorporated herein by reference) to eight amino acids, and is typically rich in lysine and arginine residues (Magin *et al.*, (2000) *Virology* 274: 11-16, incorporated herein by reference). Nuclear localization signals often comprise proline residues. A variety of nuclear localization signals have been identified and have been used to effect transport of biological molecules from the cytoplasm to the nucleus of a cell. See, e.g., Tinland *et al.*, (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:7442-46; Moede *et al.*, (1999) *FEBS Lett.* 461:229-34, which is incorporated by reference. Translocation is currently thought to involve nuclear pore proteins.

[0411] Most NLSs can be classified in three general groups: (i) a monopartite NLS exemplified by the SV40 large T antigen NLS (PKKKRKV (SEQ ID NO: 94)); (ii) a bipartite motif consisting of two basic domains separated by a variable number of spacer amino acids and exemplified by the *Xenopus nucleoplasmin* NLS (KRXXXXXXXXXXKKKL (SEQ ID NO: 110)); and (iii) noncanonical sequences such as M9 of the hnRNP Al protein, the influenza virus nucleoprotein NLS, and the yeast Gal4 protein NLS (Dingwall and Laskey 1991).

Nuclear localization signals appear at various points in the amino acid sequences of proteins. NLS's have been identified at the N-terminus, the C-terminus and in the central region of proteins. Thus, the disclosure provides prime editors that may be modified with one or more NLSs at the C-terminus, the N-terminus, as well as at in internal region of the prime editor. The residues of a longer sequence that do not function as component NLS residues should be selected so as not to interfere, for example tonically or sterically, with the nuclear localization signal itself. Therefore, although there are no strict limits on the composition of an NLS-comprising sequence, in practice, such a sequence can be functionally limited in length and composition.

[0412] The present disclosure contemplates any suitable means by which to modify a prime editor to include one or more NLSs. In one aspect, the prime editors may be engineered to express a prime editor protein that is translationally fused at its N-terminus or its C-terminus (or both) to one or more NLSs, i.e., to form a prime editor-NLS fusion construct. In other embodiments, the prime editor-encoding nucleotide sequence may be genetically modified to incorporate a reading frame that encodes one or more NLSs in an internal region of the encoded prime editor. In addition, the NLSs may include various amino acid linkers or spacer regions encoded between the prime editor and the N-terminally, C-terminally, or internally-attached NLS amino acid sequence, e.g, and in the central region of proteins. Thus, the present disclosure also provides for nucleotide constructs, vectors, and host cells for expressing fusion proteins that comprise a prime editor and one or more NLSs.

[0413] The prime editors utilized in the methods and compositions described herein may also comprise nuclear localization signals which are linked to a prime editor through one or more linkers, e.g., and polymeric, amino acid, nucleic acid, polysaccharide, chemical, or nucleic acid linker element. The linkers within the contemplated scope of the disclosure are not intended to have any limitations and can be any suitable type of molecule (e.g., polymer, amino acid, polysaccharide, nucleic acid, lipid, or any synthetic chemical linker domain) and be joined to the prime editor by any suitable strategy that effectuates forming a bond (e.g., covalent linkage, hydrogen bonding) between the prime editor and the one or more NLSs.

C. Flap endonucleases (e.g., FEN1)

[0414] In various embodiments, the PE fusion proteins may comprise one or more flap endonucleases (e.g., FEN1), which refers to an enzyme that catalyzes the removal of 5′ single strand DNA flaps (provided in trans or fused to the PE fusion proteins). These are naturally occurring enzymes that process the removal of 5′ flaps formed during cellular processes, including DNA replication. The prime editing utilized in the methods and compositions described herein may utilize endogenously supplied flap endonucleases or those provided *in trans* to remove the 5′ flap of endogenous DNA formed at the target site during prime editing. Flap endonucleases are known in the art and can be found described in Patel *et al.*, "Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends," *Nucleic Acids Research*, 2012, 40(10): 4507-4519 and Tsutakawa *et al.*, "Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily," *Cell*, 2011, 145(2): 198-211 (each of which are incorporated herein by reference). An exemplary flap endonuclease is FEN1, which can be represented by the following amino acid sequence:

Description	Sequence	SEQ ID NO:
FEN1 Wild type (wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 112

[0415] The flap endonucleases may also include any FEN1 variant, mutant, or other flap endonuclease ortholog, homolog, or variant. Non-limiting FEN1 variant examples are as follows:

Description	Sequence	SEQ ID NO:
FEN1 K168R (relative to FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALV**R**AGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD DFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKR GK	SEQ ID NO: 113
FEN1 S187A (relative to FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFG **A**PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFV DLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLD PNKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEP NEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRL DDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFK RGK	SEQ ID NO: 114
FEN1 K354R (relative to FEN1 wt)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSI YQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMME NGIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQ AQAAGAEQEVEKFTKRLVKVTKQHNDECKHLLSLMGI PYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGS PVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVD LCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESVELKWSEPN EEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLD	SEQ ID NO: 115

	DFFKVTGSLSSA**R**RKEPEPKGSTKKKAKTGAAGKFKR GK	
GEN1	MGVNDLWQILEPVKQHIPLRNLGGKTIAVDLSLWVCE AQTVKKMMGSVMKPHLRNLFFRISYLTQMDVKLVFV MEGEPPKLKADVISKRNQSRYGSSGKSWSQKTGRSHF KSVLRECLHMLECLGIPWVQAAGEAEAMCAYLNAGG HVDGCLTNDGDTFLYGAQTVYRNFTMNTKDPHVDCY TMSSIKSKLGLDRDALVGLAILLGCDYLPKGVPGVGKE QALKLIQILKGQSLLQRFNRWNETSCNSSPQLLVTKKL AHCSVCSHPGSPKDHERNGCRLCKSDKYCEPHDYEYC CPCEWHRTEHDRQLSEVENNIKKKACCCEGFPFHEVIQ EFLLNKDKLVKVIRYQRPDLLLFQRFTLEKMEWPNHY ACEKLLVLLTHYDMIERKLGSRNSNQLQPIRIVKTRIRN GVHCFEIEWEKPEHYAMEDKQHGEFALLTIEEESLFEA AYPEIVAVYQKQKLEIKGKKQKRIKPKENNLPEPDEVM SFQSHMTLKPTCEIFHKQNSKLNSGISPDPTLPQESISAS LNSLLLPKNTPCLNAQEQFMSSLRPLAIQQIKAVSKSLI SESSQPNTSSHNISVIADLHLSTIDWEGTSFSNSPAIQRN TFSHDLKSEVESELSAIPDGFENIPEQLSCESERYTANIK KVLDEDSDGISPEEHLLSGITDLCLQDLPLKERIFTKLSY PQDNLQPDVNLKTLSILSVKESCIANSGSDCTSHLSKDL PGIPLQNESRDSKILKGDQLLQEDYKVNTSVPYSVSNT VVKTCNVRPPNTALDHSRKVDMQTTRKILMKKSVCLD RHSSDEQSAPVFGKAKYTTQRMKHSSQKHNSSHFKES GHNKLSSPKIHIKETEQCVRSYETAENEESCFPDSTKSS LSSLQCHKKENNSGTCLDSPLPLRQRLKLRFQST	SEQ ID NO: 116
ERCC5	MGVQGLWKLLECSGRQVSPEALEGKILAVDISIWLNQ ALKGVRDRHGNSIENPHLLTLFHRLCKLLFFRIRPIFVF DGDAPLLKKQTLVKRRQRKDLASSDSRKTTEKLLKTF LKRQAIKTAFRSKRDEALPSLTQVRRENDLYVLPPLQE EEKHSSEEEDEKEWQERMNQKQALQEEFFHNPQAIDIE SEDFSSLPPEVKHEILTDMKEFTKRRRTLFEAMPEESDD FSQYQLKGLLKKNYLNQHIEHVQKEMNQQHSGHIRRQ YEDEGGFLKEVESRRVVSEDTSHYILIKGIQAKTVAEV DSESLPSSSKMHGMSFDVKSSPCEKLKTEKEPDATPPSP RTLLAMQAALLGSSSEEELESENRRQARGRNAPAAVD EGSISPRTLSAIKRALDDDEDVKVCAGDDVQTGGPGAE EMRINSSTENSDEGLKVRDGKGIPFTATLASSSVNSAEE HVASTNEGREPTDSVPKEQMSLVHVGTEAFPISDESMI KDRKDRLPLESAVVRHSDAPGLPNGRELTPASPTCTNS VSKNETHAEVLEQQNELCPYESKFDSSLLSSDDETKCK PNSASEVIGPVSLQETSSIVSVPSEAVDNVENVVSFNAK EHENFLETIQEQQTTESAGQDLISIPKAVEPMEIDSEESE SDGSFIEVQSVISDEELQAEFPETSKPPSEQGEEELVGTR EGEAPAESESLLRDNSERDDVDGEPQEAEKDAEDSLHE WQDINLEELETLESNLLAQQNSLKAQKQQQERIAATVT GQMFLESQELLRLFGIPYIQAPMEAEAQCAILDLTDQTS GTITDDSDIWLFGARHVYRNFFNKNKFVEYYQYVDFH NQLGLDRNKLINLAYLLGSDYTEGIPTVGCVTAMEILN	SEQ ID NO: 117

	EFPGHGLEPLLKFSEWWHEAQKNPKIRPNPHDTKVKK KLRTLQLTPGFPNPAVAEAYLKPVVDDSKGSFLWGKP DLDKIREFCQRYFGWNRTKTDESLFPVLKQLDAQQTQ LRIDSFFRLAQQEKEDAKRIKSQRLNRAVTCMLRKEKE AAASEIEAVSVAMEKEFELLDKAKRKTQKRGITNTLEE SSSLKRKRLSDSKRKNTCGGFLGETCLSESSDGSSSEDA ESSSLMNVQRRTAAKEPKTSASDSQNSVKEAPVKNGG ATTSSSSDSDDDGGKEKMVLVTARSVFGKKRRKLRRA RGRKRKT	

[0416] In various embodiments, the prime editors contemplated herein may include any flap endonuclease variant of the above-disclosed sequences having an amino acid sequence that is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to any of the above sequences.

Other endonucleases that may be utilized by the instant methods to facilitate removal of the 5´ end single strand DNA flap include, but are not limited to (1) trex 2, (2) exo1 endonuclease (e.g., Keijzers *et al.*, *Biosci Rep.* 2015, 35(3): e00206)

Trex 2

[0417] 3´ three prime repair exonuclease 2 (TREX2) - human

Accession No. NM_080701

MSEAPRAETFVFLDLEATGLPSVEPEIAELSLFAVHRSSLENPEHDESGALVLPRVLD
KLTLCMCPERPFTAKASEITGLSSEGLARCRKAGFDGAVVRTLQAFLSRQAGPICLVA
HNGFDYDFPLLCAELRRLGARLPRDTVCLDTLPALRGLDRAHSHGTRARGRQGYSL
GSLFHRYFRAEPSAAHSAEGDVHTLLLIFLHRAAELLAWADEQARGWAHIEPMYLPP
DDPSLEA (SEQ ID NO: 118).

[0418] 3´ three prime repair exonuclease 2 (TREX2) - mouse

Accession No. NM_011907

MSEPPRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRVLDK
LTLCMCPERPFTAKASEITGLSSESLMHCGKAGFNGAVVRTLQGFLSRQEGPICLVAH
NGFDYDFPLLCTELQRLGAHLPQDTVCLDTLPALRGLDRAHSHGTRAQGRKSYSLA
SLFHRYFQAEPSAAHSAEGDVHTLLLIFLHRAPELLAWADEQARSWAHIEPMYVPPD
GPSLEA (SEQ ID NO: 119).

[0419] 3´ three prime repair exonuclease 2 (TREX2) - rat

Accession No. NM_001107580

MSEPLRAETFVFLDLEATGLPNMDPEIAEISLFAVHRSSLENPERDDSGSLVLPRVLD
KLTLCMCPERPFTAKASEITGLSSEGLMNCRKAAFNDAVVRTLQGFLSRQEGPICLV

AHNGFDYDFPLLCTELQRLGAHLPRDTVCLDTLPALRGLDRVHSHGTRAQGRKSYS LASLFHRYFQAEPSAAHSAEGDVNTLLLIFLHRAPELLAWADEQARSWAHIEPMYVP PDGPSLEA (SEQ ID NO: 120).

ExoI

[0420] Human exonuclease 1 (EXO1) has been implicated in many different DNA metabolic processes, including DNA mismatch repair (MMR), micro-mediated end-joining, homologous recombination (HR), and replication. Human EXO1 belongs to a family of eukaryotic nucleases, Rad2/XPG, which also include FEN1 and GEN1. The Rad2/XPG family is conserved in the nuclease domain through species from phage to human. The EXO1 gene product exhibits both 5′ exonuclease and 5′ flap activity. Additionally, EXO1 contains an intrinsic 5′ RNase H activity. Human EXO1 has a high affinity for processing double stranded DNA (dsDNA), nicks, gaps, pseudo Y structures and can resolve Holliday junctions using its inherit flap activity. Human EXO1 is implicated in MMR and contain conserved binding domains interacting directly with MLH1 and MSH2. EXO1 nucleolytic activity is positively stimulated by PCNA, MutSα (MSH2/MSH6 complex), 14-3-3, MRN and 9-1-1 complex.

[0421] exonuclease 1 (EXO1) Accession No. NM_003686 (Homo sapiens exonuclease 1 (EXO1), transcript variant 3) – isoform A

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYV GFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITE DSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDY LSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLY QLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYN PDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVG VERVISTKGLNLPRKSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSE VFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDC VSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDH IPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQ FRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASK LSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADS LSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKF (SEQ ID NO: 121).

[0422] exonuclease 1 (EXO1) Accession No. NM_006027 (Homo sapiens exonuclease 1 (EXO1), transcript variant 3) – isoform B

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYV
GFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS
EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITE
DSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDY
LSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLY
QLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYN
PDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVG
VERVISTKGLNLPRKSSIVKRPRSAELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSE
VFVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDC
VSNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDH
IPDKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQ
FRRKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASK
LSQCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADS
LSTTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKDSEK
LPPCKKPLSPVRDNIQLTPEAEEDIFNKPECGRVQRAIFQ (SEQ ID NO: 122).

[0423] exonuclease 1 (EXO1) Accession No. NM_001319224 (Homo sapiens exonuclease 1 (EXO1), transcript variant 4) – isoform C

MGIQGLLQFIKEASEPIHVRKYKGQVVAVDTYCWLHKGAIACAEKLAKGEPTDRYV
GFCMKFVNMLLSHGIKPILVFDGCTLPSKKEVERSRRERRQANLLKGKQLLREGKVS
EARECFTRSINITHAMAHKVIKAARSQGVDCLVAPYEADAQLAYLNKAGIVQAIITE
DSDLLAFGCKKVILKMDQFGNGLEIDQARLGMCRQLGDVFTEEKFRYMCILSGCDY
LSSLRGIGLAKACKVLRLANNPDIVKVIKKIGHYLKMNITVPEDYINGFIRANNTFLY
QLVFDPIKRKLIPLNAYEDDVDPETLSYAGQYVDDSIALQIALGNKDINTFEQIDDYN
PDTAMPAHSRSHSWDDKTCQKSANVSSIWHRNYSPRPESGTVSDAPQLKENPSTVG
VERVISTKGLNLPRKSSIVKRPRSELSEDDLLSQYSLSFTKKTKKNSSEGNKSLSFSEV
FVPDLVNGPTNKKSVSTPPRTRNKFATFLQRKNEESGAVVVPGTRSRFFCSSDSTDCV
SNKVSIQPLDETAVTDKENNLHESEYGDQEGKRLVDTDVARNSSDDIPNNHIPGDHIP
DKATVFTDEESYSFESSKFTRTISPPTLGTLRSCFSWSGGLGDFSRTPSPSPSTALQQFR
RKSDSPTSLPENNMSDVSQLKSEESSDDESHPLREEACSSQSQESGEFSLQSSNASKLS
QCSSKDSDSEESDCNIKLLDSQSDQTSKLRLSHFSKKDTPLRNKVPGLYKSSSADSLS
TTKIKPLGPARASGLSKKPASIQKRKHHNAENKPGLQIKLNELWKNFGFKKDSEKLP
PCKKPLSPVRDNIQLTPEAEEDIFNKPECGRVQRAIFQ (SEQ ID NO: 123).

D. Inteins and split-inteins

[0424] It will be understood that in some embodiments (e.g., delivery of a prime editor *in vivo* using AAV particles), it may be advantageous to split a polypeptide (e.g., a deaminase or a napDNAbp) or a fusion protein (e.g., a prime editor) into an N-terminal half and a C-terminal half, delivery them separately, and then allow their colocalization to reform the complete protein (or fusion protein as the case may be) within the cell. Separate halves of a protein or a fusion protein may each comprise a split-intein tag to facilitate the reformation of the complete protein or fusion protein by the mechanism of protein trans splicing.

[0425] Protein trans-splicing, catalyzed by split inteins, provides an entirely enzymatic method for protein ligation. A split-intein is essentially a contiguous intein (e.g. a mini-intein) split into two pieces named N-intein and C-intein, respectively. The N-intein and C-intein of a split intein can associate non-covalently to form an active intein and catalyze the splicing reaction essentially in same way as a contiguous intein does. Split inteins have been found in nature and also engineered in laboratories. As used herein, the term "split intein" refers to any intein in which one or more peptide bond breaks exists between the N-terminal and C-terminal amino acid sequences such that the N-terminal and C-terminal sequences become separate molecules that can non-covalently reassociate, or reconstitute, into an intein that is functional for trans-splicing reactions. Any catalytically active intein, or fragment thereof, may be used to derive a split intein for use in the methods of the invention. For example, in one aspect the split intein may be derived from a eukaryotic intein. In another aspect, the split intein may be derived from a bacterial intein. In another aspect, the split intein may be derived from an archaeal intein. Preferably, the split intein so-derived will possess only the amino acid sequences essential for catalyzing trans-splicing reactions.

[0426] As used herein, the "N-terminal split intein (In)" refers to any intein sequence that comprises an N- terminal amino acid sequence that is functional for trans-splicing reactions. An In thus also comprises a sequence that is spliced out when trans-splicing occurs. An In can comprise a sequence that is a modification of the N-terminal portion of a naturally occurring intein sequence. For example, an In can comprise additional amino acid residues and/or mutated residues so long as the inclusion of such additional and/or mutated residues does not render the In non-functional in trans-splicing. Preferably, the inclusion of the additional and/or mutated residues improves or enhances the trans-splicing activity of the In.

[0427] As used herein, the "C-terminal split intein (Ic)" refers to any intein sequence that comprises a C- terminal amino acid sequence that is functional for trans-splicing reactions. In one aspect, the Ic comprises 4 to 7 contiguous amino acid residues, at least 4 amino acids of

which are from the last β-strand of the intein from which it was derived. An Ic thus also comprises a sequence that is spliced out when trans-splicing occurs. An Ic can comprise a sequence that is a modification of the C-terminal portion of a naturally occurring intein sequence. For example, an Ic can comprise additional amino acid residues and/or mutated residues so long as the inclusion of such additional and/or mutated residues does not render the In non-functional in trans-splicing. Preferably, the inclusion of the additional and/or mutated residues improves or enhances the trans-splicing activity of the Ic.

[0428] In some embodiments of the invention, a peptide linked to an Ic or an In can comprise an additional chemical moiety including, among others, fluorescence groups, biotin, polyethylene glycol (PEG), amino acid analogs, unnatural amino acids, phosphate groups, glycosyl groups, radioisotope labels, and pharmaceutical molecules. In other embodiments, a peptide linked to an Ic can comprise one or more chemically reactive groups including, among others, ketone, aldehyde, Cys residues and Lys residues. The N-intein and C-intein of a split intein can associate non-covalently to form an active intein and catalyze the splicing reaction when an "intein-splicing polypeptide (ISP)" is present. As used herein, "intein-splicing polypeptide (ISP)" refers to the portion of the amino acid sequence of a split intein that remains when the Ic, In, or both, are removed from the split intein. In certain embodiments, the In comprises the ISP. In another embodiment, the Ic comprises the ISP. In yet another embodiment, the ISP is a separate peptide that is not covalently linked to In nor to Ic.

[0429] Split inteins may be created from contiguous inteins by engineering one or more split sites in the unstructured loop or intervening amino acid sequence between the -12 conserved beta-strands found in the structure of mini-inteins. Some flexibility in the position of the split site within regions between the beta-strands may exist, provided that creation of the split will not disrupt the structure of the intein, the structured beta-strands in particular, to a sufficient degree that protein splicing activity is lost.

[0430] In protein trans-splicing, one precursor protein consists of an N-extein part followed by the N-intein, another precursor protein consists of the C-intein followed by a C-extein part, and a trans-splicing reaction (catalyzed by the N- and C-inteins together) excises the two intein sequences and links the two extein sequences with a peptide bond. Protein trans-splicing, being an enzymatic reaction, can work with very low (e.g. micromolar) concentrations of proteins and can be carried out under physiological conditions.

[0431] Exemplary sequences are as follows:

NAME	SEQUENCE OF LIGAND-DEPENDENT INTEIN

2-4 INTEIN:	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 124)
3-2 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGDR VAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEASM MGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWLEIL MIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATS SRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRAL DKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL YSMKYTNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDDK FLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHN C (SEQ ID NO: 125)
30R3-1 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 126)
30R3-2 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 127)
30R3-3 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPIPYSEYDPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLECAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 128)

37R3-1 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKG DRVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYNPTSPFSEA SMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWL EILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLL ATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIH RALDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGM EHLYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADAL DDKFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVV VHNC (SEQ ID NO: 129)
37R3-2 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLLARPVV SWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEGLRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 130)
37R3-3 INTEIN	CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAVAKDGTLLARPVVS WFDQGTRDVIGLRIAGGATVWATPDHKVLTEYGWRAAGELRKGD RVAGPGGSGNSLALSLTADQMVSALLDAEPPILYSEYDPTSPFSEAS MMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQAHLLERAWLEI LMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLAT SSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRA LDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEH LYSMKYKNVVPLYDLLLEMLDAHRLHAGGSGASRVQAFADALDD KFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVH NC (SEQ ID NO: 131)

[0432] Although inteins are most frequently found as a contiguous domain, some exist in a naturally split form. In this case, the two fragments are expressed as separate polypeptides and must associate before splicing takes place, so-called protein trans-splicing.

[0433] An exemplary split intein is the *Ssp* DnaE intein, which comprises two subunits, namely, DnaE-N and DnaE-C. The two different subunits are encoded by separate genes, namely *dnaE-n* and *dnaE-c*, which encode the DnaE-N and DnaE-C subunits, respectively. DnaE is a naturally occurring split intein in *Synechocytis sp.* PCC6803 and is capable of directing trans-splicing of two separate proteins, each comprising a fusion with either DnaE-N or DnaE-C.

[0434] Additional naturally occurring or engineered split-intein sequences are known in the or can be made from whole-intein sequences described herein or those available in the art. Examples of split-intein sequences can be found in Stevens *et al.*, "A promiscuous split intein with expanded protein engineering applications," PNAS, 2017, Vol.114: 8538-8543; Iwai *et*

al., "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostc punctiforme, FEBS Lett, 580: 1853-1858, each of which are incorporated herein by reference. Additional split intein sequences can be found, for example, in WO 2013/045632, WO 2014/055782, WO 2016/069774, and EP2877490, the contents each of which are incorporated herein by reference.

In addition, protein splicing in trans has been described *in vivo* and *in vitro* (Shingledecker, *et al.*, *Gene* 207:187 (1998), Southworth, *et al.*, *EMBO J.* 17:918 (1998); Mills, *et al.*, *Proc. Natl. Acad. Sci. USA,* 95:3543-3548 (1998); Lew, *et al.*, *J. Biol. Chem.,* 273:15887-15890 (1998); Wu, *et al.*, *Biochim. Biophys. Acta* 35732:1 (1998b), Yamazaki, *et al.*, *J. Am. Chem. Soc.* 120:5591 (1998), Evans, *et al.*, *J. Biol. Chem.* 275:9091 (2000); Otomo, *et al.*, *Biochemistry* 38:16040-16044 (1999); Otomo, *et al.*, *J. Biolmol. NMR* 14:105-114 (1999); Scott, *et al.*, *Proc. Natl. Acad. Sci. USA* 96:13638-13643 (1999)) and provides the opportunity to express a protein as to two inactive fragments that subsequently undergo ligation to form a functional product.

RNA-protein interaction domain

[0435] In various embodiments, two separate protein domains (e.g., a Cas9 domain and a polymerase domain) may be colocalized to one another to form a functional complex (akin to the function of a fusion protein comprising the two separate protein domains) by using an "RNA-protein recruitment system," such as the "MS2 tagging technique." Such systems generally tag one protein domain with an "RNA-protein interaction domain" (aka "RNA-protein recruitment domain") and the other with an "RNA-binding protein" that specifically recognizes and binds to the RNA-protein interaction domain, e.g., a specific hairpin structure. These types of systems can be leveraged to colocalize the domains of a prime editor, as well as to recruitment additional functionalities to a prime editor, such as a UGI domain. In one example, the MS2 tagging technique is based on the natural interaction of the MS2 bacteriophage coat protein ("MCP" or "MS2cp") with a stem-loop or hairpin structure present in the genome of the phage, i.e., the "MS2 hairpin." In the case of the MS2 hairpin, it is recognized and bound by the MS2 bacteriophage coat protein (MCP). Thus, in one exemplary scenario a deaminase-MS2 fusion can recruit a Cas9-MCP fusion.

[0436] A review of other modular RNA-protein interaction domains are described in the art, for example, in Johansson *et al.*, "RNA recognition by the MS2 phage coat protein," *Sem Virol.*, 1997, Vol. 8(3): 176-185; Delebecque *et al.*, "Organization of intracellular reactions with rationally designed RNA assemblies," *Science*, 2011, Vol. 333: 470-474; Mali *et al.*, "Cas9 transcriptional activators for target specificity screening and paired nickases for

cooperative genome engineering," *Nat. Biotechnol.*, 2013, Vol.31: 833-838; and Zalatan *et al.*, "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds," *Cell*, 2015, Vol.160: 339-350, each of which are incorporated herein by reference in their entireties. Other systems include the PP7 hairpin, which specifically recruits the PCP protein, and the "com" hairpin, which specifically recruits the Com protein. See Zalatan *et al.*

[0437] The nucleotide sequence of the MS2 hairpin (or equivalently referred to as the "MS2 aptamer") is: GCCAACATGAGGATCACCCATGTCTGCAGGGCC (SEQ ID NO: 144). The amino acid sequence of the MCP or MS2cp is: GSASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ NRKYTIKVEVPKVATQTVGGEELPVAGWRSYLNMELTIPIFATNSDCELIVKAMQGL LKDGNPIPSAIAANSGIY (SEQ ID NO: 145).

E. UGI domain

[0438] In other embodiments, the prime editors utilized in the methods and compositions described herein may comprise one or more uracil glycosylase inhibitor domains. The term "uracil glycosylase inhibitor (UGI)" or "UGI domain," as used herein, refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a UGI as set forth in SEQ ID NO: 132. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, a UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 132. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 132. In some embodiments, a UGI comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 132, or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 132. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment thereof. For example, a UGI variant is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 132. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least

80% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 132. In some embodiments, the UGI comprises the following amino acid sequence: Uracil-DNA glycosylase inhibitor:
>sp|P14739|UNGI_BPPB2
MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT SDAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 132).

[0439] The prime editors utilized in the methods and compositions described herein may comprise more than one UGI domain, which may be separated by one or more linkers as described herein.

F. Additional PE elements

[0440] In certain embodiments, the prime editors utilized in the methods and compositions described herein may comprise an inhibitor of base repair. The term "inhibitor of base repair" or "IBR" refers to a protein that is capable in inhibiting the activity of a nucleic acid repair enzyme, for example a base excision repair enzyme. In some embodiments, the IBR is an inhibitor of OGG base excision repair. In some embodiments, the IBR is an inhibitor of base excision repair ("iBER"). Exemplary inhibitors of base excision repair include inhibitors of APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGG1, hNEIL1, T7 EndoI, T4PDG, UDG, hSMUG1, and hAAG. In some embodiments, the IBR is an inhibitor of Endo V or hAAG. In some embodiments, the IBR is an iBER that may be a catalytically inactive glycosylase or catalytically inactive dioxygenase or a small molecule or peptide inhibitor of an oxidase, or variants threreof. In some embodiments, the IBR is an iBER that may be a TDG inhibitor, MBD4 inhibitor or an inhibitor of an AlkBH enzyme. In some embodiments, the IBR is an iBER that comprises a catalytically inactive TDG or catalytically inactive MBD4. An exemplary catalytically inactive TDG is an N140A mutant of SEQ ID NO: 136 (human TDG).

[0441] Some exemplary glycosylases are provided below. The catalytically inactivated variants of any of these glycosylase domains are iBERs that may be fused to the napDNAbp or polymerase domain of the prime editors utilized in the methods and compositions provided in this disclosure.

[0442] OGG (human)
MPARALLPRRMGHRTLASTPALWASIPCPRSELRLDLVLPSGQSFRWREQSPAHWSG VLADQVWTLTQTEEQLHCTVYRGDKSQASRPTPDELEAVRKYFQLDVTLAQLYHH

WGSVDSHFQEVAQKFQGVRLLRQDPIECLFSFICSSNNNIARITGMVERLCQAFGPRL
IQLDDVTYHGFPSLQALAGPEVEAHLRKLGLGYRARYVSASARAILEEQGGLAWLQ
QLRESSYEEAHKALCILPGVGTKVADCICLMALDKPQAVPVDVHMWHIAQRDYSW
HPTTSQAKGPSPQTNKELGNFFRSLWGPYAGWAQAVLFSADLRQSRHAQEPPAKRR
KGSKGPEG (SEQ ID NO: 133)

[0443] MPG (human)

MVTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAAQAP
CPRERCLGPPTTPGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPN
GTELRGRIVETEAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNI
SSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAI
NKSFDQRDLAQDEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGS
PWVSVVDRVAEQDTQA (SEQ ID NO: 134)

[0444] MBD4 (human)

MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIK
RSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISP
QGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQ
SNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRK
VRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQL
DRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSE
HNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQE
DTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKL
LIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTI
VKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLW
ENHEKLSLS (SEQ ID NO: 135)

TDG (human)

MEAENAGSYSLQQAQAFYTFPFQQLMAEAPNMAVVNEQQMPEEVPAPAPAQEPVQ
EAPKGRKRKPRTTEPKQPVEPKKPVESKKSGKSAKSKEKQEKITDTFKVKRKVDRFN
GVSEAELLTKTLPDILTFNLDIVIIGINPGLMAAYKGHHYPGPGNHFWKCLFMSGLSE
VQLNHMDDHTLPGKYGIGFTNMVERTTPGSKDLSSKEFREGGRILVQKLQKYQPRIA
VFNGKCIYEIFSKEVFGVKVKNLEFGLQPHKIPDTETLCYVMPSSSARCAQFPRAQDK
VHYYIKLKDLRDQLKGIERNMDVQEVQYTFDLQLAQEDAKKMAVKEEKYDPGYEA
AYGGAYGENPCSSEPCGFSSNGLIESVELRGESAFSGIPNGQWMTQSFTDQIPSFSNH
CGTQEQEEESHA (SEQ ID NO: 136)

[0445] In some embodiments, the fusion proteins described herein may comprise one or more heterologous protein domains (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the prime editor components). A fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Other exemplary features that may be present 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.

[0446] Examples of protein domains that may be fused to a prime editor or component thereof (e.g., the napDNAbp domain, the polymerase domain, or the NLS domain) include, without limitation, epitope tags, and reporter gene sequences. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT), beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A prime editor may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including, but not limited to, maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a prime editor are described in US Patent Publication No. 2011/0059502, published March 10, 2011 and incorporated herein by reference in its entirety.

[0447] In an aspect of the disclosure, a reporter gene which includes, but is not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP), may be introduced into a cell to encode a gene product which serves as a marker by which to measure the alteration or modification of expression of the gene product. In certain embodiments of the disclosure the gene product is luciferase. In a further embodiment of the disclosure the expression of the gene product is decreased.

[0448] 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.

[0449] In some embodiments of the present disclosure, the activity of the prime editing system may be temporally regulated by adjusting the residence time, the amount, and/or the activity of the expressed components of the PE system. For example, as described herein, the PE may be fused with a protein domain that is capable of modifying the intracellular half-life of the PE. In certain embodiments involving two or more vectors (e.g., a vector system in which the components described herein are encoded on two or more separate vectors), the activity of the PE system may be temporally regulated by controlling the timing in which the vectors are delivered. For example, in some embodiments a vector encoding the nuclease system may deliver the PE prior to the vector encoding the template. In other embodiments, the vector encoding the PEgRNA may deliver the guide prior to the vector encoding the PE system. In some embodiments, the vectors encoding the PE system and PEgRNA are delivered simultaneously. In certain embodiments, the simultaneously delivered vectors temporally deliver, e.g., the PE, PEgRNA, and/or second strand guide RNA components. In further embodiments, the RNA (such as, e.g., the nuclease transcript) transcribed from the coding sequence on the vectors may further comprise at least one element that is capable of modifying the intracellular half-life of the RNA and/or modulating translational control. In some embodiments, the half-life of the RNA may be increased. In some embodiments, the half-life of the RNA may be decreased. In some embodiments, the element may be capable of increasing the stability of the RNA. In some embodiments, the element may be capable of decreasing the stability of the RNA. In some embodiments, the element may be within the 3' UTR of the RNA. In some embodiments, the element may include a polyadenylation signal (PA). In some embodiments, the element may include a cap, e.g., an upstream mRNA or PEgRNA end. In some embodiments, the RNA may comprise no PA such that it is subject to quicker degradation in the cell after transcription. In some embodiments, the element may include at least one AU-rich element (ARE). The AREs may be bound by ARE binding proteins (ARE-BPs) in a manner that is dependent upon tissue type, cell type, timing, cellular

localization, and environment. In some embodiments the destabilizing element may promote RNA decay, affect RNA stability, or activate translation. In some embodiments, the ARE may comprise 50 to 150 nucleotides in length. In some embodiments, the ARE may comprise at least one copy of the sequence AUUUA. In some embodiments, at least one ARE may be added to the 3' UTR of the RNA. In some embodiments, the element may be a Woodchuck Hepatitis Virus (WHP).

[0450] Posttranscriptional Regulatory Element (WPRE), which creates a tertiary structure to enhance expression from the transcript. In further embodiments, the element is a modified and/or truncated WPRE sequence that is capable of enhancing expression from the transcript, as described, for example in Zufferey *et al.*, J Virol, 73(4): 2886-92 (1999) and Flajolet *et al.*, J Virol, 72(7): 6175-80 (1998). In some embodiments, the WPRE or equivalent may be added to the 3' UTR of the RNA. In some embodiments, the element may be selected from other RNA sequence motifs that are enriched in either fast- or slow-decaying transcripts.

In some embodiments, the vector encoding the PE or the PEgRNA may be self-destroyed via cleavage of a target sequence present on the vector by the PE system. The cleavage may prevent continued transcription of a PE or a PEgRNA from the vector. Although transcription may occur on the linearized vector for some amount of time, the expressed transcripts or proteins subject to intracellular degradation will have less time to produce off-target effects without continued supply from expression of the encoding vectors.

PEgRNAs

[0451] The prime editing system utilized in the methods and compositions described herein contemplates the use of any suitable PEgRNAs.

PEgRNA architecture

[0452] In some embodiments, an extended guide RNA usable in the prime editing system utilized in the methods and compositions disclosed herein whereby a traditional guide RNA includes a ~20 nt protospacer sequence and a gRNA core region, which binds with the napDNAbp. In this embodiment, the guide RNA includes an extended RNA segment at the 5′ end, i.e., a 5′ extension. In this embodiment, the 5′extension includes a reverse transcription template sequence, a reverse transcription primer binding site, and an optional 5-20 nucleotide linker sequence. The RT primer binding site hybridizes to the free 3′ end that is formed after a nick is formed in the non-target strand of the R-loop, thereby priming reverse transcriptase for DNA polymerization in the 5′-3′ direction.

[0453] In another embodiment, an extended guide RNA usable in the prime editing system utilized in the methods and compositions disclosed herein whereby a traditional guide RNA includes a ~20 nt protospacer sequence and a gRNA core, which binds with the napDNAbp. In this embodiment, the guide RNA includes an extended RNA segment at the 3′ end, i.e., a 3′ extension. In this embodiment, the 3′extension includes a reverse transcription template sequence, and a reverse transcription primer binding site. The RT primer binding site hybridizes to the free 3′ end that is formed after a nick is formed in the non-target strand of the R-loop, thereby priming reverse transcriptase for DNA polymerization in the 5′-3′ direction.

[0454] In another embodiment, an extended guide RNA usable in the prime editing system utilized in the methods and compositions disclosed herein whereby a traditional guide RNA includes a ~20 nt protospacer sequence and a gRNA core, which binds with the napDNAbp. In this embodiment, the guide RNA includes an extended RNA segment at an intermolecular position within the gRNA core, i.e., an intramolecular extension. In this embodiment, the intramolecular extension includes a reverse transcription template sequence, and a reverse transcription primer binding site. The RT primer binding site hybridizes to the free 3′ end that is formed after a nick is formed in the non-target strand of the R-loop, thereby priming reverse transcriptase for DNA polymerization in the 5′-3′ direction.

[0455] In one embodiment, the position of the intermolecular RNA extension is not in the protospacer sequence of the guide RNA. In another embodiment, the position of the intermolecular RNA extension in the gRNA core. In still another embodiment, the position of the intermolecular RNA extension is any with the guide RNA molecule except within the protospacer sequence, or at a position which disrupts the protospacer sequence.

In one embodiment, the intermolecular RNA extension is inserted downstream from the 3′ end of the protospacer sequence. In another embodiment, the intermolecular RNA extension is inserted at least 1 nucleotide, at least 2 nucleotides, at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, at least 22 nucleotides, at least 23 nucleotides, at least 24 nucleotides, at least 25 nucleotides downstream of the 3′ end of the protospacer sequence.

[0456] In other embodiments, the intermolecular RNA extension is inserted into the gRNA, which refers to the portion of the guide RNA corresponding or comprising the tracrRNA,

which binds and/or interacts with the Cas9 protein or equivalent thereof (i.e, a different napDNAbp). Preferably the insertion of the intermolecular RNA extension does not disrupt or minimally disrupts the interaction between the tracrRNA portion and the napDNAbp.

[0457] The length of the RNA extension (which includes at least the RT template and primer binding site) can be any useful length. In various embodiments, the RNA extension is at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 21 nucleotides, at least 22 nucleotides, at least 23 nucleotides, at least 24 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 40 nucleotides, at least 50 nucleotides, at least 60 nucleotides, at least 70 nucleotides, at least 80 nucleotides, at least 90 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, or at least 500 nucleotides in length.

[0458] The RT template sequence can also be any suitable length. For example, the RT template sequence can be at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 30 nucleotides, at least 40 nucleotides, at least 50 nucleotides, at least 60 nucleotides, at least 70 nucleotides, at least 80 nucleotides, at least 90 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, or at least 500 nucleotides in length.

[0459] In still other embodiments, wherein the reverse transcription primer binding site sequence is at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 30 nucleotides, at least 40 nucleotides, at least 50 nucleotides, at least 60 nucleotides, at least 70 nucleotides, at least 80 nucleotides, at least 90 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, or at least 500 nucleotides in length.

[0460] In other embodiments, the optional linker or spacer sequence is at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at

least 8 nucleotides, at least 9 nucleotides, at least 10 nucleotides, at least 11 nucleotides, at least 12 nucleotides, at least 13 nucleotides, at least 14 nucleotides, at least 15 nucleotides, at least 16 nucleotides, at least 17 nucleotides, at least 18 nucleotides, at least 19 nucleotides, at least 20 nucleotides, at least 30 nucleotides, at least 40 nucleotides, at least 50 nucleotides, at least 60 nucleotides, at least 70 nucleotides, at least 80 nucleotides, at least 90 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, or at least 500 nucleotides in length.

[0461] The RT template sequence, in certain embodiments, encodes a single-stranded DNA molecule which is homologous to the non-target strand (and thus, complementary to the corresponding site of the target strand) but includes one or more nucleotide changes. The least one nucleotide change may include one or more single-base nucleotide changes, one or more deletions, and one or more insertions.

[0462] The synthesized single-stranded DNA product of the RT template sequence is homologous to the non-target strand and contains one or more nucleotide changes. The single-stranded DNA product of the RT template sequence hybridizes in equilibrium with the complementary target strand sequence, thereby displacing the homologous endogenous target strand sequence. The displaced endogenous strand may be referred to in some embodiments as a 5′ endogenous DNA flap species. This 5′ endogenous DNA flap species can be removed by a 5′ flap endonuclease (e.g., FEN1) and the single-stranded DNA product, now hybridized to the endogenous target strand, may be ligated, thereby creating a mismatch between the endogenous sequence and the newly synthesized strand. The mismatch may be resolved by the cell's innate DNA repair and/or replication processes.

[0463] In various embodiments, the nucleotide sequence of the RT template sequence corresponds to the nucleotide sequence of the non-target strand which becomes displaced as the 5′ flap species and which overlaps with the site to be edited.

[0464] In various embodiments of the extended guide RNAs, the reverse transcription template sequence may encode a single-strand DNA flap that is complementary to an endogenous DNA sequence adjacent to a nick site, wherein the single-strand DNA flap comprises a desired nucleotide change. The single-stranded DNA flap may displace an endogenous single-strand DNA at the nick site. The displaced endogenous single-strand DNA at the nick site can have a 5′ end and form an endogenous flap, which can be excised by the cell. In various embodiments, excision of the 5′ end endogenous flap can help drive product formation since removing the 5′ end endogenous flap encourages hybridization of the single-strand 3′ DNA flap to the corresponding complementary DNA strand, and the incorporation

or assimilation of the desired nucleotide change carried by the single-strand 3´ DNA flap into the target DNA.

[0465] In various embodiments of the extended guide RNAs, the cellular repair of the single-strand DNA flap results in installation of the desired nucleotide change, thereby forming a desired product.

[0466] In still other embodiments, the desired nucleotide change is installed in an editing window that is between about -5 to +5 of the nick site, or between about -10 to +10 of the nick site, or between about -20 to +20 of the nick site, or between about -30 to +30 of the nick site, or between about -40 to + 40 of the nick site, or between about -50 to +50 of the nick site, or between about -60 to +60 of the nick site, or between about -70 to +70 of the nick site, or between about -80 to +80 of the nick site, or between about -90 to +90 of the nick site, or between about -100 to +100 of the nick site, or between about -200 to +200 of the nick site.

[0467] In other embodiments, the desired nucleotide change is installed in an editing window that is between about +1 to +2 from the nick site, or about +1 to +3, +1 to +4, +1 to +5, +1 to +6, +1 to +7, +1 to +8, +1 to +9, +1 to +10, +1 to +11, +1 to +12, +1 to +13, +1 to +14, +1 to +15, +1 to +16, +1 to +17, +1 to +18, +1 to +19, +1 to +20, +1 to +21, +1 to +22, +1 to +23, +1 to +24, +1 to +25, +1 to +26, +1 to +27, +1 to +28, +1 to +29, +1 to +30, +1 to +31, +1 to +32, +1 to +33, +1 to +34, +1 to +35, +1 to +36, +1 to +37, +1 to +38, +1 to +39, +1 to +40, +1 to +41, +1 to +42, +1 to +43, +1 to +44, +1 to +45, +1 to +46, +1 to +47, +1 to +48, +1 to +49, +1 to +50, +1 to +51, +1 to +52, +1 to +53, +1 to +54, +1 to +55, +1 to +56, +1 to +57, +1 to +58, +1 to +59, +1 to +60, +1 to +61, +1 to +62, +1 to +63, +1 to +64, +1 to +65, +1 to +66, +1 to +67, +1 to +68, +1 to +69, +1 to +70, +1 to +71, +1 to +72, +1 to +73, +1 to +74, +1 to +75, +1 to +76, +1 to +77, +1 to +78, +1 to +79, +1 to +80, +1 to +81, +1 to +82, +1 to +83, +1 to +84, +1 to +85, +1 to +86, +1 to +87, +1 to +88, +1 to +89, +1 to +90, +1 to +90, +1 to +91, +1 to +92, +1 to +93, +1 to +94, +1 to +95, +1 to +96, +1 to +97, +1 to +98, +1 to +99, +1 to +100, +1 to +101, +1 to +102, +1 to +103, +1 to +104, +1 to +105, +1 to +106, +1 to +107, +1 to +108, +1 to +109, +1 to +110, +1 to +111, +1 to +112, +1 to +113, +1 to +114, +1 to +115, +1 to +116, +1 to +117, +1 to +118, +1 to +119, +1 to +120, +1 to +121, +1 to +122, +1 to +123, +1 to +124, or +1 to +125 from the nick site.

[0468] In still other embodiments, the desired nucleotide change is installed in an editing window that is between about +1 to +2 from the nick site, or about +1 to +5, +1 to +10, +1 to +15, +1 to +20, +1 to +25, +1 to +30, +1 to +35, +1 to +40, +1 to +45, +1 to +50, +1 to +55, +1 to +100, +1 to +105, +1 to +110, +1 to +115, +1 to +120, +1 to +125, +1 to +130, +1 to

+135, +1 to +140, +1 to +145, +1 to +150, +1 to +155, +1 to +160, +1 to +165, +1 to +170, +1 to +175, +1 to +180, +1 to +185, +1 to +190, +1 to +195, or +1 to +200, from the nick site.

[0469] In various aspects, the extended guide RNAs are modified versions of a guide RNA. Guide RNAs maybe naturally occurring, expressed from an encoding nucleic acid, or synthesized chemically. Methods are well known in the art for obtaining or otherwise synthesizing guide RNAs and for determining the appropriate sequence of the guide RNA, including the protospacer sequence which interacts and hybridizes with the target strand of a genomic target site of interest.

[0470] In various embodiments, the particular design aspects of a guide RNA sequence will depend upon the nucleotide sequence of a genomic target site of interest (i.e., the desired site to be edited) and the type of napDNAbp (e.g., Cas9 protein) present in the prime editing systems utilized in the methods and compositions described herein, among other factors, such as PAM sequence locations, percent G/C content in the target sequence, the degree of microhomology regions, secondary structures, etc.

[0471] In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a napDNAbp (e.g., a Cas9, Cas9 homolog, or Cas9 variant) to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length.

[0472] In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a prime editor to a target sequence may be assessed by any suitable assay. For example, the components of a prime editor, including the guide sequence to be tested,

may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of a prime editor disclosed herein, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a prime editor, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0473] A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 298) where in the portion containing NNNNNNNNNNNNXGG, N is A, G, T, or C; and X can be anything. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 299) where in the portion containing NNNNNNNNNNNXGG, N is A, G, T, or C; and X can be anything. For the S. thermophilus CRISPR1Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 300) where in the portion containing NNNNNNNNNNNNXXAGAAW, N is A, G, T, or C; X can be anything; and W is A or T. A unique target sequence in a genome may include an S. thermophilus CRISPR 1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 301) where in the portion containing NNNNNNNNNNNXXAGAAW, N is A, G, T, or C; X can be anything; and W is A or T. For the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 302) where in the portion containing NNNNNNNNNNNXGGXG, N is A, G, T, or C; and X can be anything. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 303) where in the portion containing NNNNNNNNNNNXGGXG, N is A, G, T, or C; and X can be anything. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique.

[0474] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber *et al.*, 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62). Further algorithms may be found in U.S. application Ser. No. 61/836,080; Broad Reference BI-2013/004A); incorporated herein by reference.

[0475] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a complex at a target sequence, wherein the complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. Preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five

hairpins. In some embodiments, the single transcript further includes a transcription termination sequence; preferably this is a polyT sequence, for example six T nucleotides. Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5′ to 3′), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator:

(1)NNNNNNNNGTTTTTGTACTCTCAAGATTTAGAAATAAATCTTGCAGAAGCTACA AAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTC GTTATTTAATTTTTT (SEQ ID NO: 137);

(2)NNNNNNNNNNNNNNNNNNGTTTTTGTACTCTCAGAAATGCAGAAGCTACAAA GATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTCGT TATTTAATTTTTT (SEQ ID NO: 138);

(3)NNNNNNNNNNNNNNNNNNNNGTTTTTGTACTCTCAGAAATGCAGAAGCTACA AAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTTT T (SEQ ID NO: 139);

(4)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT (SEQ ID NO: 140);

(5)NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA GGCTAGTCCGTTATCAACTTGAAAAAGTGTTTTTTT (SEQ ID NO: 141); AND

(6)
NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTCCGTTATCATTTTTTTT (SEQ ID NO: 142).

[0476] In some embodiments, sequences (1) to (3) are used in combination with Cas9 from S. thermophilus CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from S. pyogenes. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

[0477] 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 single-stranded DNA binding protein, as disclosed herein, to a target site, e.g., a site comprising a point 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.

[0478] In some embodiments, the guide RNA comprises a structure 5′-[guide sequence]-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUGCUUUUU-3′ (SEQ ID NO: 143), 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. Additional guide sequences are well known in the art and can be used with the prime editors utilized in the methods and compositions described herein.

[0479] In some embodiments, a PEgRNA comprises three main component elements ordered in the 5′ to 3′ direction, namely: a spacer, a gRNA core, and an extension arm at the 3′ end. The extension arm may further be divided into the following structural elements in the 5′ to 3′ direction, namely: a primer binding site (A), an edit template (B), and a homology arm (C). In addition, the PEgRNA may comprise an optional 3′ end modifier region (e1) and an optional 5′ end modifier region (e2). Still further, the PEgRNA may comprise a transcriptional termination signal at the 3′ end of the PEgRNA (not depicted). These structural elements are further defined herein. The depiction of the structure of the PEgRNA is not meant to be limiting and embraces variations in the arrangement of the elements. For example, the optional sequence modifiers (e1) and (e2) could be positioned within or between any of the other regions shown, and not limited to being located at the 3′ and 5′ ends.

[0480] In some embodiments, a PEgRNA contemplated herein and may be designed in accordance with the methodology defined in Example 2. The PEgRNA comprises three main component elements ordered in the 5′ to 3′ direction, namely: a spacer, a gRNA core, and an extension arm at the 3′ end. The extension arm may further be divided into the following structural elements in the 5′ to 3′ direction, namely: a primer binding site (A), an edit template (B), and a homology arm (C). In addition, the PEgRNA may comprise an optional 3′ end modifier region (e1) and an optional 5′ end modifier region (e2). Still further, the PEgRNA may comprise a transcriptional termination signal on the 3′ end of the PEgRNA

(not depicted). These structural elements are further defined herein. The depiction of the structure of the PEgRNA is not meant to be limiting and embraces variations in the arrangement of the elements. For example, the optional sequence modifiers (e1) and (e2) could be positioned within or between any of the other regions shown, and not limited to being located at the 3′ and 5′ ends.

PEgRNA modifications

[0481] The PEgRNAs may also include additional design modifications that may alter the properties and/or characteristics of PEgRNAs thereby improving the efficacy of prime editing. In various embodiments, these modifications may belong to one or more of a number of different categories, including but not limited to: (1) designs to enable efficient expression of functional PEgRNAs from non-polymerase III (pol III) promoters, which would enable the expression of longer PEgRNAs without burdensome sequence requirements; (2) modifications to the core, Cas9-binding PEgRNA scaffold, which could improve efficacy; (3) modifications to the PEgRNA to improve RT processivity, enabling the insertion of longer sequences at targeted genomic loci; and (4) addition of RNA motifs to the 5′ or 3′ termini of the PEgRNA that improve PEgRNA stability, enhance RT processivity, prevent misfolding of the PEgRNA, or recruit additional factors important for genome editing.

[0482] In one embodiment, PEgRNA could be designed with polIII promoters to improve the expression of longer-length PEgRNA with larger extension arms. sgRNAs are typically expressed from the U6 snRNA promoter. This promoter recruits pol III to express the associated RNA and is useful for expression of short RNAs that are retained within the nucleus. However, pol III is not highly processive and is unable to express RNAs longer than a few hundred nucleotides in length at the levels required for efficient genome editing. Additionally, pol III can stall or terminate at stretches of U's, potentially limiting the sequence diversity that could be inserted using a PEgRNA. Other promoters that recruit polymerase II (such as pCMV) or polymerase I (such as the U1 snRNA promoter) have been examined for their ability to express longer sgRNAs. However, these promoters are typically partially transcribed, which would result in extra sequence 5′ of the spacer in the expressed PEgRNA, which has been shown to result in markedly reduced Cas9:sgRNA activity in a site-dependent manner. Additionally, while pol III-transcribed PEgRNAs can simply terminate in a run of 6-7 U's, PEgRNAs transcribed from pol II or pol I would require a different termination signal. Often such signals also result in polyadenylation, which would result in undesired transport of the PEgRNA from the nucleus. Similarly, RNAs expressed

from pol II promoters such as pCMV are typically 5′-capped, also resulting in their nuclear export.

[0483] Previously, Rinn and coworkers screened a variety of expression platforms for the production of long-noncoding RNA- (lncRNA) tagged sgRNAs[183]. These platforms include RNAs expressed from pCMV and that terminate in the ENE element from the MALAT1 ncRNA from humans[184], the PAN ENE element from KSHV[185], or the 3′ box from U1 snRNA[186]. Notably, the MALAT1 ncRNA and PAN ENEs form triple helices protecting the polyA-tail [184, 187]. These constructs could also enhance RNA stability. It is contemplated that these expression systems will also enable the expression of longer PEgRNAs.

[0484] In addition, a series of methods have been designed for the cleavage of the portion of the pol II promoter that would be transcribed as part of the PEgRNA, adding either a self-cleaving ribozyme such as the hammerhead[188], pistol[189], hatchet[189], hairpin[190], VS[191], twister[192], or twister sister[192] ribozymes, or other self-cleaving elements to process the transcribed guide, or a hairpin that is recognized by Csy4[193] and also leads to processing of the guide. Also, it is hypothesized that incorporation of multiple ENE motifs could lead to improved PEgRNA expression and stability, as previously demonstrated for the KSHV PAN RNA and element[185]. It is also anticipated that circularizing the PEgRNA in the form of a circular intronic RNA (ciRNA) could also lead to enhanced RNA expression and stability, as well as nuclear localization[194].

In various embodiments, the PEgRNA may include various above elements, as exemplified by the following sequence.

[0485] Non-limiting example 1 - PEgRNA expression platform consisting of pCMV, Csy4 hairpin, the PEgRNA, and MALAT1 ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTTAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATTGTTTT
CTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAAAGCAAAAGAT

GCTGGTGGTTGGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCT
TTGCTTTGACT (SEQ ID NO: 147)

[0486] Non-limiting example 2 - PEgRNA expression platform consisting of pCMV, Csy4 hairpin, the PEgRNA, and PAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCA
GACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAAT
TTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 148)

[0487] Non-limiting example 3 - PEgRNA expression platform consisting of pCMV, Csy4 hairpin, the PEgRNA, and 3xPAN ENE

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTTGTTTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCA
GACGGCAAGGTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAAT
TTAACGTGCCTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAAC
ATAAATAAAGGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAAACACACTGT
TTTGGCTGGGTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAG
GTTTTTATCCCAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGC
CTAGAGCTCAAATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAA
GGTCAATGTTTAATCCATAAAAAAAAAAAAAAAAAAAATCTCTCTGTTTTGGCTGG
GTTTTTCCTTGTTCGCACCGGACACCTCCAGTGACCAGACGGCAAGGTTTTTATCC
CAGTGTATATTGGAAAAACATGTTATACTTTTGACAATTTAACGTGCCTAGAGCTCA

AATTAAACTAATACCATAACGTAATGCAACTTACAACATAAATAAAGGTCAATGTTT
AATCCATAAAAAAAAAAAAAAAAAAA (SEQ ID NO: 149)

[0488] Non-limiting example 4 - PEgRNA expression platform consisting of pCMV, Csy4 hairpin, the PEgRNA, and 3′ box

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCC
GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTA
CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG
GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTC
CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTGGTTTAGTGAACCGTCAGATCGTTCACTGCCGTATAGGCAGGGCCCAGA
CTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGT
TATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAG
TCTGTTTGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTTGTTTGGTT
TGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 150)

[0489] Non-limiting example 5 - PEgRNA expression platform consisting of pU1, Csy4 hairpin, the PEgRNA, and 3′ box

CTAAGGACCAGCTTCTTTGGGAGAGAACAGACGCAGGGGCGGGAGGGAAAAAG
GGAGAGGCAGACGTCACTTCCCCTTGGCGGCTCTGGCAGCAGATTGGTCGGTTGA
GTGGCAGAAAGGCAGACGGGGACTGGGCAAGGCACTGTCGGTGACATCACGGAC
AGGGCGACTTCTATGTAGATGAGGCAGCGCAGAGGCTGCTGCTTCGCCACTTGCT
GCTTCACCACGAAGGAGTTCCCGTGCCCTGGGAGCGGGTTCAGGACCGCTGATCG
GAAGTGAGAATCCCAGCTGTGTGTCAGGGCTGGAAAGGGCTCGGGAGTGCGCGG
GGCAAGTGACCGTGTGTGTAAAGAGTGAGGCGTATGAGGCTGTGTCGGGGCAGA
GGCCCAAGATCTCAGTTCACTGCCGTATAGGCAGGGCCCAGACTGAGCACGTGAG
TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGGACCGAGTCGGTCCTCTGCCATCAAAGCGTGCTCAGTCTGTTTCAGCAAG
TTCAGAGAAATCTGAACTTGCTGGATTTTTGGAGCAGGGAGATGGAATAGGAGCT
TGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTGC
ACCCACTTTCTGGAGTTTCAAAAGTAGACTGTACGCTAAGGGTCATATCTTTTTTT
GTTTGGTTTGTGTCTTGGTTGGCGTCTTAAA (SEQ ID NO: 151).

[0490] In various other embodiments, the PEgRNA may be improved by introducing modifications to the scaffold or core sequences. This can be done by introducing known The core, Cas9-binding PEgRNA scaffold can likely be improved to enhance PE activity. Several such approaches have already been demonstrated. For instance, the first pairing element of the scaffold (P1) contains a GTTTT-AAAAC (SEQ ID NO: 146) pairing element. Such runs of Ts have been shown to result in pol III pausing and premature termination of the RNA transcript. Rational mutation of one of the T-A pairs to a G-C pair in this portion of P1 has been shown to enhance sgRNA activity, suggesting this approach would also be feasible

for PEgRNAs[195]. Additionally, increasing the length of P1 has also been shown to enhance sgRNA folding and lead to improved activity[195], suggesting it as another avenue for the modification of PEgRNA activity. Example modifications to the core can include:

PEgRNA containing a 6 nt extension to P1

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGCTCATGAAAATGAGCTAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTC TGCCATCAAAGCGTGCTCAGTCTGTTTTTTT (SEQ ID NO: 152)

PEgRNA containing a T-A to G-C mutation within P1

GGCCCAGACTGAGCACGTGAGTTTGAGAGCTAGAAATAGCAAGTTTAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC GTGCTCAGTCTGTTTTTTT (SEQ ID NO: 153)

[0491] In various other embodiments, the PEgRNA may be modified at the edit template region. As the size of the insertion templated by the PEgRNA increases, it is more likely to be degraded by endonucleases, undergo spontaneous hydrolysis, or fold into secondary structures unable to be reverse-transcribed by the RT or that disrupt folding of the PEgRNA scaffold and subsequent Cas9-RT binding. Accordingly, it is likely that modification to the template of the PEgRNA might be necessary to affect large insertions, such as the insertion of whole genes. Some strategies to do so include the incorporation of modified nucleotides within a synthetic or semi-synthetic PEgRNA that render the RNA more resistant to degradation or hydrolysis or less likely to adopt inhibitory secondary structures[196]. Such modifications could include 8-aza-7-deazaguanosine, which would reduce RNA secondary structure in G-rich sequences; locked-nucleic acids (LNA) that reduce degradation and enhance certain kinds of RNA secondary structure; 2'-O-methyl, 2'-fluoro, or 2'-O-methoxyethoxy modifications that enhance RNA stability. Such modifications could also be included elsewhere in the PEgRNA to enhance stability and activity. Alternatively or additionally, the template of the PEgRNA could be designed such that it both encodes for a desired protein product and is also more likely to adopt simple secondary structures that are able to be unfolded by the RT. Such simple structures would act as a thermodynamic sink, making it less likely that more complicated structures that would prevent reverse transcription would occur. Finally, one could also split the template into two, separate PEgRNAs. In such a design, a PE would be used to initiate transcription and also recruit a separate template RNA to the targeted site via an RNA-binding protein fused to Cas9 or an RNA recognition element on the PEgRNA itself such as the MS2 aptamer. The RT could either directly bind to this separate template RNA, or initiate reverse transcription on the original PEgRNA before swapping to the second template. Such an approach could enable long insertions by both

preventing misfolding of the PEgRNA upon addition of the long template and also by not requiring dissociation of Cas9 from the genome for long insertions to occur, which could possibly be inhibiting PE-based long insertions.

[0492] In still other embodiments, the PEgRNA may be modified by introducing additional RNA motifs at the 5′ and 3′ termini of the PEgRNAs, or even at positions therein between (e.g., in the gRNA core region, or the spacer). Several such motifs - such as the PAN ENE from KSHV and the ENE from MALAT1 were discussed above as possible means to terminate expression of longer PEgRNAs from non-pol III promoters. These elements form RNA triple helices that engulf the polyA tail, resulting in their being retained within the nucleus[184, 187]. However, by forming complex structures at the 3′ terminus of the PEgRNA that occlude the terminal nucleotide, these structures would also likely help prevent exonuclease-mediated degradation of PEgRNAs.

[0493] Other structural elements inserted at the 3′ terminus could also enhance RNA stability, albeit without enabling termination from non-pol III promoters. Such motifs could include hairpins or RNA quadruplexes that would occlude the 3′ terminus[197], or self-cleaving ribozymes such as HDV that would result in the formation of a 2’-3′-cyclic phosphate at the 3′ terminus and also potentially render the PEgRNA less likely to be degraded by exonucleases[198]. Inducing the PEgRNA to cyclize via incomplete splicing - to form a ciRNA - could also increase PEgRNA stability and result in the PEgRNA being retained within the nucleus[194].

Additional RNA motifs could also improve RT processivity or enhance PEgRNA activity by enhancing RT binding to the DNA-RNA duplex. Addition of the native sequence bound by the RT in its cognate retroviral genome could enhance RT activity[199]. This could include the native primer binding site (PBS), polypurine tract (PPT), or kissing loops involved in retroviral genome dimerization and initiation of transcription[199].

[0494] Addition of dimerization motifs - such as kissing loops or a GNRA tetraloop/tetraloop receptor pair - at the 5′ and 3′ termini of the PEgRNA could also result in effective circularization of the PEgRNA, improving stability. Additionally, it is envisioned that addition of these motifs could enable the physical separation of the PEgRNA spacer and primer, prevention occlusion of the spacer which would hinder PE activity. Short 5′ extensions or 3′ extensions to the PEgRNA that form a small toehold hairpin in the spacer region or along the primer binding site could also compete favorably against the annealing of intracomplementary regions along the length of the PEgRNA, e.g., the interaction between the spacer and the primer binding site that can occur. Finally, kissing loops could also be used

to recruit other template RNAs to the genomic site and enable swapping of RT activity from one RNA to the other. A number of secondary RNA structures that may be engineered into any region of the PEgRNA, including in the terminal portions of the extension arm (i.e., e1and e2), as shown.

Example modifications include, but are not limited to:

[0495] PEgRNA-HDV fusion

GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCCTCTGCCATCAAAGC GTGCTCAGTCTGGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAA CATGCTTCGGCATGGCGAATGGGACTTTTTTT (SEQ ID NO: 154)

[0496] PEgRNA-MMLV kissing loop

GGTGGGAGACGTCCCACCGGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAA TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTC GGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACC TTTTTTT (SEQ ID NO: 155)

[0497] PEgRNA-VS ribozyme kissing loop

GAGCAGCATGGCGTCGCTGCTCACGGCCCAGACTGAGCACGTGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACC GAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTCCATCAGTTGACA CCCTGAGGTTTTTTT (SEQ ID NO: 156)

[0498] PEgRNA-GNRA tetraloop/tetraloop receptor

GCAGACCTAAGTGGUGACATATGGTCTGGGCCCAGACTGAGCACGTGAGTTTTAG AGCTAUACGTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTUACGAAGTGG GACCGAGTCGGTCCTCTGCCATCAAAGCTTCGACCGTGCTCAGTCTGCATGCGATT AGAAATAATCGCATGTTTTTTT (SEQ ID NO: 157)

[0499] PEgRNA template switching secondary RNA-HDV fusion

TCTGCCATCAAAGCTGCGACCGTGCTCAGTCTGGTGGGAGACGTCCCACCGGCCG GCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCG AATGGGACTTTTTTT (SEQ ID NO: 158)

[0500] PEgRNA scaffolds could be further improved via directed evolution, in an analogous fashion to how SpCas9 and prime editors (PE) have been improved. Directed evolution could enhance PEgRNA recognition by Cas9 or evolved Cas9 variants. Additionally, it is likely that different PEgRNA scaffold sequences would be optimal at different genomic loci, either enhancing PE activity at the site in question, reducing off-target activities, or both. Finally, evolution of PEgRNA scaffolds to which other RNA motifs have been added would almost certainly improve the activity of the fused PEgRNA relative to the unevolved, fusion RNA. For instance, evolution of allosteric ribozymes composed of c-di-GMP-I aptamers and hammerhead ribozymes led to dramatically improved activity[202], suggesting that evolution would improve the activity of hammerhead-PEgRNA fusions as well. In addition, while Cas9

currently does not generally tolerate 5′ extension of the sgRNA, directed evolution will likely generate enabling mutations that mitigate this intolerance, allowing additional RNA motifs to be utilized.

The present disclosure contemplates any such ways to further improve the efficacy of the prime editing systems utilized in the methods and compositions disclosed here.

[0501] In various embodiments, it may be advantageous to limit the appearance of consecutive sequence of Ts from the extension arm as consecutive series of T's may limit the capacity of the PEgRNA to be transcribed. For example, strings of at least consecutive three T's, at least consecutive four T's, at least consecutive five T's, at least consecutive six T's, at least consecutive seven T's, at least consecutive eight T's, at least consecutive nine T's, at least consecutive ten T's, at least consecutive eleven T's, at least consecutive twelve T's, at least consecutive thirteen T's , at least consecutive fourteen T's, or at least consecutive fifteen T's should be avoided when designing the PEgRNA, or should be at least removed from the final designed sequence. In one embodiment, one can avoid the includes of unwanted strings of consecutive T's in PEgRNA extension arms but avoiding target sites that are rich in consecutive A:T nucleobase pairs.

Kits, cells, vectors, and delivery

Kits

[0502] The compositions of the present disclosure may be assembled into kits. In some embodiments, the kit comprises nucleic acid vectors for the expression of a modified prime editor as described herein. In other embodiments, the kit further comprises appropriate guide nucleotide sequences (*e.g.*, PEgRNAs and second-site gRNAs) or nucleic acid vectors for the expression of such guide nucleotide sequences, to target the Cas9 protein or prime editor to the desired target sequence.

[0503] The kit described herein may include one or more containers housing components for performing the methods described herein and optionally instructions for use. Any of the kit described herein may further comprise components needed for performing the assay methods. Each component of the kits, where applicable, may be provided in liquid form (*e.g.*, in solution) or in solid form, (*e.g.*, a dry powder). In certain cases, some of the components may be reconstitutable or otherwise processible (*e.g.*, to an active form), for example, by the addition of a suitable solvent or other species (for example, water), which may or may not be provided with the kit.

[0504] In some embodiments, the kits may optionally include instructions and/or promotion for use of the components provided. As used herein, "instructions" can define a component of instruction and/or promotion, and typically involve written instructions on or associated with packaging of the disclosure. Instructions also can include any oral or electronic instructions provided in any manner such that a user will clearly recognize that the instructions are to be associated with the kit, for example, audiovisual (*e.g.*, videotape, DVD, *etc.*), Internet, and/or web-based communications, *etc.* The written instructions may be in a form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals or biological products, which can also reflect approval by the agency of manufacture, use or sale for animal administration. As used herein, "promoted" includes all methods of doing business including methods of education, hospital and other clinical instruction, scientific inquiry, drug discovery or development, academic research, pharmaceutical industry activity including pharmaceutical sales, and any advertising or other promotional activity including written, oral and electronic communication of any form, associated with the disclosure. Additionally, the kits may include other components depending on the specific application, as described herein.

[0505] The kits may contain any one or more of the components described herein in one or more containers. The components may be prepared sterilely, packaged in a syringe and shipped refrigerated. Alternatively it may be housed in a vial or other container for storage. A second container may have other components prepared sterilely. Alternatively the kits may include the active agents premixed and shipped in a vial, tube, or other container.

[0506] The kits may have a variety of forms, such as a blister pouch, a shrink wrapped pouch, a vacuum sealable pouch, a sealable thermoformed tray, or a similar pouch or tray form, with the accessories loosely packed within the pouch, one or more tubes, containers, a box or a bag. The kits may be sterilized after the accessories are added, thereby allowing the individual accessories in the container to be otherwise unwrapped. The kits can be sterilized using any appropriate sterilization techniques, such as radiation sterilization, heat sterilization, or other sterilization methods known in the art. The kits may also include other components, depending on the specific application, for example, containers, cell media, salts, buffers, reagents, syringes, needles, a fabric, such as gauze, for applying or removing a disinfecting agent, disposable gloves, a support for the agents prior to administration, *etc.* Some aspects of this disclosure provide kits comprising a nucleic acid construct comprising a nucleotide sequence encoding the various components of the prime editing system utilized in the methods and compositions described herein (e.g., including, but not limited to, the

napDNAbps, reverse transcriptases, polymerases, fusion proteins (e.g., comprising napDNAbps and reverse transcriptases (or more broadly, polymerases), extended guide RNAs, and complexes comprising fusion proteins and extended guide RNAs, as well as accessory elements, such as second strand nicking components (e.g., second strand nicking gRNA) and 5′ endogenous DNA flap removal endonucleases for helping to drive the prime editing process towards the edited product formation). In some embodiments, the nucleotide sequence(s) comprises a heterologous promoter (or more than a single promoter) that drives expression of the prime editing system components.

[0507] Other aspects of this disclosure provide kits comprising one or more nucleic acid constructs encoding the various components of the prime editing systems utilized in the methods and compositions described herein, e.g., the comprising a nucleotide sequence encoding the components of the prime editing system capable of modifying a target DNA sequence. In some embodiments, the nucleotide sequence comprises a heterologous promoter that drives expression of the prime editing system components.

[0508] Some aspects of this disclosure provides kits comprising a nucleic acid construct, comprising (a) a nucleotide sequence encoding a napDNAbp (*e.g.,* a Cas9 domain) fused to a reverse transcriptase and (b) a heterologous promoter that drives expression of the sequence of (a).

Cells

[0509] Cells that may contain any of the compositions described herein include prokaryotic cells and eukaryotic cells.

[0510] Mammalian cells of the present disclosure include human cells, primate cells (*e.g.*, vero cells), rat cells (*e.g.*, GH3 cells, OC23 cells) or mouse cells (*e.g.*, MC3T3 cells). There are a variety of human cell lines, including, without limitation, human embryonic kidney (HEK) cells, HeLa cells, cancer cells from the National Cancer Institute's 60 cancer cell lines (NCI60), DU145 (prostate cancer) cells, Lncap (prostate cancer) cells, MCF-7 (breast cancer) cells, MDA-MB-438 (breast cancer) cells, PC3 (prostate cancer) cells, T47D (breast cancer) cells, THP-1 (acute myeloid leukemia) cells, U87 (glioblastoma) cells, SHSY5Y human neuroblastoma cells (cloned from a myeloma) and Saos-2 (bone cancer) cells. In some embodiments, rAAV vectors are delivered into human embryonic kidney (HEK) cells (*e.g.*, HEK 293 or HEK 293T cells). In some embodiments, rAAV vectors are delivered into stem cells (*e.g.*, human stem cells) such as, for example, pluripotent stem cells (*e.g.*, human pluripotent stem cells including human induced pluripotent stem cells (hiPSCs)). A stem cell refers to a cell with the ability to divide for indefinite periods in culture and to give rise to

specialized cells. A pluripotent stem cell refers to a type of stem cell that is capable of differentiating into all tissues of an organism, but not alone capable of sustaining full organismal development. A human induced pluripotent stem cell refers to a somatic (*e.g.*, mature or adult) cell that has been reprogrammed to an embryonic stem cell-like state by being forced to express genes and factors important for maintaining the defining properties of embryonic stem cells (*see*, *e.g.*, Takahashi and Yamanaka, *Cell* 126 (4): 663–76, 2006, incorporated by reference herein). Human induced pluripotent stem cell cells express stem cell markers and are capable of generating cells characteristic of all three germ layers (ectoderm, endoderm, mesoderm).

[0511] Additional non-limiting examples of cell lines that may be used in accordance with the present disclosure include 293-T, 293-T, 3T3, 4T1, 721, 9L, A-549, A172, A20, A253, A2780, A2780ADR, A2780cis, A431, ALC, B16, B35, BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C2C12, C3H-10T1/2, C6, C6/36, Cal-27, CGR8, CHO, CML T1, CMT, COR-L23, COR-L23/5010, COR-L23/CPR, COR-L23/R23, COS-7, COV-434, CT26, D17, DH82, DU145, DuCaP, E14Tg2a, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, Hepa1c1c7, High Five cells, HL-60, HMEC, HT-29, HUVEC, J558L cells, Jurkat, JY cells, K562 cells, KCL22, KG1, Ku812, KYO1, LNCap, Ma-Mel 1, 2, 3....48, MC-38, MCF-10A, MCF-7, MDA-MB-231, MDA-MB-435, MDA-MB-468, MDCK II, MG63, MONO-MAC 6, MOR/0.2R, MRC5, MTD-1A, MyEnd, NALM-1, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NW-145, OPCN/OPCT Peer, PNT-1A/PNT 2, PTK2, Raji, RBL cells, RenCa, RIN-5F, RMA/RMAS, S2, Saos-2 cells, Sf21, Sf9, SiHa, SKBR3, SKOV-3, T-47D, T2, T84, THP1, U373, U87, U937, VCaP, WM39, WT-49, X63, YAC-1 and YAR cells.

[0512] Some aspects of this disclosure provide cells comprising any of the constructs disclosed herein. In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5,

MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A 172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293. BxPC3. C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr –/–, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK 11, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof.

[0513] Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

Vectors

[0514] Some aspects of the present disclosure relate to using recombinant virus vectors (e.g., adeno-associated virus vectors, adenovirus vectors, or herpes simplex virus vectors) for the delivery of the modified prime editors as described herein into a cell. In the case of a split-PE approach, the N-terminal portion of a PE fusion protein and the C-terminal portion of a PE fusion are delivered by separate recombinant virus vectors (e.g., adeno-associated virus vectors, adenovirus vectors, or herpes simplex virus vectors) into the same cell, since the full-

length Cas9 protein or prime editors exceeds the packaging limit of various virus vectors, e.g., rAAV (~4.9 kb).

[0515] In some embodiments, the vectors used herein may encode the PE fusion proteins, or any of the components thereof (e.g., napDNAbp, linkers, or polymerases). In addition, the vectors used herein may encode the PEgRNAs, and/or the accessory gRNA for second strand nicking. The vectors may be capable of driving expression of one or more coding sequences in a cell. In some embodiments, the cell may be a prokaryotic cell, such as, e.g., a bacterial cell. In some embodiments, the cell may be a eukaryotic cell, such as, e.g., a yeast, plant, insect, or mammalian cell. In some embodiments, the eukaryotic cell may be a mammalian cell. In some embodiments, the eukaryotic cell may be a rodent cell. In some embodiments, the eukaryotic cell may be a human cell. Suitable promoters to drive expression in different types of cells are known in the art. In some embodiments, the promoter may be wild-type. In other embodiments, the promoter may be modified for more efficient or efficacious expression. In yet other embodiments, the promoter may be truncated yet retain its function. For example, the promoter may have a normal size or a reduced size that is suitable for proper packaging of the vector into a virus.

[0516] In some embodiments, the promoters that may be used in the prime editor vectors may be constitutive, inducible, or tissue-specific. In some embodiments, the promoters may be a constitutive promoters. Non-limiting exemplary constitutive promoters include cytomegalovirus immediate early promoter (CMV), simian virus (SV40) promoter, adenovirus major late (MLP) promoter, Rous sarcoma virus (RSV) promoter, mouse mammary tumor virus (MMTV) promoter, phosphoglycerate kinase (PGK) promoter, elongation factor-alpha (EFla) promoter, ubiquitin promoters, actin promoters, tubulin promoters, immunoglobulin promoters, a functional fragment thereof, or a combination of any of the foregoing. In some embodiments, the promoter may be a CMV promoter. In some embodiments, the promoter may be a truncated CMV promoter. In other embodiments, the promoter may be an EFla promoter. In some embodiments, the promoter may be an inducible promoter. Non-limiting exemplary inducible promoters include those inducible by heat shock, light, chemicals, peptides, metals, steroids, antibiotics, or alcohol. In some embodiments, the inducible promoter may be one that has a low basal (non-induced) expression level, such as, e.g., the Tet-On® promoter (Clontech). In some embodiments, the promoter may be a tissue-specific promoter. In some embodiments, the tissue-specific promoter is exclusively or predominantly expressed in liver tissue. Non-limiting exemplary tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45

promoter, CD68 promoter, desmin promoter, elastase- 1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM- 2 promoter, INF-β promoter, Mb promoter, Nphsl promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter.

[0517] In some embodiments, the prime editor vectors (e.g., including any vectors encoding the prime editor systems and/or fusion protein and/or the PEgRNAs, and/or the accessory second strand nicking gRNAs) may comprise inducible promoters to start expression only after it is delivered to a target cell. Non-limiting exemplary inducible promoters include those inducible by heat shock, light, chemicals, peptides, metals, steroids, antibiotics, or alcohol. In some embodiments, the inducible promoter may be one that has a low basal (non-induced) expression level, such as, e.g., the Tet-On® promoter (Clontech).

[0518] In additional embodiments, the prime editor vectors (e.g., including any vectors encoding the prime editors and/or prime editor fusion protein and/or the PEgRNAs, and/or the accessory second strand nicking gRNAs) may comprise tissue- specific promoters to start expression only after it is delivered into a specific tissue. Non-limiting exemplary tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45 promoter, CD68 promoter, desmin promoter, elastase- 1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM- 2 promoter, INF-β promoter, Mb promoter, Nphsl promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter.

[0519] In some embodiments, the nucleotide sequence encoding the PEgRNA (or any guide RNAs used in connection with prime editing) may be operably linked to at least one transcriptional or translational control sequence. In some embodiments, the nucleotide sequence encoding the guide RNA may be operably linked to at least one promoter. In some embodiments, the promoter may be recognized by RNA polymerase III (Pol III). Non-limiting examples of Pol III promoters include U6, HI and tRNA promoters. In some embodiments, the nucleotide sequence encoding the guide RNA may be operably linked to a mouse or human U6 promoter. In other embodiments, the nucleotide sequence encoding the guide RNA may be operably linked to a mouse or human HI promoter. In some embodiments, the nucleotide sequence encoding the guide RNA may be operably linked to a mouse or human tRNA promoter. In embodiments with more than one guide RNA, the promoters used to drive expression may be the same or different. In some embodiments, the nucleotide encoding the crRNA of the guide RNA and the nucleotide encoding the tracr RNA of the guide RNA may be provided on the same vector. In some embodiments, the nucleotide

encoding the crRNA and the nucleotide encoding the tracr RNA may be driven by the same promoter. In some embodiments, the crRNA and tracr RNA may be transcribed into a single transcript. For example, the crRNA and tracr RNA may be processed from the single transcript to form a double-molecule guide RNA. Alternatively, the crRNA and tracr RNA may be transcribed into a single-molecule guide RNA.

[0520] In some embodiments, the nucleotide sequence encoding the guide RNA may be located on the same vector comprising the nucleotide sequence encoding the PE fusion protein. In some embodiments, expression of the guide RNA and of the PE fusion protein may be driven by their corresponding promoters. In some embodiments, expression of the guide RNA may be driven by the same promoter that drives expression of the PE fusion protein. In some embodiments, the guide RNA and the PE fusion protein transcript may be contained within a single transcript. For example, the guide RNA may be within an untranslated region (UTR) of the Cas9 protein transcript. In some embodiments, the guide RNA may be within the 5´ UTR of the PE fusion protein transcript. In other embodiments, the guide RNA may be within the 3' UTR of the PE fusion protein transcript. In some embodiments, the intracellular half-life of the PE fusion protein transcript may be reduced by containing the guide RNA within its 3' UTR and thereby shortening the length of its 3' UTR. In additional embodiments, the guide RNA may be within an intron of the PE fusion protein transcript. In some embodiments, suitable splice sites may be added at the intron within which the guide RNA is located such that the guide RNA is properly spliced out of the transcript. In some embodiments, expression of the Cas9 protein and the guide RNA in close proximity on the same vector may facilitate more efficient formation of the CRISPR complex.

[0521] The vector system may comprise one vector, or two vectors, or three vectors, or four vectors, or five vector, or more. In some embodiments, the vector system may comprise one single vector, which encodes both the PE fusion protein, the PEgRNA. In other embodiments, the vector system may comprise two vectors, wherein one vector encodes the PE fusion protein and the other encodes the PEgRNA.

[0522] Some examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository

waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum component, such as serum albumin, HDL and LDL; (22) C2-C12 alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation. The terms such as "excipient", "carrier", "pharmaceutically acceptable carrier" or the like are used interchangeably herein.

Delivery methods

[0523] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a prime editor as described herein in combination with (and optionally complexed with) a guide sequence are delivered to a cell. In any of the delivery methods described herein can also be delivered along with the prime editor. In some embodiments, the inhibitor is encoded on the same vector as the prime editor. In certain embodiments, the inhibitor is fused to the prime editor. In some embodiments, the inhibitor is encoded on a second vector, which is delivered along with a vector encoding the prime editor. In some embodiments, the prime editor is delivered to a cell as proteins directly. In certain embodiments, the fusion protein is delivered directly into a cell.

[0524] Exemplary delivery strategies include vector-based strategies, PE ribonucleoprotein complex delivery, and delivery of PE by mRNA methods. In some embodiments, the method of delivery provided comprises nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA.

[0525] Exemplary methods of delivery of nucleic acids include lipofection, nucleofection, electroporation, stable genome integration (e.g., piggybac), microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™, Lipofectin™ and SF Cell Line 4D-Nucleofector X Kit™ (Lonza)). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424; WO 91/16024. Delivery may be to cells (*e.g.*, *in vitro* or *ex vivo* administration) or target tissues (*e.g.*, *in vivo* administration). Delivery may be achieved through the use of RNP complexes.

[0526] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, *e.g.*, Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0527] In other embodiments, the method of delivery and vector provided herein is an RNP complex. RNP delivery of fusion proteins markedly increases the DNA specificity of prime editing. RNP delivery of fusion proteins leads to decoupling of on- and off-target DNA editing. RNP delivery ablates off-target editing at non-repetitive sites while maintaining on-target editing comparable to plasmid delivery, and greatly reduces off-target DNA editing even at the highly repetitive VEGFA site 2. See Rees, H.A. et al., Improving the DNA specificity and applicability of prime editing through protein engineering and protein delivery, Nat. Commun. 8, 15790 (2017), U.S. Patent No. 9,526,784, issued December 27, 2016, and U.S. Patent No. 9,737,604, issued August 22, 2017, each of which is incorporated by reference herein.

[0528] Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US 2003/0087817, incorporated herein by reference.

[0529] Other aspects of the present disclosure provide methods of delivering the prime editor constructs into a cell to form a complete and functional prime editor within a cell. For example, in some embodiments, a cell is contacted with a composition described herein (*e.g.*, compositions comprising nucleotide sequences encoding the split Cas9 or the split prime editor or AAV particles containing nucleic acid vectors comprising such nucleotide

sequences). In some embodiments, the contacting results in the delivery of such nucleotide sequences into a cell, wherein the N-terminal portion of the Cas9 protein or the prime editor and the C-terminal portion of the Cas9 protein or the prime editor are expressed in the cell and are joined to form a complete Cas9 protein or a complete prime editor.

It should be appreciated that any rAAV particle, nucleic acid molecule or composition provided herein may be introduced into the cell in any suitable way, either stably or transiently. In some embodiments, the disclosed proteins may be transfected into the cell. In some embodiments, the cell may be transduced or transfected with a nucleic acid molecule. For example, a cell may be transduced (*e.g.*, with a virus encoding a split protein), or transfected (*e.g.*, with a plasmid encoding a split protein) with a nucleic acid molecule that encodes a split protein, or an rAAV particle containing a viral genome encoding one or more nucleic acid molecules. Such transduction may be a stable or transient transduction. In some embodiments, cells expressing a split protein or containing a split protein may be transduced or transfected with one or more guide RNA sequences, for example in delivery of a split Cas9 (e.g., nCas9) protein. In some embodiments, a plasmid expressing a split protein may be introduced into cells through electroporation, transient (*e.g.*, lipofection) and stable genome integration (*e.g.*, piggybac) and viral transduction or other methods known to those of skill in the art.

EXAMPLES

Example 1: Development of modified PE2 prime editor referred to as PEmax

[0530] To further improve prime editing, the PE2 protein was optimized by varying reverse transcriptase (RT) codon usages, the length and composition of the peptide linkers between nCas9 and the reverse transcriptase, the location, composition, and number of NLS sequences, and mutations within the SpCas9 domain (FIGs. 8A and 8B). Among 20 such variants tested, the greatest enhancement in editing efficiency was observed with a prime editor architecture that uses a Genscript human codon-optimized RT, a 34-aa linker containing a bipartite SV40 NLS (Wu *et al.*, 2009), an additional C-terminal c-Myc NLS (Dang and Lee, 1988), and R221K and N394K mutations in SpCas9 previously shown to improve Cas9 nuclease activity (Spencer and Zhang, 2017) (FIGs. 9 and 8A). This optimized prime editor architecture was designated as PEmax. Across seven substitution edits targeting different loci, using the PEmax architecture with the PE2 system (PE2max) increased the average frequency of intended editing by 2.3-fold in HeLa cells and 1.2-fold in HEK293T

cells over the original PE2 architecture (FIG. 9B). Similarly, PE3 using the PEmax architecture (PE3max) increased average editing efficiencies over PE3 by 3.2-fold in HeLa cells and 1.2-fold in HEK293T cells, without substantially changing product purity (FIGs. 9 and 8A).

Example 2: Engineering and Evolution of Novel and Enhanced Prime Editors

Background

[0531] Prime editing is a recently developed genome editing technology that enables the programmable installation of SNPs, insertions, and deletions into living cells. Prime editors are composed of a Cas9 (H840A) nickase fused to a reverse transcriptase (RT) enzyme: upon nicking of the genome by Cas9, the fused RT can use a 3′-extended sgRNA called a pegRNA to reverse transcribe a DNA sequence onto the end of the nicked genome. These newly synthesized bases are incorporated into the genome, leading to permanent editing. The two original versions of the prime editor are PE1 and PE2[1]. PE1 (SEQ ID NO: 3) utilizes the wild-type (WT) Moloney murine leukemia virus (M-MLV) RT; and PE2 (SEQ ID NO: 4) utilizes an engineered pentamutant of M-MLV RT (MMLV_RT with D200N, T330P, L603W, T306K, and W313F substitutions) relative to SEQ ID NO: 33) that increases editing efficiency across a wide variety of sites in human cells.

[0532] As illustrated in FIG. 27A, this Example provides engineered and PACE[2]-evolved RT variants for prime editing. Thus far, the only RT enzyme that has been utilized for prime editing in mammalian cells is M-MLV RT. M-MLV RT is a large enzyme (2.2 kB), which poses barriers for many in vivo delivery methods such as Adeno-associated Viruses (AAVs). Since RT enzymes vary widely in their size and enzymatic activity, the alternate enzymes reported here provide unique advantages for prime editing (smaller size or improved editing). In addition, this Example provides mutants of Cas9 that increase prime editing efficiency in mammalian cells. These improvements lead to prime editors that are more efficient and more easily delivered for therapeutic applications.

Approach and Results

Screening retroviral RTs for PE

[0533] The inventors hypothesized that other RT enzymes could be used instead of the M-MLV RT to either improve editing efficiencies or to decrease the size the editor. Since the M-MLV RT comes from retroviruses, the inventors identified and tested the activity of various retroviral RT enzymes in mammalian cells. Twelve (12) retroviral RTs other than M-MLV

RT were identified that exhibited activity in HEK293T cells at 2 loci (*FANCF* and *HEK3*) (FIG. 1). MMTV[3], ASLV (alpha subunit)[4], PERV[5] and HIV_MMLV[6] were identified from the literature; AVIRE, BAEMV, GALV, KORV, MPMV, POK11ERV, SRV2 and WMSV came from the UniProt database using the BLAST-P algorithm. MMTV-RT[3], PERV-RT[5], AVIRE-RT, KORV-RT and WMSV-RT had higher editing than WT M-MLV. The amino acid sequences for these alternative RTs are provided below.

Engineering retroviral RTs for improved performance

[0534] During the development of prime editors, the WT M-MLV RT enzyme was further engineered for improved activity by incorporating 5 mutations (D200N, T306K, W313F, E330P and L603W) into the enzyme to generate PE2[1]. Since PERV-RT, AVIRE-RT, KORV-RT and WMSV-RT are highly homologous to M-MLV RT (68%, 57%, 67%, 68% similar in sequence respectively), it was hypothesized that analogous mutations (i.e., mutations corresponding to D200N, T306K, W313F, E330P and L603W of M-MLV RT in PE2) could be incorporated into these RT enzymes for improved performance. On average, for all 4 RT enzymes, incorporation of each mutation increased prime editing outcome compared to WT at 4 different loci (*HEK3*, *EMX1*, *FANCF, RNF2*) (see FIG. 29).

[0535] Since all 5 individual analogous mutations improved prime editing activity, we generated a penta-mutant variant of pRT21.6 (PERV with D199N+T305K+W312F+E329P+ L603W substitutions). This variant was ~6.6x better than the WT enzyme across 9 different edits tested (FIG. 30). However, prime editing activity of pRT21.6 was on average modestly lower than PE2 (see FIG. 18).

Yeast retrotransposon Tf1 RT for PE

[0536] Next, the inventors focused on screening and engineering smaller RT enzymes to make PEs more amenable for *in vivo* delivery. From an initial screen, an RT enzyme from the yeast retrotransposon, Tf1, was identified that is 0.5 kB smaller than M-MLV RT[7]. Tf1 had significantly higher editing in mammalian cells compared to the WT M-MLV RT (PE1) but lower editing than PE2 at 3 sites tested in HEK293T cells (see FIG. 19).

Structure-guided engineering of Tf1 to improve PE

[0537] The inventors further aimed to engineer Tf1 RT to improve its performance. Tf1 belongs to the Ty3/Gypsy family of retrotransposons. Using a three-dimensional protein structure of a Ty3 reverse transcriptase bound to its RNA-DNA substrate[8] (PDB: 4OL8), a series of mutations were designed that were predicted to increase interaction of Tf1 RT with its substrates. Two mutations, K118R and S297Q, improved prime editing activity compared

to the WT enzyme (see FIG. 20). A Tf1 double mutant (K118R + S297Q) mutant further improved editing compared to the single mutants across the 5 sites tested in HEK293T cells.

[0538] Without being bound by theory, the two mutations, K118R and S297Q, were predicted to increase interaction with the RNA and DNA substrate, respectively.

Creation and Validation of a PE-PACE Circuit

[0539] Next, a PE-PACE circuit was developed to more quickly select for PE-enhancing mutations in many different RTs. Reference is made to PACE circuit design to evolve cytosine and adenine base editors[9,10]. As a first step to designing the circuit, the gIII was removed from the M13 bacteriophage genome and was placed under the control of a T7 promoter on a plasmid in host *E. coli*. A second plasmid was prepared which encoded T7 RNA polymerase (T7 RNAP) with a 1-bp deletion, which frameshifts and inactivates T7 RNAP. Correction of this frameshift by a successful prime edit would thus enable WT T7 RNAP production, which can then drive gIII transcription and phage propagation. In the initial iteration of the PE-PACE circuit, the various components of the prime editor protein were distributed between the host *E. coli* and the selection phage. A pegRNA encoding the desired T7 edit was included on the gIII plasmid, and the protein component of the editor was split between the host and phage. SpCas9(H840A) fused to an N-terminal Npu intein was included in a third and final plasmid in the host *E. coli*. The PE2 reverse transcriptase was placed on the phage genome fused to a C terminal Npu intein. Following phage infection, intein splicing reconstitutes full length prime editor. A schematic for this circuit is shown in FIG. 10.

[0540] The circuit was evaluated by overnight propagation assays. PE2 RT phage propagation exceeded that of an empty phage negative control, which strongly de-enriched; however, overnight propagation levels of the PE2 RT phage were not as robust as expected (FIG. 31A). Because prime editing efficiency in mammalian cells is heavily influenced by the PBS and RT template length of the pegRNA, we speculated that pegRNA optimization would also be important for our PACE circuit. Therefore, to enhance prime editing and in turn PE2 RT phage propagation, we tested a matrix of PBS and RT template lengths for a total of 36 pegRNAs. Strikingly, propagation of PE2 RT phage varied 10,000-fold depending on the pegRNA used (FIG. 31B). This result not only underscores the importance of pegRNA optimization, but also enabled robust phage propagation of ~100 fold in overnight propagation.

[0541] To confirm that phage propagation in our PE-PACE circuit was correlative with reverse transcriptase activity, we evaluated phage propagation using phage encoding the WT M-MLV reverse transcriptase. The reverse transcriptase used in PE2 consists of a mutant M-MLV reverse transcriptase harboring five mutations from the literature: (D200N, T306K, 313, 330, 603). The prime editor PE1, which uses the WT M-MLV reverse transcriptase, is much less efficient than PE2 when measuring prime editing in mammalian cells. For this reason, PE1 was a valuable tool to ensure that activity in our PACE circuit tracked with mammalian editing. PE1 phage propagated ~2,600-fold less than PE2 phage, showing that reverse transcriptases that are more active mammalian prime editors propagate better in the PACE circuit (FIG. 31C). Finally, to complete circuit validation, we evolved PE1 RT phage using phage-assisted noncontinuous evolution (PANCE). Encouragingly, after 12 rounds of selection, PE1 phage began to robustly propagate in PANCE (FIG. 31D).

[0542] Sequencing of these phage revealed the convergence of several mutants (FIG. 32). Two of the six mutations that converged in PANCE were mutations found in PE2. This demonstrated that PANCE could select for mutations known to improve prime editing activity and validated this novel PACE circuit. The other mutations found (D200Y, V223A, V223M, E302A, E302K, M457I, and A462S) are not in PE2; with the exception of E302K, they were not tested in the original report of prime editing.

Modifications to the PE-PACE Circuit

[0543] Several modifications were also made to the PE-PACE circuit. First, circuit stringency was tuned by modulating the expression of the T7 RNAP: the weaker the promoter and RBS of T7 RNAP, the higher the circuit stringency (FIG. 33A). Unlike previous base editing circuits, though, it was also possible to manipulate the circuit by changing the edit required for circuit turn-on. For example, in the above PE-PACE circuit, the desired prime edit was a 1 bp insertion. By changing the desired prime edit to a 20 bp insertion, the properties of the selection could be changed. In particular, this change was predicted to select for RTs with higher processivity (FIG. 33B). These changes to the circuit were incorporated into several of the evolutions below.

Directed evolution of Tf1 RT using PACE

[0544] Although the double mutant of Tf1 showed significant improvement compared to the WT enzyme, the editing of PEs with Tf1 RT was still lower than PE2. Thus, it was decided to utilize the PACE circuit described above to improve Tf1 further. Using the 1-bp deletion and 20-bp deletion circuit, the following variants were generated:

- 5.27-(V14A+L158Q+F269L+K356E) (SEQ ID NO: 197)
- 5.59-(E22K+P70T+G72V+M102I+K106R+A139T+L158Q+F269L+A363V+
- K413E+S492N) (SEQ ID NO: 199), and
- 5.60-(P70T+G72V+M102I+K106R+L158Q+F269L+A363V+K413E+S492N) (SEQ ID NO: 200),

[0545] Variants 5.60, 5.27, and 5.59 showed improved editing compared to the WT Tf1 RT enzyme. Variants 5.59 and 5.60 have comparable editing to PE2 at 5 sites tested in HEK293T cells. (See FIG. 34)

Screening other small bacterial RT enzymes for PE

[0546] Next, it was decided to screen for even smaller enzymes for PE. Seven additional RT enzymes were identified that exhibited activity in HEK293T cells at two different loci (*RNF2* and *HEK3*). The seven enzymes are CRISPR_RT, Vp96, Vc95, Ec48, Gs, Er, and Ne144, the amino acid sequences of which are provided below. All seven RT enzymes are smaller than M-MLV RT (667 amino acids long) (FIG. 24). Vp96, Vc95, Ec48 and Ne144 are bacterial retron RTs whose function have been experimentally validated[11]. The Er RT is a highly processive metazoan group II intron RT[12], whereas the CRISPR-RT was one of the smallest RT enzymes characterized by *Toro, et al.* during the phylogenetic analysis of bacterial reverse transcriptase enzymes[13]. These enzymes were further evolved as follows.

Evolution of retron Ec48 RT

[0547] Ec48 is a small bacterial RT enzyme (~0.8 kB smaller than M-MLV RT) that has low starting activity (FIG. 35). Using the 1-bp deletion and 20-bp deletion circuits, we generated variants:

- 3.8-(R267I+K318E+K326E+E328K+R372K) (SEQ ID NO: 195) (Ec48-evo1)
- 3.35-(E54K+K87E+D243N+R267I+E279K+K318E) (SEQ ID NO: 189) (Ec48-evo2)
- 3.36-(A36V+K87E+R205K+D243N+R267I+E279K+K318E) (SEQ ID NO: 190)
- 3.38-(E54K+K87E+D243N+R267I+S277F+E279K+K318E) (SEQ ID NO: 192).

[0548] These variants all show improved activity over the WT Ec48 enzymes (FIG. 24).

Evolution of retron Ne144 RT

[0549] Ne144 is another small bacterial RT enzyme (~ 0.5 kB smaller than M-MLV RT) that has very low starting activity (FIG. 35). The 20-bp deletion circuit was used to generate 38.14 Ne144 variant (A157T+A165T+G288V) (SEQ ID NO: 240) that is on average 23x fold better than the WT enzyme across 4 loci (FIG. 36).

Evolution of retron Vc95

[0550] Vc95 is another small bacterial RT enzyme (~ 1.1 kB smaller than M-MLV RT) that has very low starting activity (FIG. 35). The 1-bp deletion circuit was used to generate

[0551] 25.8 Vc95 variant (L11M+S75A+V97M+N146D+N245T) (SEQ ID NO: 242) that is on average 7-fold better than the WT enzyme across 4 loci (FIG. 37).

Evolution of a reverse transcriptase from Geobacillus stearothermophilus

[0552] In addition to the RTs included in the initial screen in FIG. 35, an additional final RT was evolved using the group II intron reverse transcriptase from the thermophilic organism, *Geobacillus stearothermophilus* (Gs RT)[14]. This RT is ~800 bp smaller than the M-MLV RT, but exhibited low WT activity in mammalian cell prime editing initially. Following rounds of PANCE (FIG. 38A) and PACE (FIG. 38B) in circuits with increasing stringency, mutants showed drastically improved prime editing activity in mammalian cells when compared to the WT enzyme.

Evolution of Sp Cas9 Variants for Prime Editing

[0553] One additional version of the circuit that has been made is to encode the entire prime editor protein, (both the Cas9 nickase and the M-MLV reverse transcriptase as shown in FIG. 13) on the phage, as opposed to all other efforts, in which only the RT was evolved. Like earlier iterations of the PE-PACE circuit, stringency can be tuned via T7 expression and examine multiple different edits. After increasingly stringent rounds of PANCE and then PACE on both the 1bp selection and the 20 bp selection, many convergent mutations in the Cas9 domain of the prime editor were found. Only a subset of these mutations, though, were helpful for mammalian cell prime editing: those mutants' mammalian activity are shown in FIG. 39.

Discussion

[0554] In this Example, a suite of reverse transcriptases have been engineered and evolved which are capable of efficient prime editing in mammalian cells.

[0555] These engineered and evolved variants exhibit drastically increased prime editing activity relative to their wild-type counterparts. The variants described here also offer unique benefits when compared to the original M-MLV mutant RT described in PE2.

[0556] Firstly, many of the RTs described here are significantly smaller than the M-MLV RT. This will be critical for eventual delivery applications, where size of the editor protein is limiting (for example, both AAV delivery and lentiviral delivery of the entire full-length editor are currently impossible due to the prime editor's large size).

[0557] In addition to decreased editor size, many of these RTs are beneficial in that, unlike M-MLV, they are not derived from mammalian viruses. This is important for downstream applications because (1) some mice used for research are known to have anti-M-MLV antibodies, and (2) M-MLV and its close structural relatives are known to interact with mammalian proteins. To minimize these unintended interactions, bacterial-derived RTs will be uniquely enabling.

[0558] In this Example, the Cas9 domain of the prime editor has also been evolved to produce useful variants. Mutations that affect interactions between the Cas9 protein and its guide RNA seem to give a slight benefit to mammalian cell prime editing, likely due to the unique nature of the pegRNA. Enhancing the Cas9 domain of the prime editor will also be crucial for achieving the high-efficiency prime editing needed for therapeutic applications of the technology.

Protein sequences of RTs tested:

[0559] MMTV-RT:

VFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWP
LKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNAT
MHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPY
QRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIV
DEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLN
DFQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLST
ARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIK
GRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTF
TLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAV
ITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGH
IRGHTGLPGPLAQGNAYADSLTRILT (SEQ ID NO: 43)

[0560] ASLV-RT:

[0561] TVALHLAIPLKWKPDHTPVWIDQWPLPEGKLVALTQLVEKELQLGHIEPSLSC
WNTPVFVIRKASGSYRLLHDLRAVNAKLVPFGAVQQGAPVLSALPRGWPLMVLDL
KDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSPTICQLVVGQVL
EPLRLKHPSLRMLHYMDDLLLAASSHDGLEAAGEEVISTLERAGFTISPDKIQREPGV
QYLGYKLGSTYVAPVGLVAEPRIATLWDVQKLVGSLQWLRPALGIPPRLMGPFYEQ
LRGSDPNEAREWNLDMKMAWREIVQLSTTAALERWDPALPLEGAVARCEQGAIGV
LGQGLSTHPRPCLWLFSTQPTKAFTAWLEVLTLLITKLRASAVRTFGKEVDILLLPAC

FREDLPLPEGILLALKGFAGKIRSSDTPSIFDIARPLHVSLKVRVTDHPVPGPTVFTDAS SSTHKGVVVWREGPRWEIKEIADSGASVQQLEARAVAMALLLWPTTPTNVVTDSAF VAKMLLKMGQEGVPSTAAAFILEDALSQRSAMAAVLHVRSHSEVPGFFTEGNDVAD SQATFQAY (SEQ ID NO: 44)

[0562] PERV-RT:

[0563] TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQL KASATPVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY RPVQDLREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQP LFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYV DDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRW LTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSW APEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPV AYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPD RWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAE LMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSL LEALHLPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 45)

[0564] HIV_MMLV:

[0565] PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPEN PYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLD VGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILE PFKKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPP FLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKL LRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQ IYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKET WETWWTEYWQATWIPEWEFVNTPPLVKLVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNK DEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI EN (SEQ ID NO: 46)

[0566] AVIRE-RT:

[0567] APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLST ALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQ

DLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEW
ADAEEGESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIA
ADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRT
QAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEE
EAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKR
LDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNAR
ITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPL
AQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALE
WSKDKSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLP
KRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQ
YSNVEEALG (SEQ ID NO: 216)

[0568] BAEMV-RT:

[0569] VSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTA
VPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQD
LREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEW
KDPERGISGQLTWTRLPQGFKNSPTLFDEALHRDLTDFRTQHPEVTLLQYVDDLLLA
APTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIE
TVARIPPPRNPREVREFLGTAGFCRLWIPGFAELAAPLYALTKESTPFTWQTEHQLAF
EALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPVAYLSKKLDP
VAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLT
HYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQEL
PDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKAL
ELSKGKKANIYTDSRYAFATAHTHGSIYERRGLLTSEGKEIKNKAEIIALLKALFLPQE
VAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHITIEHTYTSE
DQEEA (SEQ ID NO: 48)

[0570] GALV-RT:

[0571] LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSG
ASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLLPVKKPGTNDYRPV
QDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAF
EWKDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDL
LVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLT
PARKATVMKIPVPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEH
QQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMT

NARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLED
QPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQA
LRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPR
RVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKPQEPIEPAQEK
(SEQ ID NO: 49)

[0572] KORV-RT:

[0573] MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKS
DASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRP
VQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLF
AFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYV
DDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKR
WLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFT
WTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRP
VAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPP
DRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRP
DLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELI
ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLE
AIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTR
GK (SEQ ID NO: 222)

[0574] MPMV-RT:

[0575] MWGRDLLSQMKIMMCSPNDIVTAQMLAQGYSPGKGLGKKENGILHPIPNQG
QSNKKGFGNFLTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQE
QLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAI
PQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTL
CQKYVATAIHKVRHAWKQMYIIHYMDDILIAGKDGQQVLQCFDQLKQELTAAGLHI
APEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKL
TTGDLKPLFDTLKGDSDPNSHRSLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNT
ALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGRDHSKKYFGIEPSTIIQPY
SKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNA
LLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSA
YLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQR
ADLATKIVA (SEQ ID NO: 51)

[0576] POL11ERV-RT:

[0577] ATVEPPKPIPLTWKTEKPVWVNQWPLPKQKLEALHLLANEQLEKGHIEPSFSP WNSPVFVIQKKSGKWRMLTDLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKD CFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPV REKFSDCYIIHYIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLG MQIENRKIKPQKIEIRKDTLKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSD LNSKRILTPEATKEIKLVEEKIQSAQINRIDPLAPLQLLIFATAHSPTGIIIQNTDLVEWS FLPHSTVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDKIVVPLTKEQVRQAFINSGAWQ IGLANFVGIIDNHYPKTKIFQFLKMTTWILPKITRREPLENALTVFTDGSSNGKAAYTG PKERVIKTPYQSAQRAELVAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMD DQLNQLFNLLQQTVRKRNFPFYITHIRAHTNLPGPLTKANEEADLLVS (SEQ ID NO: 52)

[0578] SRV2-RT:

[0579] MWGRDLLSQMKIMMCSPNDIVTAQMLAQGYSPGKGLGKREDGILQPIPNSG QLDRKGFGNFLATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQ EQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVA IPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSP TLCQKYVAAAIEPVRKSWAQMYIIHYMDDILIAGKLGEQVLQCFAQLKQALTTTGL QIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYL HLTTGDLKPLFDILKGDSNPNSPRSLSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIF NTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILGRDNSKKYFGLEPSTI IQPYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPL DNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVLSAFPHRALNVYT DSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQG NHITDLATKVVA (SEQ ID NO: 53)

[0580] WMSV-RT:

[0581] LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSG ASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPV QDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAF EWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDL LVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLT PARKATVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEH QKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSK KLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMT NARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKD

QPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQ ALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLP KRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 228)

[0582] Tf1-RT:

[0583] ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVEL TQENYRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLR MVVDYKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLA FRCPRGVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHV KHVKDVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQP KNRKELRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCL VSPPVLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNY SVSDKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFL QDFNFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 55)

[0584] CRISPR -RT:

[0585] NSQAQSACCAGANQIVEGATLEKVVAPACLQQAWTRVRKNKGGPGGDGVTI EIFAQNAEVELEKLRAETLAGIYRPRKVRHAIVPKPKGGERKLTIPSVVDRILQTATM LSLGQTVDHHFSSASWAYREGRGVDDALADLRRLRNSGLFWTFDADIMQYFDRILH KRLIDDLFIWVDDLRIVRLIQLWLRSFSYWGRGIAQGAPISPLLANLFLHPMDRLLEL EGLASVRYADDFVVLCRSKALAQKAQLIVASHLAARGLKLNMSKTRILAPSEAFIFL GQTVEPVWDTQP (SEQ ID NO: 56)

[0586] Vp96 - RT:

[0587] NLVKRLAHHLGKSEPEVIHFLADAPNKYRVYKIPKRSYGHRVIAQPTRELKLY QKAFLELYSFPVHSSATAYCKGKSIKDNALSHVKNHYLLKTDLENFFNSITPNIFWKS IENDSIATPKFSTSEIALVERLIFWRPSKLQGGKLVLSVGAPSSPTISNFCLYQFDEYLSI ICKEQNISYTRYADDLTFSTCDKDVLHTVIPLIQSLLDYFFASELKLNHSKTVFSSKAH NRHVTGITLNNEGKLSLGRERKRYIKHLVHSFKYGKLDNTEIRHLQGMLSFAKHIEPI FIDRLKEKYTDELIKIIYEAGHE (SEQ ID NO: 57)

[0588] Vc95 - RT:

[0589] NILTTLREQLLTNNVIMPQEFERLEVRGSHAYKVYSIPKRKAGRRTIAHPSSKL KICQRHLNAILNPLLKVHDSSYAYVKGRSIKDNALVHSHSAYVLKMDFQNFFNSITP TILRQCLIQNDILLSVNELEKLEQLIFWNPSKKRNGKLILSVGSPISPLISNAIMYPFDKII NDICTKHGINYTRYADDITFSTNIKNTLNKLPEIVEQLIIQTYAGRIIINKRKTVFSSKKH

NRHVTGITLTNDSKISIGRSRKRYISSLVFKYINKNLDIDEINHMKGMLAFAYNIEPIYI HRLSHKYKVNIVEKILRGSN (SEQ ID NO: 241)

[0590] Ec48- RT:

[0591] GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDE KYTLKEIPKIDGSKRIVYSLHPKMRLLQSRINKRIFKELVVFPSFLFGSVPSKNDVLNS NVKRDYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDIC TKDDFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDF SQMQSHIERMLSEHDLPINKHKTKIFHCSSEPIKVHGLRVDYDSPRLPSDEVKRIRASI HNLKLLAAKNNTKTSVAYRKEFNRCMGRVNKLGRVGHEKYESFKKQLQAIKPMPS KRDVAVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKS RLASLKPL (SEQ ID NO: 59)

[0592] Gs-RT:

[0593] ALLERILARDNLITALKRVEANQGAPGIDGVSTDQLRDYIRAHWSTIHAQLLA GTYRPAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGR NAHDAVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIR AYLQAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDC NIYVKSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARI RLAPRSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEG WIRRRLRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQ ALGKTYWTAQGLKSLTQRYFELRQG (SEQ ID NO: 60)

[0594] Er-RT:

[0595] DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIK GQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHS YGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDV ISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYA DDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPR AHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSM KTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYV CNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC (SEQ ID NO: 185)

[0596] Ne144-RT:

[0597] AGQPTSREALYERIRSTSKEEVILEEMIRLGFWPAQGAVPHDPAEEIRRRGELE RQLSELREKSRKLYNEKALIAEQRKQRLAESRRKQKETKARRERERQERAQKWAQR KAGEILFLGEDVSGGMSHKTCDAELIKREGVPAIASAEELARAMGIALKELRFLAYN RKVSRVTHYRRFLLPKKTGGLRLISAPMPRLKRAQAWALEHIFNKLSFEPAAHGFVA

GRSIVSNARPHVGADVVVNLDLKDFFPTVSFPRVKGALRHLGYSESVATALALVCTE
PEVDEVGLDGTTWYVARGERFLPQGSPCSPAITNLLCRRLDRRLHGLAQALGFVYTR
YADDLTFSGRGEAAESKRVGKLLRGAADIVAHEGFVVHPDKTRVMRRGRRQEVTG
VVVNDKTSVPRDELRKFRATLYQIEKDGPADKRWGNGGDVLAAVHGYACFVAMV
DPSRGQPLLARARALLAKHGGPSKPPGGSGPRAPTPVQPTANAPEAPKPVAPATPAA
PAKKGWKLF (SEQ ID NO: 239)

RT variants engineered in this Example:

- AVIRE-D199N (i.e., AVIRE-RT (SEQ ID NO: 216) containing a D199N substitution)
- AVIRE- T305K (i.e., AVIRE-RT (SEQ ID NO: 216) containing a T305K substitution)
- AVIRE- W312F (i.e., AVIRE-RT (SEQ ID NO: 216) containing a W312F substitution)
- AVIRE- G329P (i.e., AVIRE-RT (SEQ ID NO: 216) containing a G329P substitution)
- AVIRE- L604W (i.e., AVIRE-RT (SEQ ID NO: 216) containing a L604W substitution)

- KORV-D197N (i.e., KORV-RT (SEQ ID NO: 222) containing a D197N substitution)
- KORV-T303K (i.e., KORV-RT (SEQ ID NO: 222) containing a T303K substitution)
- KORV-W310F (i.e., KORV-RT (SEQ ID NO: 222) containing a W310F substitution)
- KORV-E327P (i.e., KORV-RT (SEQ ID NO: 222) containing a E327P substitution)
- KORV-L599W (i.e., KORV-RT(SEQ ID NO: 222) containing a L599W substitution)

- WMSV-D197N (i.e., WMSV -RT (SEQ ID NO: 228) containing a D197N substitution)
- WMSV- T303K (i.e., WMSV -RT (SEQ ID NO: 228) containing a T303K substitution)
- WMSV – W311F (i.e., WMSV -RT (SEQ ID NO: 228) containing a W311F substitution)
- WMSV- E327P (i.e., WMSV -RT (SEQ ID NO: 228) containing a E327P substitution)

- WMSV- L599W (i.e., WMSV -RT (SEQ ID NO: 228) containing a L599W substitution)

- PERV-D199N (i.e., PERV -RT (SEQ ID NO: 45) containing a D199N substitution)
- PERV-T305K (i.e., PERV -RT (SEQ ID NO: 45) containing a 305K substitution)
- PERV-W312F (i.e., PERV -RT (SEQ ID NO: 45) containing a W312F substitution)
- PERV-E329P (i.e., PERV -RT (SEQ ID NO: 45) containing a E329P substitution)
- PERV-L602W (i.e., PERV -RT (SEQ ID NO: 45) containing a L602W substitution)
- PERV- D199N+T305K+W312F+E329P+L602W (i.e., PERV -RT (SEQ ID NO: 45) containing D199N+T305K+W312F+E329P+L602W substitutions)

- Tf1-K118R (i.e., Tf1-RT (SEQ ID NO: 55) containing a K118R substitution)
- Tf1-S297Q (i.e., Tf1-RT (SEQ ID NO: 55) containing a S297Q substitution)
- Tf1-K118R+S297Q (i.e., Tf1-RT (SEQ ID NO: 55) containing K118R+S297Q substitutions)

Protein sequences of RT variants evolved in this study:

- 5.27-(V14A+L158Q+F269L+K356E)
- 5.59-(E22K+P70T+G72V+M102I+K106R+A139T+L158Q+F269L+ A363V+K413E+S492N)
- 5.60-(P70T+G72V+M102I+K106R+L158Q+F269L+A363V+K413E+S492N)
- 3.8-(R267I+K318E+K326E+E328K+R372K)
- 3.35-(E54K+K87E+D243N+R267I+E279K+K318E)
- 3.36-(A36V+K87E+R205K+D243N+R267I+E279K+K318E)
- 3.38-(E54K+K87E+D243N+R267I+S277F+E279K+K318E)
- 38.14: Ne144 (A157T+A165T+G288V)
- 25.8 – Vc95 (L11M+S75A+V97M+N146D+N245T)

Mutant Gs-RT Prime Editors (All mutations are referring to Gs-RT; the architecture for all is Cas9(H840A)-Mutant Gs RT.)

- 809: L17P + D206V (SEQ ID NO: 159)
- 810: N12D + L37R + G78V (SEQ ID NO: 160)
- 811: A16E + L37P + A123V (SEQ ID NO: 161)
- 812: A16V + R38H + W45R + Y126F + Q412H (SEQ ID NO: 162)

- 813: A16V + R38H + W45R + R291K (SEQ ID NO: 163)
- 814: N12D + L37R + G72E + E129G + P264S + R344S + R360S (SEQ ID NO: 164)
- 815: N12D + Y40C + I67T + G73V + Q93R + R287I + R358S (SEQ ID NO: 165)
- 816: N12D + Y40C + I67T + G73V + Q93R + R358S (SEQ ID NO: 166)
- 817: N12D + I41N + P190L + A234V + K279E (SEQ ID NO: 167)
- 818: N12D + L37R + R267M + P309T + R358S + E363G (SEQ ID NO: 168)
- 819: A16V + V20G + I41S + R233K + P264S (SEQ ID NO: 169)
- 820: L17P + V20G + I41S + I67R + R263G + P264S + V374A (SEQ ID NO: 170)
- 821: L17P + V20G + I41S + I67R + K162N + R263G + P264S (SEQ ID NO: 55)

Mutant M-MLV Prime Editors (All mutations are referring to the WT MMLV RT; the architecture for all is Cas9(H840A)-Mutant M-MLV RT.)

- Clones 1 and 2: D200Y + E302A
- Clones 3 and 4: D200Y + V223A + M457I
- Clones 5-8: V223M + T306K + A462S
- Clones 9 and 10: D200N + E302K
- Clones 11 and 14: D220Y + E302K
- Clones 13 and 16: D200Y
- Clone 15: V223M

Prime Editors with a Mutant Cas9 (All mutations are in reference to Cas9; the architecture for all is Mutant Cas9(H840A)- M-MLV RT

- 1043: H721Y + R753G (SEQ ID NO: 178)
- 1044: E102K + R753G (SEQ ID NO: 179)
- 1045: E102K + H721Y + R753G (SEQ ID NO: 180)

Sequences (for Example 2)

[0598] The following amino acid sequences were obtained as a result of Example 2, described above, and includes evolved RT amino acid sequences, evolved Cas9 amino acid sequences, and evolved fusion protein sequences. This application also contemplates any additional variant sequences (e.g., variant RT or Cas9 sequences or PE fusion protein sequences) that combines one or more mutations of any one variant with that of another.

[0599] In addition, the application contemplates any amino acid sequence having at least 70%, or at least 75%, or at least 80%, or at least 85%, or at least 90%, or at least 95%, or at least 99%, or up to 100% sequence identity with any of the following amino acid sequences,

and preferably wherein the amino acid sequences having such sequence identity retain one or more mutations in the below sequences.

Evolved Gs Reverse Transcriptases (SEQ ID NOs: 159-171):

[0600] *Gs variants comprising: L17P + D206V*

EANQGAPGIDGVSTDQLRDYIRAHWSTIHAQLLAGTYRPAPVRRVEIPKPGGGTRQL
GIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDAVRQAQGYIQEGYRYVV
DMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQAGVMIEGVKVQTEEGTP
QGGPLSPLLANILLD**V**LDKELEKRGLKFCRYADDCNIYVKSLRAGQRVKQSIQRFLE
KTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSIQRLKQRIRQLTNPNWS
ISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRLRLCQWLQWKRVRTRIR
ELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTYWTAQGLKSLTQRYFEL
RQG (SEQ ID NO: 159)

[0601] *Gs variant N12D + L37R + G78V*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQL**V**IPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 160)

[0602] *Gs A16E + L37P + A123V*

ALLERILARDNLIT**E**LKRVEANQGAPGIDGVSTDQ**P**RDYIRAHWSTIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQ**V**QGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 161)

[0603] *Gs variant A16V + R38H + W45R + Y126F + Q412H*

ALLERILARDNLIT**V**LKRVEANQGAPGIDGVSTDQL**H**DYIRAH**R**STIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQAQG**F**IQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLT**H**RYFELRQG (SEQ ID NO: 162)

[0604] *Gs A16V + R38H + W45R + R291K*

ALLERILARDNLIT**V**LKRVEANQGAPGIDGVSTDQL**H**DYIRAH**R**STIHAQLLAGTYRP
APVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHDA
VRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQA
GVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVKS
LRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRSI
Q**K**LKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 163)

[0605] *Gs variant 814 N12D + L37R + G72E + E129G + P264S + R344S + R360S*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPG**E**GTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQ**G**GYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDR**S**WKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIR**S**RL
RLCQWLQWKRVRT**S**IRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 164)

[0606] *Gs variant 815 N12D + Y40C + I67T + G73V + Q93R + R287I + R358S*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRD**C**IRAHWSTIHAQLLAGTYRP
APVRRVE**T**PKPGG**V**TRQLGIPTVVDRLIQQAIL**R**ELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAP**I**SI
QRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRV**S**TRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 165)

[0607] *Gs variant 816 N12D + Y40C + I67T + G73V + Q93R + R358S*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRD**C**IRAHWSTIHAQLLAGTYRP
APVRRVE**T**PKPGG**V**TRQLGIPTVVDRLIQQAIL**R**ELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPERKARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL
RLCQWLQWKRV**S**TRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 166)

[0608] *Gs variant 817 N12D + I41N + P190L + A234V + K279E*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQLRDY**N**RAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGT**L**QGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLR**V**GQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWKRAFLGFSFTPER**E**ARIRLAPRS
IQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRRL

RLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKTY
WTAQGLKSLTQRYFELRQG (SEQ ID NO: 167)

[0609] *Gs variant 818 N12D + L37R + R267M + P309T + R358S + E363G*

ALLERILARD**D**LITALKRVEANQGAPGIDGVSTDQ**R**RDYIRAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVDRPWK**M**AFLGFSFTPERKARIRLAPR
SIQRLKQRIRQLTNPNWSISM**T**ERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRR
LRLCQWLQWKRV**S**TRIR**G**LRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKT
YWTAQGLKSLTQRYFELRQG (SEQ ID NO: 168)

[0610] *Gs variant 819 A16V + V20G + I41S + R233K + P264S*

ALLERILARDNLIT**V**LKR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVEIPKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAHD
AVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYLQ
AGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYVK
SL**K**AGQRVKQSIQRFLEKTLKLKVNEEKSAVDR**S**WKRAFLGFSFTPERKARIRLAPR
SIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRRR
LRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGKT
YWTAQGLKSLTQRYFELRQG (SEQ ID NO: 169)

[0611] *Gs variant 820 L17P + V20G + I41S + I67R + R263G + P264S + V374A*

ALLERILARDNLITA**P**KR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVE**R**PKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAH
DAVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDKRVLKLIRAYL
QAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYV
KSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVD**GS**WKRAFLGFSFTPERKARIRLAP
RSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRR
RLRLCQWLQWKRVRTRIRELRALGLKETA**A**MEIANTRKGAWRTTKTPQLHQALGK
TYWTAQGLKSLTQRYFELRQG (SEQ ID NO: 170)

[0612] *Gs variant 821 L17P + V20G + I41S + I67R + K162N + R263G + P264S*

ALLERILARDNLITA**P**KR**G**EANQGAPGIDGVSTDQLRDY**S**RAHWSTIHAQLLAGTYR
PAPVRRVE**R**PKPGGGTRQLGIPTVVDRLIQQAILQELTPIFDPDFSSSSFGFRPGRNAH
DAVRQAQGYIQEGYRYVVDMDLEKFFDRVNHDILMSRVARKVKDK**N**VLKLIRAYL
QAGVMIEGVKVQTEEGTPQGGPLSPLLANILLDDLDKELEKRGLKFCRYADDCNIYV
KSLRAGQRVKQSIQRFLEKTLKLKVNEEKSAVD**GS**WKRAFLGFSFTPERKARIRLAP
RSIQRLKQRIRQLTNPNWSISMPERIHRVNQYVMGWIGYFRLVETPSVLQTIEGWIRR
RLRLCQWLQWKRVRTRIRELRALGLKETAVMEIANTRKGAWRTTKTPQLHQALGK
TYWTAQGLKSLTQRYFELRQG (SEQ ID NO: 171)

Evolved MMLV Reverse Transcriptases (SEQ ID NOs: 172-177):

[0613] Each of the following evolved MMLV RT variants are based on the wildtype MMLV RT of SEQ ID NO: 33, but wherein each variant MMLV RT includes a C-terminal truncation of about 180 amino acids, which corresponds to the RNaseH domain.

[0614] For comparison, wildtype MMLV RT has the following amino acid sequence:

[0615] *Wildtype MMLV RT amino acid sequence:*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
VSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR
DPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQK
AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK
LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS
NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDL
TDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIA
LTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLK
ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP (SEQ ID NO: 33)

[0616] The application contemplates the following evolved MMLV RT variants (which are relative to wildtype MMLV RT).

[0617] *MMLV variant: MMLV D200S + V223A + E346K + W388C*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
VSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR
DPEMGISGQLTWTRLPQGFKNSPTLF**S**EALHRDLADFRIQHPDLILLQY**A**DDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKA
YQ**K**IKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPW**C**RPVAYLSKKL
DPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN
ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 172)

[0618] *MMLV variant: MMLV S60Y + V223A + N249S*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
V**Y**IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL

REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW
RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY**A**DDLLLA
ATSELDCQQGTRALLQTLG**S**LGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEAR
KETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQ
KAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWL
SNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO:
173)

[0619] *MMLV variant: MMLV P111L + V223A + T287A + G316R*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
VSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYR**L**VQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW
RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY**A**DDLLLA
ATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEAR
KE**A**VMGQPTPKTPRQLREFLGKAGFCRLFIP**R**FAEMAAPLYPLTKPGTLFNWGPDQQ
KAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWL
SNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO:
174)

[0620] *MMLV variant: MMLV S60Y + G138R + V223A*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
V**Y**IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLS**R**LPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW
RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY**A**DDLLLA
ATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEAR
KETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQ
KAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK
KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWL
SNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO:
175)

[0621] *MMLV variant: MMLV S60Y + Y222F + V223A + K445N*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
V**Y**IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW

RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ**FA**DDLLLAA
TSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK
ETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK
AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK
LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALV**N**QPPDRWLS
NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 176)

[0622] *MMLV variant: MMLV S60Y + C157F + V223A + T246I*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
V**Y**IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL
REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFF**F**LRLHPTSQPLFAFEW
RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY**A**DDLLLA
ATSELDCQQGTRALLQ**I**LGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK
ETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK
AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK
LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS
NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 177)

Evolved Cas9 variants:

[0623] *Evolved Cas9 variant: Cas9 H721Y + R753G*

[0624] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL
IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL**Y**EHIANLAGSPAIKKGILQTVKVVDE
LVKVMG**G**HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR

SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG (SEQ ID
NO: 178)

[0625] *Evolved Cas9 variant: Cas9 E102K + R753G*

[0626] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL**K**ESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL
IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE
LVKVMG**G**HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR
SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL

TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG (SEQ ID NO: 179)

[0627] *Evolved Cas9 variant: Cas9 E102K + H721Y + R753G*

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL**K**ESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSL**Y**EHIANLAGSPAIKKGILQTVKVVDELVK
VMG**G**HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG (SEQ ID NO: 180)

Modified PE fusion protein amino acid sequences comprising MMLV RT mutations:

[0628] *PE fusion protein comprising MMLV P111L + V223A + T287A + G316R*

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL

SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSTLNIEDEYRLHETSK
EPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARL
GIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYR**L**VQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTR
LPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY**A**DDLLLAATSELDCQQGTRALLQ
TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE**A**VMGQPTPKTPRQL
REFLGKAGFCRLFIP**R**FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPAL
GLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDT
DRVQFGPVVALNPATLLPLPEEGLQHNCLDSGGSKRTADGSEFESPKKKRKVGSGPA
AKRVKLD (SEQ ID NO: 181)

[0629] *PE fusion protein comprising Cas9 (R753G) and MMLV RT comprising rev. transcriptase mutations at S60Y + C157F + V223A + T246I*

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSTLNIEDEYRLHETSK
EPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVYIKQYPMSQEARL
GIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFFLRLHPTSQPLFAFEWRDPEMGISGQLTWTR
LPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYADDLLLAATSELDCQQGTRALLQ
ILGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLR
EFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALG
LPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMV
AAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTD

RVQFGPVVALNPATLLPLPEEGLQHNCLDSGGSKRTADGSEFESPKKKRKVGSGPAA
KRVKLD (SEQ ID NO: 182)

Additional MMLV variants (SEQ ID NOs: 183-184):

[0630] MMLV variant: V223M + T306K + A462S

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
VSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR
DPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYMDDLLLAA
TSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARK
ETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK
AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK
LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS
NARMTHYQSLLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 183)

[0631] MMLV variant: D200N + E302K

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTP
VSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWR
DPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLRKFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK
AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK
LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS
NARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD (SEQ ID NO: 184)

Er RT wild-type:

DTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKGQLRT
RKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQFHDHSYGFRP
NRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIGRTIKDGDVISIVRK
YLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCII
MVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQF
KAKPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGSMKTLCK
ELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKG
AVNVAISNKRLASFGLISMLDYYIEKCVTC (SEQ ID NO: 185)

Rs09415 RT (or "CRISPR-RT") wild-type:

NSQAQSACCAGANQIVEGATLEKVVAPACLQQAWTRVRKNKGGPGGDGVTIEIFAQ
NAEVELEKLRAETLAGIYRPRKVRHAIVPKPKGGERKLTIPSVVDRILQTATMLSLGQ
TVDHHFSSASWAYREGRGVDDALADLRRLRNSGLFWTFDADIMQYFDRILHKRLID
DLFIWVDDLRIVRLIQLWLRSFSYWGRGIAQGAPISPLLANLFLHPMDRLLELEGLAS
VRYADDFVVLCRSKALAQKAQLIVASHLAARGLKLNMSKTRILAPSEAFIFLGQTVE
PVWDTQP (SEQ ID NO: 56)

HIV-MMLV:

PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTP
VFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDA
YFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFKK
QNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLW
MGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRG
TKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQ
EPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWET
WWTEYWQATWIPEWEFVNTPPLVKLVVALNPATLLPLPEEGLQHNCLDILAEAHGT
RPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRA
ELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEIL
ALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIEN
(SEQ ID NO: 46)

Ec48 variants: 3.23, 3.35, 3.36, 3.37, 3.38, 3.5, 3.501, 3.8 (SEQ ID NOs: 188-195):

[0632] Ec48 variant 3.23:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQKKGLVYTRLLDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDEVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVSELGRVGQEEYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 188)

[0633] Ec48 variant 3.35 (or Ec48-evo2):

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 189)

[0634] Ec48 variant 3.36:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKVLSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQKKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 190)

[0635] Ec48 variant 3.37:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 191)

[0636] Ec48 variant 3.38:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDKKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD

FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPFDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSKRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 192)

[0637] Ec48 variant 3.500:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALDYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 193)

[0638] Ec48 variant 3.501:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRTVFEEILHIKDEALDYLVDICTKD
DFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQM
QSHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLK
LLAAKNNTKTSMAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDV
AVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASL
KPL (SEQ ID NO: 194)

[0639] Ec48 variant 3.8 (or Ec48-evo1):

[0640] GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDE
KYTLKEIPKIDGSKRIVYSLHPKMRLLQSRINKRIFKELVVFPSFLFGSVPSKNDVLNS
NVKRDYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDIC
TKDDFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDF
SQMQSHIERMLSEHDLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDEVKRIRASIH
NLKLLAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEEYKSFKKQLQAIKPMPSK

RDVAVIDAAIKSLELSYSKGNQNKHWYKKKYDLTRYKMIILTRSESFKEKLECFKSR
LASLKPL (SEQ ID NO: 195)

Ec48 variants comprising: E60K, E165D, S151T, V303M, K343N (SEQ ID NOs: 193-194):

[0641] Ec48 variant 3.500:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALDYLVDICTKDD
FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL
LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDVA
VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK
PL (SEQ ID NO: 193)

[0642] Ec48 variant 3.501:

GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRTVFEEILHIKDEALDYLVDICTKD
DFVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQM
QSHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLK
LLAAKNNTKTSMAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDV
AVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASL
KPL (SEQ ID NO: 194)

Tf1 variants: 5.131, 5.27, 5.47, 5.59, 5.60, 5.612, 5.618 (SEQ ID NOs: 196-202):

[0643] Tf1 variant 5.131:

ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL

RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 196)

[0644] Tf1 variant 5.27:

ISSSKHTLSQMNKASNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKEILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 197)

[0645] Tf1 variant 5.47:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPRKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTVEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 198)

[0646] Tf1 variant 5.59:

ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK

EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 199)

[0647] Tf1 variant 5.60:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFEPKK
KRKV (SEQ ID NO: 200)

[0648] Tf1 variant 5.612:

ISSSKHTLSQMNKVSNIVKEPKLPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPLRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLTKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNRK
ELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 201)

[0649] Tf1 variant 5.618:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD

KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 202)

Tf1 variants comprising: S188K, I260L, R288Q, Q293K, I64L, I64W, N316Q, K321R, L133N (SEQ ID NOs: 203-213):

[0650] Tf1 variant S188K:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRK
ELRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 203)

[0651] Tf1 variant I260L:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 204)

[0652] Tf1 variant R288Q:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD

VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNQKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 205)

[0653] Tf1 variant Q293K:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRKFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 206)

[0654] Tf1 variant I64L:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPLRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 207)

[0655] Tf1 variant I64W:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPWRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVV
DYKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCP
RGVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRK

ELRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 208)

[0656] Tf1 variant N316Q:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLQKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 209)

[0657] Tf1 variant K321R:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKRDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 210)

[0658] Tf1 variant L133N:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YKPLNKYVKPNIYPLPNIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV

LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 211)

[0659] Tf1 variant K118R:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVD
YRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 212)

[0660] Tf1 variant S297Q:

[0661] ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVEL
TQENYRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLR
MVVDYKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLA
FRCPRGVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHV
KHVKDVLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGFTPCQENIDKVLQWKQP
KNRKELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCL
VSPPVLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNY
SVSDKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFL
QDFNFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 213)

[0662] Tf1_1 max: PE fusion protein comprising TF1 variant

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG

TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSISSSKHTLSQMNKVS
NIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQENYRLPIRNYPLPPGKM
QAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVVDYRPLNKYVKPNIYPL
PLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCPRGVFEYLVMPYGIKT
APAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDVLQKLKNANLIINQ
AKCEFHQSQVKFLGYHISEKGFTPCQENIDKVLQWKQPKNQKELRQFLGQVNYLRK
FIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVLRHFDFSKKILLET
DASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKSLKHWR
HYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHIA
DALSRIVDETEPIPKDSEDNSINFVNQISISGGSKRTADGSEFESPKKKRKVGSGPAAK
RVKLD (SEQ ID NO: 246)

[0663] Tf1_2 max: PE fusion protein comprising TF1 variant

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL

SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSISSSKHTLSQMNKVS
NIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQENYRLPIRNYPLTPVKM
QAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDYRPLNKYVKPNVYPL
PLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRGVFEYLVMPYGIKT
APAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDVLQKLKNANLIINQ
AKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQKELRQFLGQVNYLRK
FIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVLRHFDFSKKILLET
DVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKSLEHWR
HYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHIA
DALSRIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFESPKKKRKVGSGPAAK
RVKLD (SEQ ID NO: 247)

[0664] Tf1_3 max: PE fusion protein comprising TF1 variant

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSISSSKHTLSQMNKVS
NIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQENYRLPIRNYPLTPVKM
QAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDYRPLNKYVKPNIYPLP
LIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRGVFEYLVMPYGIKTA
PAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDVLQKLKNANLIINQA
KCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQKELRQFLGQVNYLRKFI
PKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVLRHFDFSKKILLETDV
SDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKSLEHWRHY
LESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHIADAL

SRIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFESPKKKRKVGSGPAAKRVK
LD (SEQ ID NO: 248)

[0665] Tf1_4 max: PE fusion protein comprising TF1 variant

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSISSSKHTLSQMNKVS
NIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQENYRLPIRNYPLTPVKM
QAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDYKPLNKYVKPNIYPLP
LIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRGVFEYLVMPYGISTA
PAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDVLQKLKNANLIINQA
KCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKELRQFLGSVNYLRKFIP

KTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVLRHFDFSKKILLETDVS
DVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKSLEHWRHYL
ESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHIADALS
RIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFESPKKKRKVGSGPAAKRVKL
D (SEQ ID NO: 249)

[0666] Tf1_5 max: PE fusion protein comprising TF1 variant

MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNT
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLI
YLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL
TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDN
LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSISSSKHTLSQMNKVS
NIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQENYRLPIRNYPLTPVKM
QAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDYRPLNKYVKPNVYPL

PLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRGVFEYLVMPYGIST
APAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDVLQKLKNANLIINQ
AKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKELRQFLGSVNYLRKF
IPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVLRHFDFSKKILLETD
VSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKSLEHWRH
YLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEINYRPGSANHIAD
ALSRIVDETEPIPKDNEDNSINFVNQISISGGSKRTADGSEFESPKKKRKVGSGPAAKR
VKLD (SEQ ID NO: 250)

PERV variants: 21 and 21.6 (SEQ ID NOs: 214-215):

[0667] PERV variant 21:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 214)

[0668] PERV variant 21.6 (pentamutant comprising D199N, T305K, W312F, E329P, and L602W substitutions):
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQK

AFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNA
RMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALTQ
ALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHL
PKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 215)

AVIRE variants comprising: D199N, T305K, W312F, G329P, L604W (SEQ ID NOs: 216-221):

[0669] AVIRE wildtype

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 216)

[0670] AVIRE-RT (D199N):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA

GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 217)

[0671] AVIRE-RT (T305K):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGKIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSL
KLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVA
AGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQ
VLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAE
ATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKD
KSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 218)

[0672] AVIRE-RT (W312F):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLFIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV

MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 219)

[0673] AVIRE-RT (G329P):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 220)

[0674] AVIRE-RT (L604W):

APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE
GESGQLTWTRLPQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 221)

KORV variants comprising: D197N, T303K, W310F, E327P, L599W (SEQ ID NOs: 222-227):

[0675] KORV wildtype:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 222)

[0676] KORV-RT D197N:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 223)

[0677] KORV-RT T303K:

MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGKAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQ
EAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTN
ARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPL
PGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALR
LAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKR
VAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 224)

[0678] KORV-RT W310F:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLFIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 225)

[0679] KORV-RT E327P:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA

APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 226)

[0680] KORV-RT L599W:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 227)

WMSV variants comprising: D197N, T303K, W311F, E327P, L599W (SEQ ID NOs: 228-233):

[0681] WMSV-RT wildtype:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA

TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 228)

[0682] WMSV-RT D197N:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 229)

[0683] WMSV-RT T303K:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGKAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL

AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 230)

[0684] WMSV-RT W311F:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLFIPGFASLAAPLYPLTKESIPFIWTEEHQKAFDR
IKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPVA
SGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMT
HYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLPG
VPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLA
EGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAII
HCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 231)

[0685] WMSV-RT E327P:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVA
IIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 232)

[0686] WMSV-RT L599W:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE

INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV
ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM
THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP
GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL
AEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO:
233)

PERV variants comprising: D199N, T305K, E329P, L602W (SEQ ID NO: 234-238):

[0687] PERV-RT D199N:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 234)

[0688] PERV-RT T305K:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL

AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGKAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 235)

[0689] PERV-RT W313F:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLFIPGFATLAAPLYPLTKEKGEFSWAPEHQK
AFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKK
LDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNA
RMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALTQ
ALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLP
KRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 236)

[0690] PERV-RT E329P:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT

QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 237)

[0691] PERV-RT L602W:
TLQLDDEYRLYSPQVKPDQDIQSWLEQFPQAWAETAGMGLAKQVPPQVIQLKASAT
PVSVRQYPLSREAREGIWPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQD
LREVNKRVQDIHPTVPNPYNLLSALPPERNWYTVLDLKDAFFCLRLHPTSQPLFAFE
WRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLL
AGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRGGQRWLTEAR
KKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQ
KAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK
KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTN
ARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTHTIWASSLPEGTSAQKAELMALT
QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALH
LPKRLAIIHCPGHQKAKDLISRGNQMADRVAKQAAQAVNLLPI (SEQ ID NO: 238)

Ne144 comprising: 38.14 variant (SEQ ID NOs: 239-240):

[0692] Ne144 RT wildtype
AGQPTSREALYERIRSTSKEEVILEEMIRLGFWPAQGAVPHDPAEEIRRRGELERQLSE
LREKSRKLYNEKALIAEQRKQRLAESRRKQKETKARRERERQERAQKWAQRKAGEI
LFLGEDVSGGMSHKTCDAELIKREGVPAIASAEELARAMGIALKELRFLAYNRKVSR
VTHYRRFLLPKKTGGLRLISAPMPRLKRAQAWALEHIFNKLSFEPAAHGFVAGRSIVS
NARPHVGADVVVNLDLKDFFPTVSFPRVKGALRHLGYSESVATALALVCTEPEVDE
VGLDGTTWYVARGERFLPQGSPCSPAITNLLCRRLDRRLHGLAQALGFVYTRYADD
LTFSGRGEAAESKRVGKLLRGAADIVAHEGFVVHPDKTRVMRRGRRQEVTGVVVN
DKTSVPRDELRKFRATLYQIEKDGPADKRWGNGGDVLAAVHGYACFVAMVDPSRG
QPLLARARALLAKHGGPSKPPGGSGPRAPTPVQPTANAPEAPKPVAPATPAAPAKKG
WKLF (SEQ ID NO: 239)

[0693] Ne144 RT 38.14:
AGQPTSREALYERIRSTSKEEVILEEMIRLGFWPAQGAVPHDPAEEIRRRGELERQLSE
LREKSRKLYNEKALIAEQRKQRLAESRRKQKETKARRERERQERAQKWAQRKAGEI

LFLGEDVSGGMSHKTCDAELIKREGVPAIASAEELARAMGITLKELRFLTYNRKVSR
VTHYRRFLLPKKTGGLRLISAPMPRLKRAQAWALEHIFNKLSFEPAAHGFVAGRSIVS
NARPHVGADVVVNLDLKDFFPTVSFPRVKGALRHLGYSESVATALALVCTEPEVDE
VVLDGTTWYVARGERFLPQGSPCSPAITNLLCRRLDRRLHGLAQALGFVYTRYADD
LTFSGRGEAAESKRVGKLLRGAADIVAHEGFVVHPDKTRVMRRGRRQEVTGVVVN
DKTSVPRDELRKFRATLYQIEKDGPADKRWGNGGDVLAAVHGYACFVAMVDPSRG
QPLLARARALLAKHGGPSKPPGGSGPRAPTPVQPTANAPEAPKPVAPATPAAPAKKG
WKLF (SEQ ID NO: 240)

Vc95 comprising: 25.8 variant (SEQ ID NOs: 241-242):

[0694] Vc95 RT wildtype:
NILTTLREQLLTNNVIMPQEFERLEVRGSHAYKVYSIPKRKAGRRTIAHPSSKLKICQR
HLNAILNPLLKVHDSSYAYVKGRSIKDNALVHSHSAYVLKMDFQNFFNSITPTILRQC
LIQNDILLSVNELEKLEQLIFWNPSKKRNGKLILSVGSPISPLISNAIMYPFDKIINDICT
KHGINYTRYADDITFSTNIKNTLNKLPEIVEQLIIQTYAGRIIINKRKTVFSSKKHNRHV
TGITLTNDSKISIGRSRKRYISSLVFKYINKNLDIDEINHMKGMLAFAYNIEPIYIHRLS
HKYKVNIVEKILRGSN (SEQ ID NO: 241)

[0695] Vc95 RT variant - 25.8:
NILTTLREQLMTNNVIMPQEFERLEVRGSHAYKVYSIPKRKAGRRTIAHPSSKLKICQ
RHLNAILNPLLKVHDASYAYVKGRSIKDNALVHSHSAYMLKMDFQNFFNSITPTILR
QCLIQNDILLSVNELEKLEQLIFWNPSKKRDGKLILSVGSPISPLISNAIMYPFDKIINDI
CTKHGINYTRYADDITFSTNIKNTLNKLPEIVEQLIIQTYAGRIIINKRKTVFSSKKHNR
HVTGITLTTDSKISIGRSRKRYISSLVFKYINKNLDIDEINHMKGMLAFAYNIEPIYIHR
LSHKYKVNIVEKILRGSN (SEQ ID NO: 242)

Sequences for FIG. 59 (SEQ ID NOs 243-245)

[0696] AVIRE_penta:
APLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVR
VRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV
NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE

GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA
ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIP
VPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLK
LALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA
GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV
LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEA
TLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDK
SVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAV
MHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISDAPDMPDTETPQYSNVE
EALG (SEQ ID NO: 243)

[0697] KORV_penta:
MNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPV
AVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLR
EVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWR
DPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVA
APTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPAR
KATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQE
AFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKL
DPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNA
RMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLP
GVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRL
AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV
AIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKNQEHFEPTRGK (SEQ
ID NO: 244)

[0698] WMSV_penta:
LNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVA
VRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE
INKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDP
EKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAP
TYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKA
TVMKIPPPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFD
RIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPV

ASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARM THYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLP GVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRL AEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRV AIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKPQELI (SEQ ID NO: 245)

References (for Example 2)

1. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).
2. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, doi:10.1038/nature09929 (2011).
3. Taube, R., Loya, S., Avidan, O., Perach, M. & Hizi, A. Reverse transcriptase of mouse mammary tumour virus: expression in bacteria, purification and biochemical characterization. *Biochem J* **332 (Pt 3)**, 807-808, doi:10.1042/bj3320807w (1998).
4. Hizi, A. & Herschhorn, A. Retroviral reverse transcriptases (other than those of HIV-1 and murine leukemia virus): a comparison of their molecular and biochemical properties. *Virus Res* **134**, 203-220, doi:10.1016/j.virusres.2007.12.008 (2008).
5. Avidan, O., Loya, S., Tonjes, R. R., Sevilya, Z. & Hizi, A. Expression and characterization of a recombinant novel reverse transcriptase of a porcine endogenous retrovirus. *Virology* **307**, 341-357, doi:10.1016/s0042-6822(02)00131-9 (2003).
6. Misra, H. S., Pandey, P. K. & Pandey, V. N. An enzymatically active chimeric HIV-1 reverse transcriptase (RT) with the RNase-H domain of murine leukemia virus RT exists as a monomer. *J Biol Chem* **273**, 9785-9789, doi:10.1074/jbc.273.16.9785 (1998).
7. Kirshenboim, N., Hayouka, Z., Friedler, A. & Hizi, A. Expression and characterization of a novel reverse transcriptase of the LTR retrotransposon Tf1. *Virology* **366**, 263-276, doi:10.1016/j.virol.2007.04.002 (2007).
8. Nowak, E. *et al.* Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat Struct Mol Biol* **21**, 389-396, doi:10.1038/nsmb.2785 (2014).

9. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol* **37**, 1070-1079, doi:10.1038/s41587-019-0193-0 (2019).
10. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883-891, doi:10.1038/s41587-020-0453-z (2020).
11. Simon, A. J., Ellington, A. D. & Finkelstein, I. J. Retrons and their applications in genome engineering. *Nucleic Acids Res* **47**, 11007-11019, doi:10.1093/nar/gkz865 (2019).
12. Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183-195, doi:10.1261/rna.063479.117 (2018).
13. Toro, N. & Nisa-Martinez, R. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS One* **9**, e114083, doi:10.1371/journal.pone.0114083 (2014).
14. Stamos, J.L. *et al.* Structure of a Thermostable Group II Intron Reverse Transcriptase with Template-Prime and Its Functional and Evolutionary Implications. *Mol. Cell.* **68**, 926-939 (2017).

Example 3: Improved Tf1 Reverse Transcriptases Using Rational Engineering

[0699] Further rational engineering of Tf1 revealed 3 additional mutations that improved the editing efficiency of the Tf1-based prime editor. In total, 5 mutations, K118R, S188K, I260L, S297Q and R288Q improved PE (FIG. 46). Combining all five mutations further improved editing, and the final rationally designed variant of Tf1, Tf1-rat4 demonstrated editing comparable to PE2 at many sites (FIG. 47).

[0700] Further evolution has resulted in two additional variants that demonstrate modest improvements in editing, Tf1evo3.1 and Tf1evo3.2 (FIG. 48).

[0701] The rational mutation identified was combined with the best evolved variant. Further small improvements in editing compared to the Tf1evo3.1 and Tf1evo3.2 variants were observed. Some of these final variants (Tf1evo3.1, Tf1evo3.2, Tf1evo+rat-1, Tf1evo+rat2) exhibited higher editing than PE2 on average across 8 different sites (FIG. 49).

[0702] Given the success of our rational engineering efforts for Tf1, a similar strategy was applied to improve the activity of the Ec48-based prime editor. Utilizing an AlphaFold

structure of Ec48, 6 mutations were predicted to improved editing: T189N, R378K, K307R, T385R, L182N and R315K (FIGs. 50A-50B). Combining L182N, R315K and T189N further improved editing (FIG. 51). This variant was named Ec48-v2.

[0703] An additionally evolved variant, Ec48-evo3, was generated which exhibited further improved editing (Ec48-evo3) (FIG. 52). The best variants were then implemented in the PEmax architecture (FIG. 53).

[0704] Tf1-rat4:

MISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQE
NYRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINACPVMFVPKKEGTLRMVV
DYRPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHLIRVRKGDEHKLAFRCP
RGVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGFTPCQENIDKVLQWKQPKNQK
ELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPP
VLRHFDFSKKILLETDASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISI (SEQ ID NO: 251)

[0705] Tf1evo3.1:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
KPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGISTAPAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKDV
LQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKEL
RQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPVL
RHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDK
EMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNF
EINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 252)

[0706] Tf1evo3.2:

ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNVYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGISTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD

VLQKLKNANLIINQAKCEFHQSQVKFIGYHISEKGLTPCQENIDKVLQWKQPKNRKE
LRQFLGSVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 253)

[0707] Tf1evo+rat-1:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKGGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNVYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPR
GVFEYLVMPYGIKTAPAHFQYFINTILGEAKESHVVCYMDDILIHSKSESEHVKHVK
DVLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQ
KELRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSP
PVLRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVS
DKEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDF
NFEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 254)

[0708] Tf1evo+rat2:
ISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITADTNTEKLPKPIKGLEFEVELTQEN
YRLPIRNYPLTPVKMQAMNDEINQGLKSGIIRESKAINACPVIFVPRKEGTLRMVVDY
RPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLDLKSAYHQIRVRKGDEHKLAFRCPRG
VFEYLVMPYGIKTAPAHFQYCINTILGEAKESHVVCYMDDILIHSKSESEHVKHVKD
VLQKLKNANLIINQAKCEFHQSQVKFLGYHISEKGLTPCQENIDKVLQWKQPKNQKE
LRQFLGQVNYLRKFIPKTSQLTHPLNKLLKKDVRWKWTPTQTQAIENIKQCLVSPPV
LRHFDFSKKILLETDVSDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSD
KEMLAIIKSLEHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFN
FEINYRPGSANHIADALSRIVDETEPIPKDNEDNSINFVNQISI (SEQ ID NO: 255)

[0709] Ec48-v2:
GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL
KEIPKIDGSKRIVYSLHPKMRLLQSRINKRIFKELVVFPSFLFGSVPSKNDVLNSNVKR
DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALEYLVDICTKDD
FVVQGANTSSYIANLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ
SHIERMLSEHDLPINKHKTKIFHCSSEPIKVHGLRVDYDSPRLPSDEVKRIRASIHNLK

LLAAKNNTKTSVAYRKEFNRCMGKVNKLGRVGHEKYESFKKQLQAIKPMPSKRDV AVIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASL KPL (SEQ ID NO: 256)

[0710] Ec48-evo3:
GRPYVTLNLNGMFMDKFKPYSKSNAPITTLEKLSKALSISVEELKAIAELSLDEKYTL KKIPKIDGSKRIVYSLHPKMRLLQSRINERIFKELVVFPSFLFGSVPSKNDVLNSNVKR DYVSCAKAHCGAKTVLKVDISNFFDNIHRDLVRSVFEEILHIKDEALDYLVDICTKDD FVVQGALTSSYIATLCLFAVEGDVVRRAQRKGLVYTRLVDDITVSSKISNYDFSQMQ SHIERMLSEHNLPINKHKTKIFHCSSEPIKVHGLIVDYDSPRLPSDKVKRIRASIHNLKL LAAKNNTKTSVAYRKEFNRCMGRVNELGRVGHEKYESFKKQLQAIKPMPSNRDVA VIDAAIKSLELSYSKGNQNKHWYKRKYDLTRYKMIILTRSESFKEKLECFKSRLASLK PL (SEQ ID NO: 257)

Example 4: Improved M-MLV Reverse Transcriptases

[0711] To improve M-MLV to be more efficient than PE2 in mammalian cells, individual PANCE and PACE-evolved mutants were screened in N2A cells. The mutants behaved in different ways, depending on the target edit: some mutations were helpful for small edits encoded by short RTTs. As used herein, "short RTTs" or "small RTT class of mutants" refers to the group of MMLV mutants that improve prime editing when the pegRNA has a short RT template (RTT or RT template). Other mutations were helpful for long RTT edits, such as collapsing the CAG expansion for HTT and doing some twinPE edits. Starting with the small RTT class of mutants, 13 mutations evolved or engineered in M-MLV improved editing, typically from 1.1 fold to 1.3 fold relative to PE2 (FIG. 54). Importantly, these mutants are all truncated M-MLV variants, lacking their RNaseH domain. The presence or absence of the RNAseH domain effected different mammalian edits differently: in general, it was either equivalent or better than FL PE2 for short edits, but also caused a decrease in editing for longer RTT edits.

[0712] A different group of mutants did not help with short RTT edits, but they did help with long RTT edits, such as correction of the CAG expansion that causes Huntington's disease, and some twinPE edits. All of our mutants are truncated (lacking an RNaseH domain) because it was seen that truncation improved editing for the mutants, and was better for delivery purposes. When truncated mutants were compared to full-length PE2 in HEK293T

cells, there was a small improvement from these mutants on long RTT edits (FIG. 55A). Additionally, there was improvements see relative to the WT truncated enzyme (FIG. 55B). At sites like these, the truncated PE2 enzyme performed worse than WT. The truncated mutants recovered this activity.

[0713] Additional PACE- and PANCE-evolved and engineered Cas9 mutants were identified that improved mammalian prime editing. The results of the evolution procedures and subsequent mammalian characterizations showed that the target edit used in an evolution greatly influenced the outputs of that evolution, and a given mutation's effect in mammalian cells varied between target edits (FIG. 56). To use these insights to maximize the therapeutic potential of PE-PACE, a disease-specific circuit was developed that selected for correction of the precise DNA sequence that causes the majority of Tay-Sachs disease (TSD): the HEXA 1278insTATC mutation. To create this PACE circuit (TSD-PACE), a fragment of the pathogenic human HEXA allele was inserted into an otherwise wild-type T7RNAP gene. The insertion was positioned to occur at residue 601 of T7 RNAP protein which is the residue at the center of a disordered loop on the T7RNAP that has previously been manipulated for splitting T7RNAP and other applications. If the inserted HEXA fragment harbored the frameshifting TSD allele, then it frameshifted the remainder of the T7 RNAP gene downstream, leading to an inactive enzyme. However, if the TSD mutation was correctly repaired by prime editing, the frame of the HEXA-T7RNAP fusion was restored, which enabled gIII transcription and phage propagation (FIG. 57A-57C).

[0714] A PANCE campaign was initiated to evolve compact Ec48 and Gs RTs specifically on the TSD mutation. Sequencing of both of these RTs revealed unique mutations that were not enriched in previous selections. To evaluate the impact of the TSD-PANCE mutations in mammalian cells, the newly-evolved editors were tested as well as the WT enzymes and other variants that were produced in a HEK293T cell line that had previously been manipulated to harbor the 1278TATCins mutation. Mutations from the disease-specific evolution further improved activity over generalist-evolved counterparts. Specifically, disease-specific evolution allowed Ec48 to reach PE2 levels of editing (FIG. 58A). Additionally, the outputs of a very low-stringency, disease-specific PANCE evolution of Gs RT outperformed Gs RT variants that were evolved in a high-stringency PACE on a different target (FIG. 58B).

EQUIVALENTS AND SCOPE

[0715] In the articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Embodiments or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

[0716] Furthermore, the disclosure encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claims that is dependent on the same base claim. Where elements are presented as lists, *e.g.*, in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should it be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the disclosure or aspects of the disclosure consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth *in haec verba* herein. It is also noted that the terms "comprising" and "containing" are intended to be open and permits the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub–range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[0717] This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the embodiments. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any

particular embodiment of the invention can be excluded from any embodiment, for any reason, whether or not related to the existence of prior art.

[0718] Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended embodiments. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present invention, as defined in the following embodiments.

CLAIMS

What is claimed is:

1. A prime editor comprising a nucleic acid-programmable DNA-binding protein (napDNAbp) and a mouse mammary tumor virus (MMTV) reverse transcriptase or a variant thereof, an avian sarcoma leukosis virus (ASLV) reverse transcriptase or a variant thereof, a porcine endogenous retrovirus (PERV) reverse transcriptase or a variant thereof, an HIV-MMLV reverse transcriptase or a variant thereof, an AVIRE reverse transcriptase or a variant thereof, a baboon endogenous virus (BAEVM) reverse transcriptase or a variant thereof, a gibbon ape leukemia virus (GALV) reverse transcriptase or a variant thereof, a koala retrovirus (KORV) reverse transcriptase or a variant thereof, a Mason-Pfizer monkey virus (MPMV) reverse transcriptase or a variant thereof, a POK11ERV reverse transcriptase or a variant thereof, a simian retrovirus type 2 (SRV2) reverse transcriptase or a variant thereof, a woolly monkey sarcoma virus (WMSV) reverse transcriptase or a variant thereof, a Vp96 reverse transcriptase or a variant thereof, a Vc95 reverse transcriptase or a variant thereof, an Ec48 reverse transcriptase or a variant thereof, a Gs reverse transcriptase or a variant thereof, an Er reverse transcriptase or a variant thereof, an Ne144 reverse transcriptase or a variant thereof, a Tf1 reverse transcriptase or a variant thereof, or an Rs09415 reverse transcriptase ("CRISPR-RT") or a variant thereof.

2. The prime editor of claim 1, wherein the prime editor comprises an AVIRE reverse transcriptase of SEQ ID NO: 216, or an AVIRE reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 216, wherein the AVIRE reverse transcriptase variant comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, G329P, and L604W.

3. The prime editor of claim 2, wherein the AVIRE reverse transcriptase variant comprises the mutation D199N.

4. The prime editor of claim 2 or 3, wherein the AVIRE reverse transcriptase variant comprises the mutation T305K.

5. The prime editor of any one of claims 2-4, wherein the AVIRE reverse transcriptase variant comprises the mutation W312F.

6. The prime editor of any one of claims 2-5, wherein AVIRE reverse transcriptase variant comprises the mutation G329P.

7. The prime editor of any one of claims 2-6, wherein the AVIRE reverse transcriptase variant comprises the mutation L604W.

8. The prime editor of any one of claims 2-7, wherein the AVIRE reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 217-221, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 217-221, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 604W.

9. The prime editor of any one of claims 2-7, wherein the AVIRE reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 243, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 243, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 604W.

10. The prime editor of claim 1, wherein the prime editor comprises a KORV reverse transcriptase of SEQ ID NO: 222, or a KORV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 222, wherein the KORV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W310F, E327P, and L599W.

11. The prime editor of claim 10, wherein the KORV reverse transcriptase variant comprises the mutation D197N.

12. The prime editor of claim 10 or 11, wherein the KORV reverse transcriptase variant comprises the mutation T303K.

13. The prime editor of any one of claims 10-12, wherein the KORV reverse transcriptase variant comprises the mutation W310F.

14. The prime editor of any one of claims 10-13, wherein KORV reverse transcriptase variant comprises the mutation E327P.

15. The prime editor of any one of claims 10-14, wherein the KORV reverse transcriptase variant comprises the mutation L599W.

16. The prime editor of any one of claims 10-15, wherein the KORV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 223-227, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 223-227, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 310F, 327P, and 599W.

17. The prime editor of any one of claims 10-15, wherein the KORV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 244, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 244, wherein the amino acid sequence comprises the residues 197N, 303K, 310F, 327P, and 599W.

18. The prime editor of claim 1, wherein the prime editor comprises a WMSV reverse transcriptase of SEQ ID NO: 228, or a WMSV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 228, wherein the WMSV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W311F, E327P, and L599W.

19. The prime editor of claim 18, wherein the WMSV reverse transcriptase variant comprises the mutation D197N.

20. The prime editor of claim 18 or 19, wherein the WMSV reverse transcriptase variant comprises the mutation T303K.

21. The prime editor of any one of claims 18-20, wherein the WMSV reverse transcriptase variant comprises the mutation W311F.

22. The prime editor of any one of claims 18-21, wherein WMSV reverse transcriptase variant comprises the mutation E327P.

23. The prime editor of any one of claims 18-22, wherein the WMSV reverse transcriptase variant comprises the mutation L599W.

24. The prime editor of any one of claims 18-23, wherein the WMSV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 229-233, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 229-233, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 311F, 327P, and 599W.

25. The prime editor of any one of claims 18-23, wherein the WMSV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 245, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 245, wherein the amino acid sequence comprises the residues 197N, 303K, 311F, 327P, and 599W.

26. The prime editor of claim 1, wherein the prime editor comprises a PERV reverse transcriptase of SEQ ID NO: 45, or a PERV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 45, wherein the PERV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, E329P, and L602W.

27. The prime editor of claim 26, wherein the PERV reverse transcriptase variant comprises the mutation D199N.

28. The prime editor of claim 26 or 27, wherein the PERV reverse transcriptase variant comprises the mutation T305K.

29. The prime editor of any one of claims 26-28, wherein the PERV reverse transcriptase variant comprises the mutation W312F.

30. The prime editor of any one of claims 26-29, wherein PERV reverse transcriptase variant comprises the mutation E329P.

31. The prime editor of any one of claims 26-30, wherein the PERV reverse transcriptase variant comprises the mutation L602W.

32. The prime editor of any one of claims 26-31, wherein the PERV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 214 and 234-238, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 214 and 234-238, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 602W.

33. The prime editor of any one of claims 26-31, wherein the PERV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 215, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 215, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 602W.

34. The prime editor of claim 1, wherein the prime editor comprises a Tf1 reverse transcriptase of SEQ ID NO: 55, or a Tf1 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 55, wherein the Tf1 reverse transcriptase variant comprises one or more mutations selected from the group consisting of V14A, E22K, I64L, I64W, P70T, G72V, M102I, K106R, K118R, L133N, A139T, L158Q, S188K, I260L, F269L, E274R, R288Q, Q293K, S297Q, N316Q, K321R, K356E, A363V, K413E, I423V, and S492N relative to SEQ ID NO: 55.

35. The prime editor of claim 34, wherein the Tf1 reverse transcriptase variant comprises an I64L mutation, an I64W mutation, a K118R mutation, an L133N mutation, an S188K mutation, an I260L mutation, an E274R mutation, an R288Q mutation, a Q293K mutation, an S297Q mutation, an N316Q mutation, or a K321R mutation.

36. The prime editor of claim 34, wherein the Tf1 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 55:

K118R and S297Q;

V14A, L158Q, F269L, and K356E;

E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, and S492N;

P70T, G72V, M102I, K106R, L158Q, F269L, A363V, K413E, and S492N;

K106R, L158Q, F269L, A363V, and I423V;

K118R, S297Q, S188K, I64L, I260L, and R288Q;

E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, S492N, K118R, S297Q, S188K, I64L, and I260L;

K118R and S188K;

K118R, S188K, and I260L;

K118R, S188K, I260L, and S297Q; or

K118R, S188K, I260L, R288K, and S297Q.

37. The prime editor of any one of claims 34-36, wherein the Tf1 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 196-213 and 251-255, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 196-213 and 251-255, wherein the amino acid sequence comprises at least one of residues 14A, 22K, 64L, 64W, 70T, 72V, 102I, 106R, 118R, 133N, 139T, 158Q, 188K, 260L, 269L, 274R, 288Q, 293K, 297Q, 316Q, 321R, 356E, 363V, 413E, 423V, and 492N.

38. The prime editor of claim 1, wherein the prime editor comprises an Ec48 reverse transcriptase of SEQ ID NO: 59, or an Ec48 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 59, wherein the Ec48 reverse transcriptase variant comprises one or more mutations selected from the group consisting of A36V, E54K, E60K, K87E, S151T, E165D, L182N, T189N, R205K, V214L, D243N, R267I, S277F, E279K, V303M, K307R, R315K, N317S, K318E, H324Q, K326E, E328K, K343N, R372K, R378K, and T385R relative to SEQ ID NO: 59.

39. The prime editor of claim 38, wherein the Ec48 reverse transcriptase variant comprises an L182N mutation, a T189N mutation, a K307R mutation, an R315K mutation, an R378K mutation, or a T385R mutation.

40. The prime editor of claim 39, wherein the Ec48 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 59:

R267I, K318E, K326E, E328K, and R372K;

K87E, R205K, V214L, D243N, R267I, N317S, K318E, H324Q, and K326E;

E54K, K87E, D243N, R267I, E279K, and K318E;

A36V, K87E, R205K, D243N, R267I, E279K, and K318E;

E54K, K87E, D243N, R267I, E279K, and K318E;

E54K, K87E, D243N, R267I, S277F, E279K, and K318E;

E60K, K87E, E165D, D243N, R267I, E279K, K318E, and K343N;

E60K, K87E, S151T, E165D, D243N, R267I, E279K, V303M, K318E, and K343N;

or

R315K, L182N, and T189N.

41. The prime editor of any one of claims 38-40, wherein the Ec48 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 188-195, 256, and 257, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 188-195, 256, and 257, wherein the amino acid sequence comprises at least one of residues 36V, 54K, 60K, 87E, 151T, 165D, 182N, 189N, 205K, 214L, 243N, 267I, 277F, 279K, 303M, 307R, 315K, 317S, 318E, 324Q, 326E, 328K, 343N, 372K, 378K, and 385R.

42. The prime editor of claim 1, wherein the prime editor comprises an Ne144 reverse transcriptase of SEQ ID NO: 239, or an Ne144 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 239, wherein the Ne144 reverse transcriptase variant comprises one or more mutations selected from the group consisting of A157T, A165T, and G288V relative to SEQ ID NO: 239.

43. The prime editor of claim 42, wherein the Ne144 reverse transcriptase variant comprises the mutations A157T, A165T, and G288V.

44. The prime editor of claim 42 or 43, wherein the Ne144 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 240, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 240, wherein the amino acid sequence comprises at least one of residues 157T, 165T, and 288V.

45. The prime editor of claim 1, wherein the prime editor comprises a Vc95 reverse transcriptase of SEQ ID NO: 241, or a Vc95 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 241, wherein the Vc95 reverse transcriptase variant comprises one or more mutations selected from the group consisting of L11M, S75A, V97M, N146D, and N245T relative to SEQ ID NO: 241.

46. The prime editor of claim 45, wherein the Vc95 reverse transcriptase variant comprises the mutations L11M, S75A, V97M, N146D, and N245T.

47. The prime editor of claim 45 or 46, wherein the Vc95 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 242, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 242, wherein the amino acid sequence comprises at least one of residues 11M, 75A, 97M, 146D, and 245T.

48. The prime editor of claim 1, wherein the prime editor comprises a Gs reverse transcriptase of SEQ ID NO: 60, or a Gs reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 60, wherein the Gs reverse transcriptase variant comprises one or more mutations selected from the group consisting of N12D, A16E, A16V, L17P, V20G, L37R, L37P, R38H, Y40C, I41N, I41S, W45R, I67T, I67R, G72E, G73V, G78V, Q93R, A123V, Y126F, E129G, K162N, P190L, D206V, R233K, A234V, R263G, P264S, R267M, K279E, R287I, R291K, P309T, R344S, R358S, R360S, E363G, V374A, and Q412H relative to SEQ ID NO: 60.

49. The prime editor of claim 48, wherein the Gs reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 60:

L17P and D206V;

N12D, L37R, and G78V;

A16E, L37P, and A123V;

A16V, R38H, W45R, Y126F, and Q412H;

A16V, R38H, W45R, and R291K;

N12D, L37R, G72E, E129G, P264S, R344S, and R360S;

N12D, Y40C, I67T, G73V, Q93R, R287I, and R358S;

N12D, Y40C, I67T, G73V, Q93R, and R358S;

N12D, I41N, P190L, A234V, and K279E;

N12D, L37R, R267M, P309T, R358S, and E363G;

A16V, V20G, I41S, R233K, and P264S;

L17P, V20G, I41S, I67R, R263G, P264S, and V374A; or

L17P, V20G, I41S, I67R, K162N, R263G, and P264S.

50. The prime editor of claim 48 or 49, wherein the Gs reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 159-171, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 159-171, wherein the amino acid sequence comprises at least one of residues 12D, 16E, 16V, 17P, 20G, 37R, 37P, 38H, 40C, 41N, 41S, 45R, 67T, 67R, 72E, 73V, 78V, 93R, 123V, 126F, 129G, 162N, 190L, 206V, 233K, 234V, 263G, 264S, 267M, 279E, 287I, 291K, 309T, 344S, 358S, 360S, 363G, 374A, and 412H.

51. A prime editor comprising a nucleic acid-programmable DNA-binding protein (napDNAbp) and an MMLV reverse transcriptase variant comprising one or more mutations relative to SEQ ID NO: 33 selected from the group consisting of T13I, V19I, A32T, G38V, S60Y, P111L, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, G138R, C157F, P175Q, P175S, D200S, D200Y, D200C, Y222F, V223A, V223M, V223T, V223W, V223Y, L234I, T246I, N249S, T287A, P292T, E302A, E302K, G316R, E346K, K373N, W388C, V402A, K445N, M457I, and A462S.

52. The prime editor of claim 51, wherein the MMLV reverse transcriptase variant comprises a single mutation relative to SEQ ID NO: 33 selected from the group consisting of T13I, G38V, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, P175Q, P175S, D200C, D200Y, V223M, V223T, V223W, V223Y, L234I, P292T, G316R, K373N, M457I, and V402A.

53. The prime editor of claim 51, wherein the MMLV reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 33:

D200Y and E302A;
D200Y, V223A, and M457I;
V223M, T306K, and A462S;
D200N and E302K;
D200Y and E302K;
T128N and V223A;
V19I, A32T, and D200Y;
D200S, V223A, E346K, and W388C;
S60Y, V223A, and N249S;
P111L, V223A, T287A, and G316R;
S60Y, G138R, and V223A;
S60Y, Y222F, V223A, and K445N; or
S60Y, C157F, V223A, and T246I.

54. The prime editor of any one of claims 51-53, wherein the MMLV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, wherein the amino acid sequence comprises at least one of residues 13I, 19I, 32T, 38V, 60Y, 111L, 120R, 126Y, 128N, 128F, 128H, 129S, 132S, 138R, 157F, 175Q, 175S, 200S, 200Y, 200C, 222F, 223A, 223M, 223T, 223W, 223Y, 234I, 246I, 249S, 287A, 292T, 302A, 302K, 316R, 346K, 373N, 388C, 402A, 445N, 457I, and 462S.

55. The prime editor of any one of claims 1-54, wherein the napDNAbp is a Cas protein.

56. The prime editor of any one of claims 1-55, wherein the napDNAbp is a Cas9 nickase (nCas9) or a nuclease-inactive Cas9 (dCas9).

57. The prime editor of any one of claims 1-56, wherein the napDNAbp comprises the amino acid sequence of any one of SEQ ID NOs: 9-32, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 9-32.

58. The prime editor of any one of claims 1-57, wherein the napDNAbp comprises a Cas9 variant comprising one or more mutations relative to SEQ ID NO: 9 or SEQ ID NO: 11 selected from the group consisting of D23G, H99Q, H99R, E102K, E102S, E102R, N175K, D177G, K218R, N309D, I312V, E471K, G485S, K562N, D608N, I632V, D645N, D645E, R654C, G687D, G715E, H721Y, R753K, R753G, H754R, K775R, E790K, T804A, K918A, K1003R, M1021Y, E1071K, and E1260D.

59. The prime editor of claim 58, wherein the Cas9 variant comprises an R753G mutation.

60. The prime editor of claim 58, wherein the Cas9 variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 11:

H721Y and R753G;

E102K and R753G; and

E102K, H721Y, and R753G.

61. The prime editor of claim 58, wherein the Cas9 variant comprises the amino acid sequence of any one of SEQ ID NOs: 178-180.

62. An AVIRE reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 216, wherein the AVIRE reverse transcriptase variant comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, G329P, and L604W.

63. The AVIRE reverse transcriptase variant of claim 62, wherein the AVIRE reverse transcriptase variant comprises the mutation D199N.

64. The AVIRE reverse transcriptase variant of claim 62 or 63, wherein the AVIRE reverse transcriptase variant comprises the mutation T305K.

65. The AVIRE reverse transcriptase variant of any one of claims 62-64, wherein the AVIRE reverse transcriptase variant comprises the mutation W312F.

66. The AVIRE reverse transcriptase variant of any one of claims 62-65, wherein AVIRE reverse transcriptase variant comprises the mutation G329P.

67. The AVIRE reverse transcriptase variant of any one of claims 62-66, wherein the AVIRE reverse transcriptase variant comprises the mutation L604W.

68. The AVIRE reverse transcriptase variant of any one of claims 62-67, wherein the AVIRE reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 217-221, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 217-221, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 604W.

69. The AVIRE reverse transcriptase variant of any one of claims 62-67, wherein the AVIRE reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 243, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 243, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 604W.

70. A KORV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 222, wherein the KORV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W310F, E327P, and L599W.

71. The KORV reverse transcriptase variant of claim 70, wherein the KORV reverse transcriptase variant comprises the mutation D197N.

72. The KORV reverse transcriptase variant of claim 70 or 71, wherein the KORV reverse transcriptase variant comprises the mutation T303K.

73. The KORV reverse transcriptase variant of any one of claims 70-72, wherein the KORV reverse transcriptase variant comprises the mutation W310F.

74. The KORV reverse transcriptase variant of any one of claims 70-73, wherein KORV reverse transcriptase variant comprises the mutation E327P.

75. The KORV reverse transcriptase variant of any one of claims 70-74, wherein the KORV reverse transcriptase variant comprises the mutation L599W.

76. The KORV reverse transcriptase variant of any one of claims 70-75, wherein the KORV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 223-227, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 223-227, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 310F, 327P, and 599W.

77. The KORV reverse transcriptase variant of any one of claims 70-75, wherein the KORV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 244,

or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 244, wherein the amino acid sequence comprises the residues 197N, 303K, 310F, 327P, and 599W.

78. A WMSV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 228, wherein the WMSV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D197N, T303K, W311F, E327P, and L599W.

79. The WMSV reverse transcriptase variant of claim 78, wherein the WMSV reverse transcriptase variant comprises the mutation D197N.

80. The WMSV reverse transcriptase variant of claim 78 or 79, wherein the WMSV reverse transcriptase variant comprises the mutation T303K.

81. The WMSV reverse transcriptase variant of any one of claims 78-80, wherein the WMSV reverse transcriptase variant comprises the mutation W311F.

82. The WMSV reverse transcriptase variant of any one of claims 78-81, wherein WMSV reverse transcriptase variant comprises the mutation E327P.

83. The WMSV reverse transcriptase variant of any one of claims 78-82, wherein the WMSV reverse transcriptase variant comprises the mutation L599W.

84. The WMSV reverse transcriptase variant of any one of claims 78-83, wherein the WMSV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 229-233, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 229-233, wherein the amino acid sequence comprises at least one of the residues 197N, 303K, 311F, 327P, and 599W.

85. The WMSV reverse transcriptase variant of any one of claims 78-83, wherein the WMSV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 245, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 245, wherein the amino acid sequence comprises the residues 197N, 303K, 311F, 327P, and 599W.

86. A PERV reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 45, wherein the PERV reverse transcriptase variant comprises one or more mutations selected from the group consisting of D199N, T305K, W312F, E329P, and L602W.

87. The PERV reverse transcriptase variant of claim 86, wherein the PERV reverse transcriptase variant comprises the mutation D199N.

88. The PERV reverse transcriptase variant of claim 86 or 87, wherein the PERV reverse transcriptase variant comprises the mutation T305K.

89. The PERV reverse transcriptase variant of any one of claims 86-88, wherein the PERV reverse transcriptase variant comprises the mutation W312F.

90. The PERV reverse transcriptase variant of any one of claims 86-89, wherein PERV reverse transcriptase variant comprises the mutation E329P.

91. The PERV reverse transcriptase variant of any one of claims 86-90, wherein the PERV reverse transcriptase variant comprises the mutation L602W.

92. The PERV reverse transcriptase variant of any one of claims 86-91, wherein the PERV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 214 and 234-238, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 214 and 234-238, wherein the amino acid sequence comprises at least one of the residues 199N, 305K, 312F, 329P, and 602W.

93. The PERV reverse transcriptase variant of any one of claims 86-91, wherein the PERV reverse transcriptase variant comprises an amino acid sequence of SEQ ID NO: 215, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 215, wherein the amino acid sequence comprises the residues 199N, 305K, 312F, 329P, and 602W.

94. A Tf1 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 55, wherein the Tf1 reverse transcriptase variant comprises one or more mutations selected from the group consisting of V14A, E22K, I64L, I64W, P70T, G72V, M102I, K106R, K118R, L133N, A139T, L158Q, S188K, I260L, F269L, E274R, R288Q, Q293K, S297Q, N316Q, K321R, K356E, A363V, K413E, I423V, and S492N relative to SEQ ID NO: 55.

95. The Tf1 reverse transcriptase variant of claim 94, wherein the Tf1 reverse transcriptase variant comprises an I64L mutation, an I64W mutation, a K118R mutation, an L133N mutation, an S188K mutation, an I260L mutation, an E274R mutation, an R288Q mutation, a Q293K mutation, an S297Q mutation, an N316Q mutation, or an K321R mutation.

96. The Tf1 reverse transcriptase variant of claim 94, wherein the Tf1 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 55:

K118R and S297Q;

V14A, L158Q, F269L, and K356E;

E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, and S492N;

P70T, G72V, M102I, K106R, L158Q, F269L, A363V, K413E, and S492N;

K106R, L158Q, F269L, A363V, and I423V;

K118R, S297Q, S188K, I64L, I260L, and R288Q;

E22K, P70T, G72V, M102I, K106R, A139T, L158Q, F269L, A363V, K413E, S492N, K118R, S297Q, S188K, I64L, and I260L;

K118R and S188K;
K118R, S188K, and I260L;
K118R, S188K, I260L, and S297Q; or
K118R, S188K, I260L, R288K, and S297Q.

97. The Tf1 reverse transcriptase variant of any one of claims 94-96, wherein the Tf1 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 196-213 and 251-255, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 196-213 and 251-255, wherein the amino acid sequence comprises at least one of residues 14A, 22K, 64L, 64W, 70T, 72V, 102I, 106R, 118R, 133N, 139T, 158Q, 188K, 260L, 269L, 274R, 288Q, 293K, 297Q, 316Q, 321R, 356E, 363V, 413E, 423V, and 492N.

98. An Ec48 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 59, wherein the Ec48 reverse transcriptase variant comprises one or more mutations selected from the group consisting of A36V, E54K, E60K, K87E, S151T, E165D, L182N, T189N, R205K, V214L, D243N, R267I, S277F, E279K, V303M, K307R, R315K, N317S, K318E, H324Q, K326E, E328K, K343N, R372K, R378K, and T385R relative to SEQ ID NO: 59.

99. The Ec48 reverse transcriptase variant of claim 98, wherein the Ec48 reverse transcriptase variant comprises an L182N mutation, a T189N mutation, a K307R mutation, an R315K mutation, an R378K mutation, or a T385R mutation.

100. The Ec48 reverse transcriptase variant of claim 99, wherein the Ec48 reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 59:
R267I, K318E, K326E, E328K, and R372K;
K87E, R205K, V214L, D243N, R267I, N317S, K318E, H324Q, and K326E;
E54K, K87E, D243N, R267I, E279K, and K318E;
A36V, K87E, R205K, D243N, R267I, E279K, and K318E;
E54K, K87E, D243N, R267I, E279K, and K318E;

E54K, K87E, D243N, R267I, S277F, E279K, and K318E;

E60K, K87E, E165D, D243N, R267I, E279K, K318E, and K343N;

E60K, K87E, S151T, E165D, D243N, R267I, E279K, V303M, K318E, and K343N;

or

R315K, L182N, and T189N.

101. The Ec48 reverse transcriptase variant of any one of claims 98-100, wherein the Ec48 reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 188-195, 256, and 257, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 188-195, 256, and 257, wherein the amino acid sequence comprises at least one of residues 36V, 54K, 60K, 87E, 151T, 165D, 182N, 189N, 205K, 214L, 243N, 267I, 277F, 279K, 303M, 307R, 315K, 317S, 318E, 324Q, 326E, 328K, 343N, 372K, 378K, and 385R.

102. An Ne144 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 239, wherein the Ne144 reverse transcriptase variant comprises one or more mutations selected from the group consisting of A157T, A165T, and G288V relative to SEQ ID NO: 239.

103. The Ne144 reverse transcriptase variant of claim 102, wherein the Ne144 reverse transcriptase variant comprises the mutations A157T, A165T, and G288V.

104. The Ne144 reverse transcriptase variant of claim 102 or 103, wherein the Ne144 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 240, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 240, wherein the amino acid sequence comprises at least one of residues 157T, 165T, and 288V.

105. A Vc95 reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 241, wherein the Vc95 reverse transcriptase variant

comprises one or more mutations selected from the group consisting of L11M, S75A, V97M, N146D, and N245T relative to SEQ ID NO: 241.

106. The Vc95 reverse transcriptase variant of claim 105, wherein the Vc95 reverse transcriptase variant comprises the mutations L11M, S75A, V97M, N146D, and N245T.

107. The Vc95 reverse transcriptase variant of claim 105 or 106, wherein the Vc95 reverse transcriptase variant comprises the amino acid sequence of SEQ ID NO: 242, or an amino acid sequence 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%, or at least 99% identical to SEQ ID NO: 242, wherein the amino acid sequence comprises at least one of residues 11M, 75A, 97M, 146D, and 245T.

108. A Gs reverse transcriptase variant having 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%, or at least 99% sequence identity with SEQ ID NO: 60, wherein the Gs reverse transcriptase variant comprises one or more mutations selected from the group consisting of N12D, A16E, A16V, L17P, V20G, L37R, L37P, R38H, Y40C, I41N, I41S, W45R, I67T, I67R, G72E, G73V, G78V, Q93R, A123V, Y126F, E129G, K162N, P190L, D206V, R233K, A234V, R263G, P264S, R267M, K279E, R287I, R291K, P309T, R344S, R358S, R360S, E363G, V374A, and Q412H relative to SEQ ID NO: 60.

109. The Gs reverse transcriptase variant of claim 108, wherein the Gs reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 60:

L17P and D206V;

N12D, L37R, and G78V;

A16E, L37P, and A123V;

A16V, R38H, W45R, Y126F, and Q412H;

A16V, R38H, W45R, and R291K;

N12D, L37R, G72E, E129G, P264S, R344S, and R360S;

N12D, Y40C, I67T, G73V, Q93R, R287I, and R358S;

N12D, Y40C, I67T, G73V, Q93R, and R358S;

N12D, I41N, P190L, A234V, and K279E;

N12D, L37R, R267M, P309T, R358S, and E363G;
A16V, V20G, I41S, R233K, and P264S;
L17P, V20G, I41S, I67R, R263G, P264S, and V374A; or
L17P, V20G, I41S, I67R, K162N, R263G, and P264S.

110. The Gs reverse transcriptase variant of claim 108 or 109, wherein the Gs reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 159-171, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 159-171, wherein the amino acid sequence comprises at least one of residues 12D, 16E, 16V, 17P, 20G, 37R, 37P, 38H, 40C, 41N, 41S, 45R, 67T, 67R, 72E, 73V, 78V, 93R, 123V, 126F, 129G, 162N, 190L, 206V, 233K, 234V, 263G, 264S, 267M, 279E, 287I, 291K, 309T, 344S, 358S, 360S, 363G, 374A, and 412H.

111. An MMLV reverse transcriptase variant comprising one or more mutations relative to SEQ ID NO: 33 selected from the group consisting of T13I, V19I, A32T, G38V, S60Y, P111L, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, G138R, C157F, P175Q, P175S, D200S, D200Y, D200C, Y222F, V223A, V223M, V223T, V223W, V223Y, L234I, T246I, N249S, T287A, P292T, E302A, E302K, G316R, E346K, K373N, W388C, V402A, K445N, M457I, and A462S.

112. The MMLV reverse transcriptase variant of claim 111, wherein the MMLV reverse transcriptase variant comprises a single mutation relative to SEQ ID NO: 33 selected from the group consisting of T13I, G38V, K120R, H126Y, T128N, T128F, T128H, V129S, P132S, P175Q, P175S, D200C, D200Y, V223M, V223T, V223W, V223Y, L234I, P292T, G316R, K373N, M457I, and V402A.

113. The MMLV reverse transcriptase variant of claim 111, wherein the MMLV reverse transcriptase variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 33:
D200Y and E302A;
D200Y, V223A, and M457I;
V223M, T306K, and A462S;

D200N and E302K;
D200Y and E302K;
T128N and V223A;
V19I, A32T, and D200Y;
D200S, V223A, E346K, and W388C;
S60Y, V223A, and N249S;
P111L, V223A, T287A, and G316R;
S60Y, G138R, and V223A;
S60Y, Y222F, V223A, and K445N; or
S60Y, C157F, V223A, and T246I.

114. The MMLV reverse transcriptase variant of any one of claims 111-113, wherein the MMLV reverse transcriptase variant comprises the amino acid sequence of any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 35-42, 172-177, 183, and 184, wherein the amino acid sequence comprises at least one of residues 13I, 19I, 32T, 38V, 60Y, 111L, 120R, 126Y, 128N, 128F, 128H, 129S, 132S, 138R, 157F, 175Q, 175S, 200S, 200Y, 200C, 222F, 223A, 223M, 223T, 223W, 223Y, 234I, 246I, 249S, 287A, 292T, 302A, 302K, 316R, 346K, 373N, 388C, 402A, 445N, 457I, and 462S.

115. A Cas9 variant comprising one or more mutations relative to SEQ ID NO: 9 or SEQ ID NO: 11 selected from the group consisting of D23G, H99Q, H99R, E102K, E102S, E102R, N175K, D177G, K218R, N309D, I312V, E471K, G485S, K562N, D608N, I632V, D645N, D645E, R654C, G687D, G715E, H721Y, R753K, R753G, H754R, K775R, E790K, T804A, K918A, K1003R, M1021Y, E1071K, and E1260D.

116. The Cas9 variant of claim 115, wherein the Cas9 variant comprises an R753G mutation.

117. The Cas9 variant of claim 115, wherein the Cas9 variant comprises any one of the following groups of mutations relative to the amino acid sequence of SEQ ID NO: 9 or SEQ ID NO: 11:

H721Y and R753G;

E102K and R753G; and

E102K, H721Y, and R753G.

118. The Cas9 variant of claim 115, wherein the Cas9 variant comprises the amino acid sequence of any one of SEQ ID NOs: 178-180.

119. A prime editor comprising the Cas9 variant of any one of claims 115-118 and a reverse transcriptase.

120. The prime editor of claim 119, wherein the reverse transcriptase comprises the amino acid sequence of any one of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241, or an amino acid sequence 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%, or at least 99% identical to any one of SEQ ID NOs: 33-46, 48, 49, 51-53, 55-57, 59, 60, 63-78, 185, 216, 222, 228, 239, and 241.

121. The prime editor of claim 119, wherein the reverse transcriptase is a reverse transcriptase of any one of claims 62-114.

122. The prime editor of any one of claims 1-61 or 119-121, wherein the napDNAbp and the reverse transcriptase are provided *in trans* or are not fused to one another.

123. The prime editor of any one of claims 1-61 or 119-121, wherein the napDNAbp and the reverse transcriptase are provided as a fusion protein or are fused to one another.

124. The prime editor of claim 123, wherein the napDNAbp and the reverse transcriptase are fused via a linker.

125. The prime editor of claim 124, wherein the linker comprises any one of SEQ ID NOs: 79-93.

126. The prime editor of any one of claims 1-61 or 119-125 further comprising a nuclear localization sequence (NLS).

127. The prime editor of any one of claims 123-126, wherein the fusion protein comprises the amino acid sequence of any one of SEQ ID NOs: 246-250.

128. A complex comprising the prime editor of any one of claims 1-61 or 119-127 and a PEgRNA.

129. The complex of claim 128, wherein the PEgRNA comprises a guide RNA and a nucleic acid extension arm at the 3′ or 5′ end of the guide RNA.

130. The complex of claim 128 or 129, wherein the PEgRNA is capable of binding to the napDNAbp and directing the napDNAbp to a target DNA sequence.

131. One or more polynucleotides encoding the prime editor of any one of claims 1-61 or 119-127.

132. A vector comprising the one or more polynucleotides of claim 131.

133. A cell comprising a prime editor of any one of claims 1-61 or 119-127, a complex of any one of claims 128-130, the one or more polynucleotides of claim 131, or the vector of claim 132.

134. A pharmaceutical composition comprising a prime editor of any of claims 1-61 or 119-127, a complex of any one of claims 128-130, the one or more polynucleotides of claim 131, or the vector of claim 132.

135. A method for editing a nucleic acid molecule by prime editing comprising contacting a nucleic acid molecule with a prime editor of any one of claims 1-61 or 119-127 or a complex of any one of claims 128-130, thereby installing one or more modifications to the nucleic acid molecule at a target site.

136. The method of claim 135, wherein the method further comprises contacting the nucleic acid molecule with a second