

REVERSE TRANSCRIPTION-MEDIATED GENE EDITING SYSTEMS AND USES THEREOF

5 CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of the filing dates of U.S. Provisional Application No. 63/580,188, filed September 1, 2023, U.S. Provisional Application No. 63/580,168, filed September 1, 2023, U.S. Provisional Application No. 63/553,974, filed February 15, 2024; U.S. Provisional Application No. 63/638,559, filed April 25, 2024. Each of the priority
10 applications is incorporated by reference herein in their entities.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been filed electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created
15 on August 28, 2024, is named 063586-525001WO_SeqList_ST26.xml and is 597.0 kilobytes in size.

BACKGROUND

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) genes, collectively known as CRISPR-Cas or CRISPR/Cas systems, are
20 adaptive immune systems in archaea and bacteria that defend particular species against foreign genetic elements.

Reverse transcriptases (RTs) are enzymes that generate a strand of DNA that is complementary to an RNA template. The combination of reverse transcriptases and CRISPR/Cas
25 systems has shown great potentials in genetic editing. CRISPR-guided reverse transcription allows for introduction of desired nucleotide substitutions at a genomic site.

It is therefore of great interest to develop efficient and accurate reverse transcriptase-CRISPR gene editing systems for use in disease treatment.

30 SUMMARY OF THE PRESENT DISCLOSURE

The present disclosure provides reverse transcriptase-CRISPR-mediated gene editing systems, which successfully introduced designed nucleotide substitutions into target genetic sites. In some embodiments, the gene editing systems disclosed herein involve a fusion polypeptide comprising a CRISPR nuclease fragment and a reverse transcriptase (RT) fragment,
35 and optionally one or more nuclear localization sequences (NLS) and/or peptide linkers. The

CRISPR nuclease polypeptide can be genetically engineered to possess advantageous enzymatic activities (*e.g.*, high indel activities and/or DNA cleavage activities and precise gene editing as designed). Accordingly, the gene editing systems provided herein would be expected to show superior effectiveness in inserting desired base substitutions at a genomic site of interest.

5 Accordingly, one aspect of the present disclosure features a gene editing system comprising: (a) a fusion polypeptide comprising a CRISPR nuclease polypeptide and a reverse transcriptase (RT) polypeptide, or a first nucleic acid encoding the fusion polypeptide; and (b) an RNA molecule comprising a guide RNA (gRNA) and a reverse transcription donor RNA (RT donor RNA), or a second nucleic acid encoding the RNA molecule. The CRISPR nuclease may
10 be the reference nuclease of SEQ ID NO: 1 or a variant thereof. The gRNA comprises a scaffold sequence recognizable by the CRISPR nuclease and a spacer sequence specific to a target sequence within a genomic site of interest, the target sequence being upstream to a protospacer adjacent motif (PAM). The RT donor RNA comprises a primer binding site (PBS) and a template sequence.

15 In some embodiments, the CRISPR nuclease polypeptide comprises the amino acid sequence of SEQ ID NO: 1. In other embodiments, the CRISPR nuclease is a variant of SEQ ID NO: 1, the variant comprising: (i) one or more mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1 that reduce or eliminate the nuclease activity thereof; (ii) one or more arginine and/or lysine substitutions, optionally one or more arginine
20 substitutions; or (iii) a combination of (i) and (ii);

 In some instances, the fusion polypeptide comprises one or more nuclear localization signal (NLS) upstream or downstream to the CRISPR nuclease polypeptide, the RT polypeptide, or both. For example, the fusion polypeptide, from N-terminus to C-terminus, comprises a first NLS, the CRISPR nuclease polypeptide, the RT polypeptide, and a second NLS. In some
25 examples, the fusion polypeptide may further comprise a peptide linker between the CRISPR nuclease polypeptide and the RT polypeptide.

 In some examples, the fusion polypeptide comprises a first peptide linker located between the CRISPR nuclease polypeptide and the RT polypeptide. Further, the fusion polypeptide may comprise a first NLS, a second NLS, which are located at the N-terminus
30 and/or the C-terminus of the fusion polypeptide. In some instances, the fusion polypeptide may further comprise additional NLSs, for example, a third NLS and optionally a fourth NLS. Alternatively or in addition, the fusion polypeptide may further comprise a second peptide linker and optionally a third peptide linker. These peptide linkers may be located (*e.g.*, connecting) the CRISPR nuclease polypeptide and/or the RT polypeptide, and the first and/or second NLS. In

some instances, these peptide linkers may be located (*e.g.*, connecting) two NLSs.

Provided below are specific exemplary configurations for the fusion polypeptide disclosed herein (from N-terminus to C-terminus):

- 5 (i) the first NLS, the second NLS, the CRISPR nuclease polypeptide, the first peptide linker, and the RT polypeptide;
- (ii) the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the first NLS, and the second NLS;
- (iii) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third NLS, the second peptide linker, and the second NLS;
- 10 (iv) the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third NLS, the second NLS, the second peptide linker, and the first NLS;
- (v) the first NLS, the second peptide linker, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third peptide linker, and the second NLS;
- 15 (vi) the first NLS, the second peptide linker, the CRISPR nuclease, the first peptide linker, the RT polypeptide, the third peptide linker, the third NLS, the fourth NLS, and the second NLS;
- (vii) the first NLS, the second peptide linker, the RT polypeptide, the first peptide linker, the CRISPR nuclease polypeptide, the third peptide linker, the third NLS, the fourth NLS, and the second NLS;
- 20 (viii) the RT polypeptide, the first peptide linker, the CRISPR nuclease polypeptide, the third NLS, the second NLS, the second peptide linker, and the first NLS;
- (ix) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the second peptide linker, the second NLS, and the RT polypeptide;
- 25 (x) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the third NLS, the second peptide linker, the RT polypeptide, the third peptide linker, and the second NLS;
- (xi) the first NLS, the RT polypeptide, the first peptide linker, the second NLS, the second peptide linker, and the CRISPR nuclease polypeptide,
- 30 (xii) the first NLS, the RT polypeptide, the first peptide linker, the third NLS, the second peptide linker, the CRISPR nuclease polypeptide, the third peptide linker, and the second NLS;
- (xiii) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, and the second NLS; or

- (xiv) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the second peptide linker, the third peptide linker, and the second NLS.

In specific examples, the fusion polypeptide may have the configuration (from N-terminus to C-terminus) of (iv), (v), (vii), (ix), or (x). See also **Table 16** below.

In some instances, wherein the peptide linker(s) between the CRISPR nuclease polypeptide and the RT polypeptide is about 20-80 amino acids in length.

In some embodiments, the CRISPR nuclease polypeptide comprises the variant of SEQ ID NO: 1. For example, the variant of SEQ ID NO: 1 may comprise one or more mutations in the HNH nuclease domain at positions D844, H845, and/or N868 relative to SEQ ID NO: 1. In one example, the mutation is at position H845 (*e.g.*, H845A substitution). In some examples, the mutation at D844 is an amino acid substitution of D844A, D844G, D844L, or D844S. In some examples, the mutation at H845 is an amino acid substitution of H845A, H845G, H845L, or H845S. In a specific example, the CRISPR nuclease polypeptide comprises the mutation at position H845 (*e.g.*, H845A) relative to SEQ ID NO: 1 (*e.g.*, comprising the amino acid sequence of SEQ ID NO: 32). In some examples, the mutation at N868 is an amino acid substitution of N868A, N868G, N868L, or N868S.

Alternatively or in addition, the CRISPR nuclease polypeptide comprises a bridge helix (BH) domain, a nucleic acid recognition (REC) domain, a phosphate lock loop (PLL), a wedge (WED) domain, and a PAM-interacting (PID) domain, and wherein one or more arginine and/or lysine substitutions, optionally arginine substitutions, are located in the BH domain, in the REC domain, in the PLL domain, in the WED domain, in the PID domain, or a combination thereof. For example, the CRISPR nuclease polypeptide contains up to 20 arginine and/or lysine substitutions relative to the reference CRISPR nuclease. In specific examples, the CRISPR nuclease polypeptide contains up to 15 arginine and/or lysine substitutions relative to the reference CRISPR nuclease. In more specific examples, the one or more arginine and/or lysine substitutions are at positions K736, L784, Q812, N813, I857, and/or A919.

In some examples, the CRISPR nuclease polypeptide contains at least two arginine and/or lysine substitutions relative to the reference CRISPR nuclease. The two arginine and/or lysine substitutions are at positions K736, L784, Q812, N813, I857, and/or A919. In some instances, the CRISPR nuclease polypeptide contains arginine and/or lysine substitutions at the following positions relative to the reference CRISPR nuclease:

- (a) I857, L784, and K736 (*e.g.*, I857R, L784R, and K736R);
- (b) I857, A919, and K736 (I857R, A919R, and K736R);

- (c) I857, N813, and L784 (I857R, N813R, and L784R);
- (d) I857, L784, and A919 (I857R, L784R, and A919R);
- (e) I857, N813, and K736 (I857R, N813R, and K736R);
- (f) I857 and N813 (I857R and N813R);
- 5 (g) L784, A919, and K736 (L784R, A919R, and K736R);
- (h) I857 and L784 (I857R and L784R); or
- (i) I857 and A919 (I857R and A919R).

In specific examples, the CRISPR nuclease polypeptide comprises the arginine substitutions of I857R, L784R, and K736R relative to SEQ ID NO: 1.

- 10 In some instances, the CRISPR nuclease polypeptide disclosed herein may further comprise one or more mutations that enhance double-strand nuclease activity relative to the reference CRISPR nuclease, in which the mutations are introduced. Examples are provided in **Table 19** below.

- Alternatively or in addition, the engineered CRISPR nuclease polypeptide disclosed
 15 herein may comprise or further comprise the one or more mutations for reducing PAM recognition stringency. In some instances, the one or more mutations for reducing PAM recognition stringency may be at position D61, A68, H494, L1117, D1144, S1145, G1227, E1228, S1327, A1332, R1343, R1345, and/or T1347 of SEQ ID NO: 1. In some examples, such mutations may comprise: (i) one or more arginine and/or lysine substitutions, optionally
 20 arginine substitutions, at position D61, A68, H494, L1117, G1227, S1327, A1332, and/or T1347 of SEQ ID NO: 1; (ii) one or more amino acid substitutions at position D1144, S1145, E1228, R1343, and/or R1345, of SEQ ID NO: 1; or (iii) a combination of (i) and (ii). In specific examples, the one or more amino acid substitutions of (ii) may comprise optionally D1144L, S1145W, E1228Q, R1343P, R1345V, and/or R1345Q relative to SEQ ID NO: 1.

- 25 In specific examples, the engineered CRISPR nuclease polypeptide with less PAM recognition stringency may comprise the following combination of mutations: L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and A68R relative to SEQ ID NO: 1. In other specific examples, the engineered CRISPR nuclease polypeptide with less PAM recognition stringency may comprise the following combination of mutations: L1117R,
 30 D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and D61R relative to SEQ ID NO: 1. In yet other specific examples, the engineered CRISPR nuclease polypeptide with less PAM recognition stringency may comprise the following combination of mutations: L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and H494R relative to SEQ ID NO: 1. Other exemplary engineered CRISPR nuclease polypeptides can be found in **Table 20**, each

of which is within the scope of the present disclosure.

The engineered CRISPR nuclease polypeptide as disclosed herein recognize a PAM sequence of 5'-NDR-3', in which N represents A, C, G, or U, D represents A, G, or T, and R represents G or A. In some instances, the engineered CRISPR nuclease polypeptides having reduced PAM recognition stringency as disclosed herein may recognize a PAM sequence of 5'-NGN-3'. See **Example 10** below. In some examples, the PAM is 5'-NRG-3' or 5'-NRR-3', in which N and R are defined herein. In some specific examples, the PAM is 5'-NGG-3', in which N represents any nucleotide. In other specific examples, the PAM can be 5'-TGC-3' or 5'-GGA-3'.

Any of the CRISPR nuclease polypeptides may comprise the arginine and/or lysine substitutions disclosed herein, any of the nickase mutations in either the HNH or RuvC nuclease domains also disclosed herein, any of the mutations leading to reduced PAM recognition stringency, or a combination thereof. For example, the CRISPR nuclease polypeptide may comprise (a) the one or more nickase mutations in the HNH nuclease domain at positions D844, H845, and/or N868 relative to SEQ ID NO: 1 (*e.g.*, the mutation is at position H845); and (b) one or more arginine and/or lysine substitutions relative to SEQ ID NO: 1 (*e.g.*, at positions I857, L784, and K736). For example, in some examples, the CRISPR nuclease polypeptide may comprise (*e.g.*, consists of) a nickase mutation at position H845 (*e.g.*, an H845A mutation) and an arginine and/or lysine substitution at position I857 (*e.g.*, an I857R substitution) relative to SEQ ID NO: 1. In other examples, the CRISPR nuclease polypeptide may comprise or further comprise the one or more mutations that result in reduced PAM recognition stringency (*e.g.*, at positions L1117, D1144, G1227, E1228, A1332, R1345, and/or T1347 of SEQ ID NO: 1, and optionally at one or more positions of D61, A68, and H494 of SEQ ID NO: 1).

Any of the CRISPR nuclease polypeptide provided herein may comprise an amino acid sequence at least 90% identical to SEQ ID NO: 1. In some examples, the CRISPR nuclease polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 1. In yet other examples, the CRISPR nuclease polypeptide comprises an amino acid sequence at least 98% identical to SEQ ID NO: 1.

The RT polypeptide is Moloney Murine Leukemia Virus (MMLV)-RT or a variant thereof. In one example, the MMLV-RT comprises the amino acid sequence of SEQ ID NO: 53.

In some embodiments, the gene editing system disclosed herein comprises the fusion polypeptide. Alternatively, the system comprises the first nucleic acid encoding the fusion polypeptide. In some examples, the first nucleic acid is located on a vector, which optionally is a viral vector. In other examples, the first nucleic acid is a first messenger RNA (mRNA).

In any of the gene editing systems provided herein, the spacer sequence in the gRNA of (b) can be 15-30-nucleotide in length. In one example, the spacer sequence may be 15-20-nucleotide in length. In one specific example, the spacer sequence may be about 17-nucleotide in length.

5 In some embodiments, the scaffold sequence comprises a nucleotide sequence at least 85% identical to SEQ ID NO: 2. In one example, the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2.

 In some embodiments, the PBS in the RT donor RNA portion of the RNA molecule is 5-50-nucleotide in length. In some examples, the PBS is 5-20-nucleotide in length. In specific
10 examples, the PBS can be 7-17-nucleotide in length. In some embodiments, the PBS binds a PBS-targeting site that is adjacent to or overlaps with the target sequence. For example, the PBS-targeting site is adjacent to or overlaps with the target sequence. In other examples, the PBS-targeting site is adjacent to the 5' of the PAM.

 In some embodiments, the template sequence in the RT donor RNA portion of the RNA
15 molecule can be 5-100-nucleotide in length. In some examples, the template sequence can be 15-25-nucleotide in length. In some embodiments, the template sequence in the RT donor RNA is homologous to the genomic site of interest and comprises one or more nucleotide variations relative to the genomic site of interest. In some examples, at least one nucleotide variation may be located within the target sequence. Alternatively, at least one nucleotide variation may be
20 located in the PAM.

 In some embodiments, any of the RNA molecules of (b) provided herein may further comprise a 3' end extension. In some examples, the RNA molecule may further comprise a 5' end protection fragment, a 3' protection fragment, or both, each of the 5' end protection fragment and the 3' end protection fragment forming a secondary structure, which optionally is a
25 hairpin, a circularization, a pseudoknot, or a triplex structure.

 In some examples, the RNA molecule of (b) comprises, from 5' to 3': the spacer sequence, the scaffold sequence, the template sequence, and the PBS. In other examples, the RNA molecule may comprise, from 5' to 3', the spacer sequence, the scaffold sequence, the template sequence, the PBS, and the 3' extension.

30 In some embodiments, the gene editing system disclosed herein may comprise the RNA molecule that comprises the gRNA, the RT donor RNA, and optionally one or more of the additional elements disclosed herein. Alternatively, the gene editing system may comprise the nucleic acid encoding the RNA molecule. In some examples, the nucleic acid is located on a vector, which optionally is a viral vector.

In some embodiments, the gene editing system disclosed herein may comprise one or more lipid nanoparticles (LNPs) associated with one or more of elements (a)-(b). Alternatively, the gene editing system may comprise one or more viral vectors, for example, optionally one or more adeno-associated viral (AAV) vectors encoding one or more of elements (a)-(b).

5 Also provided herein are a pharmaceutical composition comprising any of the gene editing systems provided herein, and a kit comprising the elements (a) and (b) of the gene editing system as disclosed herein.

In other aspects, the present disclosure features a gene editing method, comprising delivering the gene editing system disclosed herein to a host cell to edit a genomic site targeted
10 by the gRNA of the gene editing system. In some embodiments, the host cell is cultured *in vitro*. In other embodiments, the host cell is located in a subject who needs the gene editing.

Further, the present disclosure provides a fusion polypeptide, comprising any of the CRISPR nuclease polypeptide set forth herein and any of the reverse transcriptase polypeptide also set forth herein. Such a fusion polypeptide may comprise the amino acid sequence of SEQ
15 ID NO: 55 or 57.

In addition, the present disclosure provides a nucleic acid encoding the fusion polypeptide disclosed herein. Such a nucleic acid may comprise the nucleotide sequence of SEQ ID NO: 54 or 56. In some examples, the nucleic acid is a vector, such as an expression vector.

The details of one or more embodiments of the invention are set forth in the description
20 below. Other features or advantages of the present invention will be apparent from the following drawings and detailed description of several embodiments, and also from the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

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

FIG. 1 is a diagram showing gene editing efficacy of reference CRISPR nuclease SEQ ID NO: 1 on exemplary target genes AAVS1, EMX1, and VEGFA.

30 **FIGs. 2A-2D** include gel images showing quantification of nuclease activities. **FIG. 2A:** a gel image captured using a 700 nm channel showing *in vitro* cleavage of the target strand (labelled on the 5' end with an IR700 dye) of the target DNA substrate by the reference CRISPR nuclease, putative HNH-knockout nickases, or putative RuvC-knockout nickases. **FIG. 2B:** a gel image captured using an 800 nm channel showing *in vitro* cleavage of the non-target strand

(labelled on the 5' end with IR800) of the target DNA substrate by the reference CRISPR nuclease, putative HNH-knockout nickases, or putative RuvC-knockout nickases. **FIG. 2C:** Overlaid images captured using 700 nm and 800 nm channels of **FIG. 2A** and **FIG. 2B**. **FIG. 2D:** Quantification of the percent of cleaved target and non-target DNA generated by the reference CRISPR nuclease, the putative HNH-knockout nickases, and the putative RuvC-knockout nickases tested in **Example 4**.

FIGs. 3A and **3B** include diagrams showing reverse transcription-mediated gene editing efficiency using the reference CRISPR nuclease-reverse transcriptase fusion polypeptide or CRISPR nickase variant-reverse transcriptase fusion polypeptide. **FIG. 3A:** Percentage of NGS reads comprising indels (black bars) and edits encoded by editing template RNAs (grey bars) using the reference CRISPR nuclease-reverse transcriptase fusion polypeptide. **FIG. 3B:** Percentage of NGS reads comprising indels (black bars) and edits encoded by editing template RNAs (grey bars) using the CRISPR nickase variant-reverse transcriptase fusion polypeptide.

FIG. 4A and **FIG. 4B** include diagrams showing gene editing efficiencies of CRISPR nuclease-reverse transcriptase fusion polypeptides with various configurations as indicated in **Table 16** and **Table 17**. **FIG. 4A:** gene editing efficiency of fusion polypeptides containing the CRISPR nuclease of SEQ ID NO: 1. The constructs correspond to those listed in **Table 17**, except that the nickase sequence therein is replaced with the reference nuclease of SEQ ID NO: 1. **FIG. 4B:** gene editing efficiency of the fusion polypeptides listed in **Table 17**, each of which comprises the CRISPR nickase of SEQ ID NO: 32.

FIG. 5 is a diagram showing gene editing efficiency of the exemplary CRISPR nickase-reverse transcriptase fusion polypeptides listed in **Table 17** as indicated by percentages of eGFP-positive cells relative to mCherry-positive cells.

FIG. 6 is a diagram showing gene editing efficiency of converting BFP to eGFP by the exemplary CRISPR nickase-reverse transcriptase fusion polypeptides listed in **Table 17** relative to the percentage of eGFP-positive cells measured for each exemplary CRISPR nickase-reverse transcriptase fusion polypeptide.

FIG. 7 is a diagram showing gene editing efficiency at the EMX1_T2 site by the exemplary CRISPR nickase-reverse transcriptase fusion polypeptides listed in **Table 17** relative to the percentage of eGFP-positive cells measured for each exemplary CRISPR nickase-reverse transcriptase fusion polypeptide.

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure provides a gene editing system involving both a CRISPR nuclease

polypeptide and a reverse transcriptase (RT) polypeptide, as well as a guide RNA, which directs gene editing at a desired genomic site, and an RT donor RNA, which serves as the RNA template for the RT polypeptide to synthesize DNA strands carrying desired base substitutions. In some embodiments, the gene editing system provided herein may comprise a fusion
5 polypeptide comprising the CRISPR nuclease fragment and the RT fragment, or a nucleic acid encoding the fusion polypeptide. Alternatively or in addition, the gene editing system may comprise a single RNA molecule comprising the guide RNA and the RT donor RNA, or a nucleic acid encoding the single RNA molecule.

The gene editing system provided herein has shown successful substitution of
10 nucleotides at target sites. See Examples below. Such gene editing systems are expected to be effective in introducing desired nucleotide substitutions at genetic sites of interests, thereby achieving desired therapeutic effects (*e.g.*, correcting genetic defects). The gene editing system provided herein can also be used in other areas, for example, in breeding and genomic functional studies of animals and plants.

15

I. RT-CRISPR Mediated Gene Editing System

In some aspects, provided herein is an RT-CRISPR mediated gene editing system, which involves at least two protein components, *i.e.*, a CRISPR nuclease polypeptide and an RT polypeptide, and at least two RNA components, *i.e.*, a guide RNA and an RT donor RNA. In
20 specific embodiments, the two protein components can be located on a fusion polypeptide. Alternatively or in addition, the two RNA components may be located on a single RNA molecule. In some instances, the gene editing system may comprise the protein components and/or the RNA components. In other instances, the gene editing system may comprise nucleic acid(s) encoding the protein components, and/or nucleic acid(s) encoding the RNA components.

25 In some instances, the RT-CRISPR mediated gene editing system comprises a CRISPR nuclease having nickase activity as disclosed herein. Such a gene editing system is expected to achieve precise gene editing at a desired genomic target site.

A. Protein Components

30 The gene editing systems provided herein involve at least two enzymes, a CRISPR nuclease and an RT. In some embodiments, the gene editing system comprises the two enzymes. In specific examples, the gene editing system may comprise a fusion polypeptide comprising the two enzyme components. Alternatively, the gene editing system may comprise one or more nucleic acids encoding the two enzyme components. For example, the gene editing system may

comprise one or more expression vectors (*e.g.*, viral vectors such as retroviral vectors, adenoviral vectors, or adeno-associated viral vectors) capable of expressing the CRISPR nuclease, the RT, or the fusion polypeptide comprising such. In other examples, the gene editing system may comprise one or more mRNA molecules coding for the CRISPR nuclease, the RT, or the fusion polypeptide comprising such.

In some embodiments, the CRISPR nuclease polypeptide and the RT polypeptide as disclosed herein may form a complex, which may be a heterodimer of the two protein components via a dimerization domain (*e.g.*, a leucine zipper), an antibody, a nanobody, or an aptamer.

10

(i) CRISPR Nuclease Polypeptides

The CRISPR nuclease polypeptide for use in the gene editing systems disclosed herein may be a CRISPR nuclease comprising the amino acid sequence of SEQ ID NO: 1 (the reference CRISPR nuclease), or a variant thereof.

As used herein, the term “CRISPR nuclease” refers to an RNA-guided effector that is capable of binding a nucleic acid and introducing a single-stranded break or double-stranded break. A CRISPR nuclease typically comprises multiple functional domains, *e.g.*, nuclease domains (*e.g.*, RuvC and HNH), bridge helix (BH) domain, nucleic acid recognition (REC) domain, phosphate lock loop (PLL), wedge domain (WED), PAM-interacting domain (PID), or a combination thereof. As used herein, the term “domain” refers to a distinct functional and/or structural unit of a polypeptide. In some instances, a functional domain may be linear. In other instances, a functional domain can be discontinuous and conformational. In some embodiments, a domain may comprise a conserved amino acid sequence across different CRISPR nucleases.

The reference CRISPR nuclease of SEQ ID NO: 1 (see **Table 1** below) is a CRISPR nuclease that comprises both a RuvC nuclease domain (located at residues 1-59, 722-771, and 927-1101 of SEQ ID NO: 1) and a HNH domain (located at residues 772-926 of SEQ ID NO: 1). The RuvC nuclease domain and the HNH nuclease domain coordinate cleavage of the DNA strand adjacent to the 5'-NDR-3' PAM motif, in which N represents any nucleotide, D represents A, G, or T, and R represents G or A. In some examples, the PAM is 5'-NRG-3' or 5'-NRR'3', in which N and R are defined herein. In one specific example, the PAM is 5'-NGG-3', in which N represents any nucleotide. Positions D10, E763 and D991 are deemed the active sites in the RuvC domain and positions D844, H845, and N868 are deemed the active sites in the HNH domain. In addition to the nuclease domains, the reference CRISPR nuclease of SEQ ID NO: 1 also includes a BH domain (residues 60-93 of SEQ ID NO: 1), a REC domain (residues

94-721 of SEQ ID NO: 1), a PLL domain (residues 1102-1148 of SEQ ID NO: 1), a WED domain (residues 1149-1208 of SEQ ID NO: 1), and a PID domain (residues 1209-1378 of SEQ ID NO: 1).

In some embodiments, the gene editing system disclosed herein comprises a variant CRISPR nuclease polypeptide derived from SEQ ID NO: 1, *e.g.*, a variant CRISPR nuclease polypeptide comprising one or more arginine or lysine substitutions, one or more mutations in one of the nuclease domains such as in the HNH nuclease domain or the RuvC nuclease domain, or a combination thereof, as relative to the reference CRISPR nuclease of SEQ ID NO: 1. Such a protein component may form a complex with the gRNA(s) in the same gene editing system.

10

Variants of CRISPR Nuclease Polypeptide

In some embodiments, the CRISPR nuclease polypeptide in the gene editing system disclosed herein is a variant of the reference CRISPR nuclease of SEQ ID NO: 1, *e.g.*, via introducing one or more mutations to the reference CRISPR nuclease to modulate (*e.g.*, enhance or reduce) one or more activities of the nuclease. As used herein, the term “variant CRISPR nuclease polypeptide” refers to a CRISPR nuclease polypeptide comprising an alteration, *e.g.*, a substitution, insertion, deletion and/or fusion, at one or more residue positions, compared to the reference CRISPR nuclease (SEQ ID NO: 1).

The variant CRISPR nuclease polypeptides may comprise one or more mutations (*e.g.*, arginine substitutions) relative to the reference CRISPR nuclease. Alternatively or in addition, the variant CRISPR nuclease polypeptide may comprise one or more mutations in either the RuvC nuclease domain or the HNH nuclease domain. Such mutations may reduce or eliminate the nuclease activity of either the RuvC or the HNH nuclease domain, leading to a variant exhibiting nickase activity. The variant CRISPR nuclease polypeptides may share a high sequence homology relative to the reference CRISPR nuclease (*e.g.*, at least 85% sequence identity).

As used herein, the term “nickase” refers to an enzyme that cuts one strand of a double-stranded DNA at a specific recognition nucleotide sequence (*e.g.*, the target sequence disclosed herein). A nickase may interact with one strand of the DNA duplex to produce DNA molecules that are cut at one strand (*a.k.a.*, nicked). In some embodiments, a nickase is a variant of a CRISPR nuclease that comprises a deactivated HNH domain. In some embodiments, a nickase is a variant of a CRISPR nuclease that comprises a deactivated RuvC domain. In other embodiments, the variant CRISPR nuclease polypeptide may comprise one or more mutations relative to SEQ ID NO: 1 that result in reduced PAM recognition stringency as compared with a

counterpart CRISPR nuclease polypeptide without such mutations. The variant CRISPR nuclease polypeptides may share a high sequence homology relative to the reference CRISPR nuclease (*e.g.*, at least 85% sequence identity).

5 The variant CRISPR nuclease polypeptides provided herein are expected to possess advantageous features relative to the reference CRISPR nuclease, for example, exhibiting nickase activity and/or higher nuclease activity, etc. As such, the variant CRISPR nuclease polypeptides disclosed herein would be expected to exhibit improved gene editing relative to the reference CRISPR nuclease, *e.g.*, higher efficiency and accuracy in gene editing involving strand replacement.

10 In some embodiments, the variant CRISPR nuclease polypeptides provided herein, relative to the reference CRISPR nuclease SEQ ID NO: 1, comprises one or more mutations in either the RuvC nuclease domain or the HNH nuclease domain (*e.g.*, in the HNH nuclease domain) to reduce or eliminate the nuclease activity, and/or comprises one or more arginine and/or lysine substitutions to improve nuclease features suitable for use in gene editing.

15 The variant CRISPR nuclease polypeptides provided herein are expected to exhibit one or more modulated activities (*e.g.*, enhanced or reduced) relative to the reference CRISPR nuclease. As used herein, the term “activity” refers to a biological activity. In some embodiments, activity includes enzymatic activity, *e.g.*, catalytic ability of an effector. For example, activity can include nuclease activity. In some embodiments, activity includes nickase activity. For example, the variant CRISPR nuclease polypeptides may cut substantially at only one strand of the target DNA duplex.

20 In some embodiments, activity includes binding activity, *e.g.*, binding of an effector (*e.g.*, a CRISPR nuclease) to an RNA guide and/or target nucleic acid. In some examples, the variant CRISPR nuclease polypeptides disclosed herein have an enhanced binding to a cognate guide RNA (gRNA) as compared with the reference CRISPR nuclease, *e.g.*, having a binding activity at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 2-fold, 2-fold, 5-fold, 10-fold, or greater than that of the reference CRISPR nuclease. A cognate gRNA refers to a gRNA having a scaffold recognizable by the CRISPR nuclease.

25 In some examples, the variant CRISPR nuclease polypeptides disclosed herein have an enhanced enzymatic activity relative to the reference CRISPR nuclease, *e.g.*, having an enzymatic activity at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 2-fold, 2-fold, 5-fold, 10-fold, or greater than that of the reference CRISPR nuclease. In other examples, the variant CRISPR nuclease polypeptides disclosed herein have a decreased enzymatic activity (*e.g.*, the enzymatic activity for cleaving both strands of a target DNA duplex) relative to the

reference CRISPR nuclease, *e.g.*, having an enzymatic activity at least 20%, 30%, 40%, 50%, 60%, or 70% lower than that of the reference CRISPR nuclease. In some instances, the decreased enzymatic activity is achieved by reducing or diminishing the nuclease activity of the RuvC domain. In other instances, the decreased enzymatic activity is achieved by reducing or
5 diminishing the nuclease activity of the HNH domain.

In some instances, the variant CRISPR nuclease polypeptides disclosed herein have enhanced indel activity relative to the reference CRISPR nuclease. As used herein, the term “indel activity” refers to the ability of a CRISPR nuclease to introduce an indel (insertion/deletion) into a sequence (*e.g.*, a genomic target). For example, in some embodiments,
10 the CRISPR nuclease introduces a double-strand break into a sequence (*e.g.*, a genomic target in a cell), and through DNA repair mechanisms, an indel is created.

In some embodiments, the variant CRISPR nuclease polypeptide provided herein share a high sequence homology relative to the reference CRISPR nuclease. For example, the variant CRISPR nuclease polypeptide may comprise an amino acid sequence at least 70% (*e.g.*, at least
15 80%, 85%, 90%, 95%, or higher) identical to SEQ ID NO: 1. In some instances, the variant CRISPR nuclease polypeptide may comprise an amino acid sequence at least 90% identical to SEQ ID NO: 1. In some instances, the variant CRISPR nuclease polypeptide may comprise an amino acid sequence at least 95% identical to SEQ ID NO: 1. In other instances, the variant CRISPR nuclease polypeptide may comprise an amino acid sequence at least 97% (*e.g.*, 98%,
20 99%, 99.5%, or greater) identical to SEQ ID NO: 1.

The “percent identity” (*a.k.a.*, sequence identity) of two nucleic acids or of two amino acid sequences is determined using the algorithm of Karlin and Altschul Proc. Natl. Acad. Sci. USA 87:2264-68, 1990, modified as in Karlin and Altschul Proc. Natl. Acad. Sci. USA 90:5873-77, 1993. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version
25 2.0) of Altschul, et al. J. Mol. Biol. 215:403-10, 1990. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, word length=3 to obtain amino acid sequences homologous to the protein molecules of the invention. Where gaps exist between two
30 sequences, Gapped BLAST can be utilized as described in Altschul et al., Nucleic Acids Res. 25(17):3389-3402, 1997. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used.

The variant CRISPR nuclease polypeptide provided herein can contain one or more alterations relative to the reference CRISPR nuclease of SEQ ID NO: 1, *e.g.*, one or more amino

acid residue substitutions, one or more deletions, one or more insertions, fusion, or a combination thereof. In some instances, the alterations may be introduced into the BH domain, the PLL domain, the WED domain, the PID domain, or a combination thereof.

In some embodiments, the variant CRISPR nuclease polypeptide provided herein may comprise one or more arginine substitutions relative to SEQ ID NO: 1. “Arginine substitutions” or “lysine substitution” refers to the replacement of a non-arginine or non-lysine residue in SEQ ID NO: 1 with an arginine or lysine residue. In some examples, the variant CRISPR nuclease polypeptide may contain up to 20 arginine and/or lysine substitutions, *e.g.*, up to 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 arginine and/or lysine substitutions. In specific examples, the variant CRISPR nuclease polypeptide may contain 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 arginine and/or lysine substitutions.

In some instances, one or more of the substituting arginine residues may be replaced by a conservative amino acid residue such as lysine or histidine. In some embodiments, the variant CRISPR nuclease polypeptide provided herein may comprise one or more arginine substitutions, one or more lysine substitutions, or a combination thereof.

In some instances, the arginine substitutions may be located in the BH domain, in the PLL domain, in the WED domain, in the PID domain, or in any of the combination thereof. In some examples, the variant CRISPR nuclease polypeptide may contain one or more arginine and/or lysine substitutions at one or more of positions: I79, E331, Y348, S473, F501, I581, D720, A730, G731, Q741, V752, M753, Q809, Q840, Q849, S872, S898, E982, K918, D985, Y986, Y1015, E1037, K1091, S1094, P1096, N1099, T1104, E1105, I1106, T1108, L1117, K1131, I1147, E1179, M1205, P1208, E1214, A1226, Q1230, A1236, P1238, F1241, L1281, D1284, F1285, A1292, N1295, K1298, G1329, A1333, K1344, S1348, Q1360, and I1370 of SEQ ID NO: 1. In specific examples, the arginine and/or lysine substitutions can be at one or more of positions I857R, K736, L784, N813, Q812, I857, and A919 of SEQ ID NO: 1.

In some specific examples, the variant CRISPR nuclease polypeptide may contain one or more arginine substitutions of I79R, E331R, Y348R, S473R, F501R, I581R, D720R, A730R, G731R, Q741R, V752R, M753R, Q809R, Q840R, Q849R, S872R, S898R, E982R, K918R, D985R, Y986R, Y1015R, E1037R, K1091R, S1094R, P1096R, N1099R, T1104R, E1105R, I1106R, T1108R, L1117R, K1131R, I1147R, E1179R, M1205R, P1208, E1214R, A1226R, Q1230R, A1236R, P1238R, F1241R, L1281R, D1284R, F1285R, A1292R, N1295R, K1298R, G1329R, A1333R, K1344R, S1348R, Q1360R, and I1370R relative to SEQ ID NO: 1.

In some examples, the variant CRISPR nuclease polypeptide may contain one or more arginine substitutions at one or more of the above-noted positions. Examples include I857R,

N813R, L784R, K736R, A919R, Q812R, or a combination thereof. In other examples, the variant CRISPR nuclease polypeptide may contain one or more lysine substitutions at one or more of the above-noted positions. Examples include I857K, N813K, L784K, A919K, Q812K, or a combination thereof.

- 5 In other examples, the variant CRISPR nuclease polypeptide may contain a combination of arginine and/or lysine substitutions at: I79, E331, Y348, S473, F501, I581, D720, A730, G731, Q741, V752, M753, Q809, Q840, Q849, S872, S898, E982, K918, D985, Y986, Y1015, E1037, K1091, S1094, P1096, N1099, T1104, E1105, I1106, T1108, L1117, K1131, I1147, E1179, M1205, P1208E1214, A1226, Q1230, A1236, P1238, F1241, L1281, D1284, F1285, 10 A1292, N1295, K1298, G1329, A1333, K1344, S1348, Q1360, and/or I1370 of SEQ ID NO: 1. In specific examples, the variant CRISPR nuclease polypeptide may contain a combination of arginine and/or lysine substitutions (*e.g.*, combination of arginine substitutions) at I857R, K736, L784, N813, Q812, I857, and/or A919 of SEQ ID NO: 1.

- In specific examples, the engineered CRISPR nuclease polypeptide may contain the 15 arginine and/or lysine substitutions at the following positions relative to SEQ ID NO: 1: (a) I857, L784, and K736; (b) I857, A919 and K736; (c) I857, N813, and L784; (d) I857, L784, and A919; (e) I857, N813, and K736; (f) I857 and N813; (g) L784, A919, and K736; (h) I857, and L784; or (i) I857 and A919. In some instances, the engineered CRISPR nuclease polypeptide may contain arginine substitutions at any of the combinations of positions in SEQ ID NO: 1. In 20 one specific example, the engineered CRISPR nuclease polypeptide may contain arginine substitutions I857R, L784R, and K736R relative to SEQ ID NO: 1. Other examples of arginine and/or lysine substitutions can be found in **Table 4** below.

- Alternatively or in addition, the variant CRISPR nuclease polypeptide provided herein may comprise one or more mutations within either the RuvC or the HNH nuclease 25 domain to reduce or eliminate the nuclease activity of the target domain, thereby producing a variant with nickase activity. Such mutations (nickase mutations) may be deletions, insertions, amino acid substitutions, or a combination thereof. In some embodiments, the mutations within either the RuvC or the HNH nuclease domain are amino acid substitutions, of which the substituting amino acid residue is not a conservative substitution of the native 30 amino acid residue at the position of the mutation. For example, if the native amino acid residue is R, the substituting residue can be any amino acid residue except for K. Similarly, if the native amino acid residue is K, the substituting residue can be any amino acid residue except for R. Groups of conservative amino acid residue substitutions are provided herein.

In some instances, the one or more nickase mutations may be within the HNH

nuclease domain, for example, at D844, H845, and/or N868 of SEQ ID NO: 1. In some examples, the mutations may be amino acid residue substitutions and the native amino acid residues in SEQ ID NO: 1 may be replaced by an amino acid residue not of the same type as the native residues. For example, a positively charged residue may be replaced by a non-charged amino acid residue, or *vice versa*. In some examples, the amino acid residue substitution at D844 may be D844G, D844A, D844L, or D844S. In one specific example, the mutation can be D844A. In another example, the amino acid residue substitution at H845 may be H845G, H845A, H845I, H845L, H845M, H845V, or H845S. In one specific example, the mutation at position H845 can be H845A. Alternatively or in addition, the amino acid residue substitution may be at position N868, for example, N868G, N868A, N868L, or N868S. In one example, the mutation at position N868 is N868A.

In some instances, the one or more mutations may be with the RuvC nuclease domain, for example, at position D10, E763, D991, or a combination thereof, of SEQ ID NO: 1 (*e.g.*, at position E763 and/or D991). In some examples, the mutations may be amino acid residue substitutions and the native amino acid residues in SEQ ID NO: 1 may be replaced by an amino acid residue not of the same type as the native residues. For example, a positively charged residue may be replaced by a non-charged amino acid residue, or *vice versa*. In some examples, the amino acid residue substitution at D10 may be D10G, D10A, D10L, or D10S. In some examples, the amino acid residue substitution at E763 may be E763G, E763A, E765L, or E763S. Alternatively or in addition, the amino acid residue substitution at D991 may be D991G, D991A, D991L, or D991S.

In some examples, the variant CRISPR nuclease polypeptide provided herein may be a nickase variant, which comprises one or more mutations in one nuclease domain (*e.g.*, the HNH nuclease domain such as at position H845, *e.g.*, H845A). Exemplary nickase variants are provided in **Table 5** below. Such a nickase variant may further comprise one or more arginine or lysine substitutions (*e.g.*, arginine substitutions) for enhancing certain features, such as indel activities. Exemplary arginine and/or lysine substitutions are provided herein, for example, at one or more of positions I857R, K736, L784, N813, Q812, I857, and A919 of SEQ ID NO: 1 (*e.g.*, the arginine substitutions at positions I857, L784, and K736). In some examples, the CRISPR nuclease polypeptide may comprise (*e.g.*, consists of) a nickase mutation at position H845 (*e.g.*, an H845A mutation) and an arginine and/or lysine substitution at position I857 (*e.g.*, an I857R substitution) relative to SEQ ID NO: 1.

In some examples, the variant CRISPR nuclease polypeptide disclosed herein exhibits enhanced double-strand nuclease activity. Examples of such CRISPR nuclease polypeptides are

provide in **Table 19** below, each of which is within the scope of the present disclosure.

In some embodiments, the variant CRISPR nuclease polypeptide disclosed herein may comprise one or more mutations that reduce stringency of PAM recognition relative to the reference CRISPR nuclease. In some instances, the one or more mutations for reducing PAM recognition stringency may be at position of L1117, D1144, S1145, G1227, E1228, S1327, A1332, R1343, R1345, and/or T1347 of SEQ ID NO: 1. In some examples, the one or more mutations may comprise: (i) one or more arginine and/or lysine substitutions, optionally arginine substitutions, at position L1117, G1227, S1327, A1332, and/or T1347 of SEQ ID NO: 1; (ii) one or more amino acid substitutions at position D1144, S1145, E1228, R1343, and/or R1345, of SEQ ID NO: 1; or (iii) a combination of (i) and (ii).

In some examples, such a variant CRISPR nuclease polypeptide may contain mutations (*e.g.*, amino acid residue substitutions) at position D61 (*e.g.*, D61R or D61K), L1117 (*e.g.*, L1117R or L1117K), D1144 (*e.g.*, D1144V, D1144A, D1144G, or D1144S), S1145 (*e.g.*, S1145W, S1145Y, or S1145F), G1227 (*e.g.*, G1227K or G1227R), E1228 (*e.g.*, E1228F, E1228Y, or E1228W), A1327 (*e.g.*, A1327R or A1332K), A1332 (*e.g.*, A1332R or A1332K), R1345 (*e.g.*, R1345Q or R1345N), R1345 (*e.g.*, R1345V, R1345A, R1345G, or R1345S, or R1345Q, R1345N), T1347 (*e.g.*, T1347R or T1347K), or a combination thereof in SEQ ID NO: 1.

In specific examples, the one or more amino acid substitutions of (ii) may comprise D1144L, S1145W, E1228Q, R1343P, R1345V and/or R1345Q relative to SEQ ID NO: 1. Alternatively, the substituting amino acid residues at one or more of positions D1114, S1145, E1228, R1343, and R1345, may be a conservative substitution of L, W, Q, P, V, and Q, respectively. For example, the substitutions at position D1114 may be D1114M, D1114I, or D1114V. The substitutions at position S1145 may be S1145F or S1145Y. The substitutions at position E1228 may be E1228N. The substitutions at position R1345 may be R1345M, R1345I, R1345L, or R1345N.

Specific examples of such variant CRISPR nuclease polypeptides are provided in **Table 20**, each of which is within the scope of the present disclosure.

In some specific examples, the engineered CRISPR nuclease polypeptide comprises (or consists of) L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, and T1347R relative to SEQ ID NO: 1. Such CRISPR nuclease polypeptide has the amino acid sequence set forth in SEQ ID NO: 236.

In some examples, the engineered CRISPR nuclease polypeptide disclosed herein (*e.g.*, those having mutations at positions L1117, D1144, G1227, E1228, A1332, R1343, R1345,

and/or T1347) may further comprise one or more single arginine substitutions (e.g., a single arginine substitution) at position D61, A68, H494, L64, S410, T67, Q849, G1110, F501, T659, L784, Y516, G55, E1037, N57, D720, A919, A1294, Q812, N700, H657, T73, Q899, I857, K751, D327, I581, D462, E331, A589, D471, I699, N1295, T470, I1147, E130, S473, A353, K40, K334, A60, S1348, K367, A1118, K31, Q349, K341, Q83, K585, Q840, G660, K527, G727, Y42, L1281, L122, Q123, T1108, E41, K1131, K30, S872, I1206, D1132, K460, L80, E459, K1182, M696, K918, K126, N721, Q809, K1091, K736, K783, N498, K723, H1119, F463, L594, D472, K744, E365, G595, K45, Y348, K964, S1181, N813, D407, S839, Y658, E586, G754, A730, Y1015, D903, A1333, S461, or H1359 relative to SEQ ID NO: 236.

10 In some specific examples, the engineered CRISPR nuclease polypeptide comprises L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and A68R relative to SEQ ID NO: 1. In other specific examples, the engineered CRISPR nuclease polypeptide comprises L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and D61R relative to SEQ ID NO: 1. In still other specific examples, the engineered CRISPR nuclease polypeptide
15 comprises L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and H494R relative to SEQ ID NO: 1.

In some instances, the variant CRISPR nuclease polypeptide exhibiting less stringent recognition of PAM sequences may recognize 5'-NGN-3', 5'-NRN-3', or 5'-NYN-3' PAM sequences, in which N represents any nucleotide, R represents A or G, and Y represents C or T.

20 In some examples, the variant CRISPR nuclease polypeptide may comprise one or more of the mutations disclosed herein (e.g., one or more arginine and/or lysine substitutions, one or more nickase mutations, and/or one or more mutations resulting less PAM recognition stringency). Any of the variant CRISPR nuclease polypeptides disclosed herein may share a sequence identity at least 90% (e.g., 95%, 97%, 98%, 99%, 99.5%, or greater) with SEQ ID NO:
25 1.

In some instances, the variant CRISPR nuclease polypeptide may comprise one or more conservative amino acid residue substitutions, in addition to the mutations in the HNH or RuvC nuclease domain, the arginine/lysine substitutions, and/or the mutations that result in reduced PAM recognition stringency.

30 As used herein, a "conservative amino acid substitution" refers to an amino acid substitution that does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Variants can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references which compile such methods, e.g., Molecular Cloning: A Laboratory Manual, J. Sambrook, et

al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or Current Protocols in Molecular Biology, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, 5 T; (f) Q, N; and (g) E, D.

Exemplary CRISPR nuclease polypeptides for use in the gene editing systems provided herein are disclosed in **Tables 1, 5, 19, and 20** below, each of which is within the scope of the present disclosure.

In some embodiments, the CRISPR nuclease polypeptide in the gene editing systems 10 disclosed herein (*e.g.*, the reference nuclease of SEQ ID NO: 1 or any of the variants thereof as disclosed herein) may be a fusion polypeptide comprising a CRISPR nuclease and one or more additional functional moieties. As used herein, the terms “fusion” and “fused” refer to the joining of at least two nucleotide or protein molecules. For example, “fusion” and “fused” can refer to the joining of at least two polypeptide domains that are encoded by separate genes in 15 nature. The fusion can be an N-terminal fusion, a C-terminal fusion, or an intramolecular fusion. In some aspects, the domains are transcribed and translated to produce a single polypeptide.

Exemplary functional moieties to include in the fusion polypeptide include a peptide tag, a fluorescent protein, a base-editing domain, a DNA methylation domain, a histone residue modification domain, a localization factor, a transcription modification factor, a light-gated 20 control factor, a chemically inducible factor, a chromatin visualization factor, or a combination thereof.

In some embodiments, the additional functional moiety may comprise a nuclear localization signal (NLS), a nuclear export signal (NES), or a combination thereof. In some examples, the fusion polypeptide may comprise an NLS, which may be located at either the N- 25 terminus or the C-terminus. In specific examples, the fusion polypeptide may comprise a first NLS located at the N-terminus and a second NLS located at the C-terminus. The first and second NLS fragments may be identical. Alternatively, the two NLS fragments may be different. In some embodiments, the fusion polypeptide may comprise an NLS near the N-terminus and/or near the C-terminus (*e.g.*, within about 1, 2, 3, 4, or 5 of the first amino acid or last amino acid of 30 the CRISPR nuclease). In some embodiments, the fusion polypeptide may comprise an NLS within a flexible loop of the CRISPR nuclease.

In some embodiments, the additional functional moiety may be a flexible peptide linker, for example, an XTEN peptide linker, or a G/S rich peptide linker. Examples of such peptide linkers are provided in **Example 1** below.

In some embodiments, the gene editing system provided herein may comprise the CRISPR nuclease polypeptide, which may form a ribonucleoprotein (RNP) complex with the cognate guide RNA. As used herein, the term “complex” refers to a grouping of two or more molecules. In some embodiments, the complex comprises a polypeptide and a nucleic acid molecule interacting with (*e.g.*, binding to, coming into contact with, adhering to) one another. In other embodiments, the gene editing system provided herein may comprise a nucleic acid encoding the CRISPR nuclease polypeptide. In some examples, the nucleotide sequence encoding the CRISPR nuclease polypeptide described herein can be codon-optimized for use in a particular host cell or organism. For example, the nucleic acid can be codon-optimized for any non-human eukaryote including mice, rats, rabbits, dogs, livestock, or non-human primates. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at the world wide web site of kazusa.or.jp/codon/ and these tables can be adapted in a number of ways. See Nakamura *et al. Nucl. Acids Res.* 28:292 (2000), which is incorporated herein by reference in its entirety. Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA). In some examples, the nucleic acid encoding the CRISPR nuclease polypeptides as disclosed herein can be an mRNA molecule, which can be codon optimized. Exemplary codon-optimized nucleotide sequences encoding exemplary CRISPR nuclease polypeptides can be found in **Tables 1** and **5** below, any of which is within the scope of the present disclosure.

In some examples, the gene editing system may comprise a vector (*e.g.*, a viral vector such as an AAV vector, an AdV vector, or a retroviral vector) encoding the CRISPR nuclease polypeptide.

(ii) Reverse Transcriptase Polypeptides

The gene editing system disclosed herein also comprise a reverse transcriptase (RT) polypeptide, which may be a wild-type RT or a variant thereof. In some instances, the RT polypeptide and the CRISPR nuclease polypeptide disclosed herein may form a fusion protein. As used herein, the terms “reverse transcriptase” or “RT” refer to a multi-functional enzyme that typically has three enzymatic activities including RNA- and DNA-dependent DNA polymerization activity and an RNase H activity that catalyzes the cleavage of RNA in RNA-DNA hybrids. A reverse transcriptase can generate DNA from an RNA template.

In some embodiments, the reverse transcriptase polypeptide is any wild-type reverse transcriptase obtained from any naturally-occurring organism or virus, or obtained from a commercial or non-commercial source. The reverse transcriptase polypeptide may also be a variant reverse transcriptase polypeptide.

The reverse transcriptase polypeptide can be obtained from a number of different sources. For instance, the gene may be obtained from eukaryotic cells which are infected with retrovirus or from a plasmid that comprises either a portion of or the entire retrovirus genome. In addition, RNA that comprises the reverse transcriptase gene can be obtained from retroviruses.

5 In some embodiments, the reverse transcriptase is expressed or otherwise provided as an individual component, *i.e.*, not as a fusion protein with the CRISPR nuclease polypeptide provided herein.

A person of ordinary skill in the art will recognize that reverse transcriptases are known in the art, including, but not limited to, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, Human Immunodeficiency Virus (HIV) reverse transcriptase, and avian Sarcoma-Leukosis Virus (ASLV) reverse transcriptase, which includes but is not limited to Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Avian Erythroblastosis Virus (AEV) Helper Virus MCAV reverse transcriptase, Avian Myelocytomatosis Virus MC29 Helper Virus MCAV reverse transcriptase, Avian

10 Reticuloendotheliosis Virus (REV-T) Helper Virus REV-A reverse transcriptase, Avian Sarcoma Virus UR2 Helper Virus UR2AV reverse transcriptase, Avian Sarcoma Virus Y73 Helper Virus YAV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, and Myeloblastosis Associated Virus (MAV) reverse transcriptase may be suitably used in the composition described herein.

20 In some embodiments, the reverse transcriptase is MMLV-RT, MarathonRT from *Eubacterium rectale*, or RTX reverse transcriptase or a variant of MMLV-RT, MarathonRT, or RTX reverse transcriptase. In some embodiments, the reverse transcriptase is a sequence shown in **Table 10**, a variant thereof, or an ortholog thereof.

25 **Table 10. Exemplary Reverse Transcriptase Sequences**

Reverse Transcriptase	Sequence	SEQ ID NO
MMLV-RT	<i>See Table 9 below</i>	53
MMLV-RT_2	MTWLSDFPQAWAETGGMGLAVRQAPLIIPKATSTPVS IKQYPMSQEARLGI KPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMG ISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLAA TSELDCCQGTRALLQTLGNLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTE ARKETVMGQFTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLF NWGPDQQKAYQEILQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTRVQFGPVVALNPATLLPLPEEG LQHNCCLDILAEAHGTRPDL	266
MMTV-RT	MFTLWGRDIMKDIKVRMLTSPDDSQDLMIGAIESNLFADQISWKS DQP VWL NQWPLKQEKLQALQQLVTEQLQLGHLEESNTPWNTPVFVIKKKSGKWRLLQD	267

	LRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIIDLQDCFFNIKLHPEDCKRF AFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYI VHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVSVTEKIQKYDNLKYLGT IQGDSVSYQKLQIRTDKLRNLDFQKLLGNINWIRPFLKLTGTGELKPLFEIL NGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTP TACLWQDGVVEWHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDI VVPYTKVQFDLLQEKEDWPISLLGFLGEVHFHLPKDPDLLTTLQTATIFPH MTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAIEIVAVIT AFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHQLRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRIL	
MarathonRT	MDTSNLMEQILSSDNLNRAYLQVVRNKAEGVDGMKYTELKEHLAKNGETIK GQLRTRKYKPPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEE QFHDHSYGRFPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKIMT LIGRTIKDGDVISIVRKYLVSIMIDDEYEDSIVGTPQGNNLSPILANIMLN ELDKEMEKRGLNFRVYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMT KSKVDRPSGLKYLGFYFDPRAHQFKAHPAKSVAKFKKRMKEITCRSWGVS SNSYKVEKLNQLIRGWINYFKIGSMKTLCKELDSRIIRYLRMCIWQWKTPO NQEKNLVKLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRSLASFLGISM LDYYIEKCVTC	268
RTX reverse transcriptase	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVK KITAERHGTVVTVKRVEKVQKKFLGRPVVEVWKLYFTHPQDVPALMDKIREHP AVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIETLYHEGEEFAEGP ILMISYADEEGARVITWKNVDLPYVDVSTEREMIKRFLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINFALGRDGESEPKIQRMGDRFAVEVKGRIHFDL YPVIRRTINLPTYTLEAVYEAVFGQPKKEKVYAAEITTAWETGENLERVARYS MEDAKVTYELGKEFLPMEAQLSRLIGQSLWDVSRSTGNLVEWFLLRKAYER NELAPNKPDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITH NVSPDTLNREGCKEYDVAPQVGHFRCKDFPGFIPSLGDLLEERQKIKKRMK ATIDPIERKLLDYRQRAIKILANSLYGYGYARARWYCKEACSVIAWGREY LTMTIKEIEEKYGFKVIYSDTDGFFATIPGADAEVKKKAMEFLKYINAKLP GALELEYEGFYKRGFLVTKKYYAVIDEEGKITTRGLEIVRRDWEIAKETQA RVLEALLKDG DVEKAVRIVKEVTEKLSKYEVPPPEKLVIHKQITRDLKDYKAT GPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDP TKHKYD AEYYIEKQVLP AVERILRAF GYRKEDLRYQKTRQVGLSARLKPKGT	269
RTX-exoMinus	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVK KITAERHGTVVTVKRVEKVQKKFLGRPVVEVWKLYFTHPQDVPALMDKIREHP AVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIETLYHEGEEFAEGP ILMISYADEEGARVITWKNVDLPYVDVSTEREMIKRFLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINFALGRDGESEPKIQRMGDRFAVEVKGRIHFDL YPVIRRTINLPTYTLEAVYEAVFGQPKKEKVYAAEITTAWETGENLERVARYS MEDAKVTYELGKEFLPMEAQLSRLIGQSLWDVSRSTGNLVEWFLLRKAYER NELAPNKPDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITH NVSPDTLNREGCKEYDVAPQVGHFRCKDFPGFIPSLGDLLEERQKIKKRMK ATIDPIERKLLDYRQRAIKILANSLYGYGYARARWYCKEACSVIAWGREY LTMTIKEIEEKYGFKVIYSDTDGFFATIPGADAEVKKKAMEFLKYINAKLP GALELEYEGFYKRGFLVTKKYYAVIDEEGKITTRGLEIVRRDWEIAKETQA RVLEALLKDG DVEKAVRIVKEVTEKLSKYEVPPPEKLVIHKQITRDLKDYKAT GPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDP TKHKYD AEYYIEKQVLP AVERILRAF GYRKEDLRYQKTRQVGLSARLKPKGT	270
B11 RT	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVK KITAERHSTVVTVKRVEKVQKKFLGRSVEVWKLYFTHPQDVPALMDKIREHP AVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLALDIGTPCHEGEVFAEGP ILMISYADEEGTRVITWRNVLDLPYVDVSTEREMIQRFRLRVVKEKDPDVLIT YNGDNFDFAYLKKRCEKLGINFALGREGSEPKIQRMGDRFAVEVKGRIHFDL YPVIRRTVNLPIY TLEAVYEAVFGQPKKEKVYAAEITTAWETGENLERVARYS MEDAKVTYELGKEFMPMEAQLSRLIGQSLWDVSRSTGNLVEWFLLRKAYER NELAPNKPDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITH NVSPDTLNREGCKEYDVAPQVGHFRCKDFPGFIPSLGDLLEERQKIKKRMK ATIDPIERKLLDYRQRAIKILANSLYGYGYARARWYCKEACSVIAWGREY ITMTIKEIEEKYGFKLIYSDTDGFFATIPGAEAEVKKKAMEFLKYINAKLP GALELEYEGFYKRGFLVTKKYYAVIDEEGKITTRGLEIVRRDWEIAKETQA RVLEALLKDG DVEKAVRIVKEVTEKLSKYEVPPPEKLVIHKQITRDLKDYKAT	271

	GPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDP TKHKYD AEYYIENQVLP AVERILRAYGYRKEDLWYQKTRQVGLSARLKPKGT	
RT CaMV	MDHLLLTQTQTEQVMNVTNPNSIYIKGRLYFKGYKKIELHCFVDTGASLCI ASKFVIPEEHVNAERPIMVKIADGSSITISKVCKDIDLIIAGEIFRIPTVY QQESGIDFIIGNNFCQLYEPFIQFTDRVIFTKNKSYPVHIAKLTRAVRVGTE GFLESMKKRSKTQQPEPVNISTNKIENPLEEIALSEGRRLSEEKLFITQQR MQKIEELLEKVCSENPLDPNKTQWMKASIKLSDPSKAIVKPKMYSMPDRE EFDKQIKELLDLKVIKPSKSPHMAPFLVNNEAEKRRGKKRMVNVYKAMKA TVGDAYNLPNKDELLTLIRGKKIFSSFDCKSGFWQVLLDQESRPLTAFTCPQ GHYEWNVVPFGLKQAPSFQRHMDFAFRVFRKFCCVYVDDILVFSNNEEDHL LHVAMILQKCNQHGIIILSKKKAQLFKKKINFLGLEIDEGTHKPQGHILEHIN KFPDTLEDKKQLQRFLGILTYASDYIPKLAQIRKPLQAKLKENVPWRWKED TLYMQVKVKNLQGFPLHHPLEEKLIETDASDDYWGMLKAIKINEGTNT ELICRYASGSFKAEEKNYHSNDKETLAVINTIKKF'SIYLPVHFLRTDNTH FKSFVNLYKGD SKLGRNIRWQAWLSHYSF DVEHIKGTDNHFADFLSREFNK VNSSGGS	272
RT retron	MGIHGVPAAMKSAEYLNTFRLRNGLPVMNNLHDMSKATRISVETLRLLIYT ADFRYRIYVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKLSSPFSIG FEKHQSILNNATPHIGANFILNIDLEDFP SLTANKVFGVFHSLGYNRLISS VLTKICCYKNLLPQAGPSSPKLANLICS KLDYRIQGYAGSRGLIYTRYADDL TLAQSMKKVVKARDFLFSIIPSEGLVINSKKTCSGPRSQRKVTGLVISQE KVGIGREKYKEIRAKIHIFCGKSSEIEHV RGWLSF ILSVDSKSHRRLITYI SKLEKKY GKNPLNKAKT	273
HIV2	MEKEGQLEEAPPTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVTQDFTEIQL GIPHPAGLAKKRRITVLDVGDAYFSIPLHEDFRPYTAFTLPSVNNAEPGRKY IYKVLPPQGWKGSPAIFQHTMRQVLEPFRKANKDVII IQYMDLILASDRTDL EHDRVVLQLKELLNGLGFS TPDEKFQKDPPIHWMGYELWPTKWKLQKIQLPQ KEIWTVNDIQKLVGVLNWAALYPGIK	274
HIV1	MEEGKISKIGPENPYNTPVFAIKKKDKSTKWRKLVDFRELNKRTQDFWEVQL GIPHPAGLKKKSVTVLDVGDAYFSVPLDESF RKYTAFTIP SINNETPGIRY QYNVLPQGWKGSPAIFQSSMTKILEPFRIKNPEIYIYQYMDLIVGSDLEIG QHRTKIEELRAHLLSWGFTTPDKKHQKEPFLWMGYELHPDRWTVQPIDLPE KDSWTVNDIQKLVGKLNWASQIYAGIK	275
HIV-1 RT p51	MPISPIETVPVKLPGMDGPKVKQWPLTEEEKIKALVEICTEMEKEGKISKIG PENPYNTPVFAIKKKDKSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKK KSVTVLDVGDAYFSVPLDEDFRKYTAFTIP SINNETPGIRYQYNVLPQGWKG SPAIFQSSMTKILEPFRKQNPDIYIYQYMDLIVGSDLEIGQHRTKIEELRQ HLLRWGLTPDKKHQKEPFLWMGYELHPDKWTVQPIVLEPKDSWTVNDIQK LVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEALELAENREILK EPVHGVYDPSKDLIAETIQKQGQGWTYQIYQEPFKNLKTGKYARMRGAHTN DVQQLTEAVQKITTESIYIWGKTPKFKLP IQKETWETWWTEYWQATWPEWE FVNTPLVLKLWYQLEKEP IVGAETF	276
MLLV RT lacking the RNase H domain	STLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP LKATSTPVSIIKQYPMSEARLGIKPHIQRLLDQILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRVEDIHP TVPNPNLLSGLPPSHQWYTVLDLKDFAF FCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLAATSELDCCQGTALLQTLGNLGYRASAKKA QICQKQVKYLG YLLKEGQRWLTEARKETVMGQTPKTPRQLREFLGKAGFCR LFIPGFAEMAAPLYPLTKPGTLFNWGPDPQKAYQEIQAALLTAPALGLPDLT KPFELFVDEKQGYAKGVL TQKLG PWRRPVAYLSKKLDPVAAGWPPCLRMVAA IAVLTKDAGKLTMGQPLVILAPHAVEALVKOPPDRWLSNARMTHYQALLLDT DRVQFGPVVALNPATLLP	277

- In some embodiments, the reverse transcriptase polypeptide is an “error-prone” reverse transcriptase variant. Error-prone reverse transcriptases that are known and/or available in the art may be used. It will be appreciated that reverse transcriptases naturally do not have any proofreading function; thus, the error rate of reverse transcriptases is generally higher than DNA

polymerases comprising a proofreading activity. In some embodiments, the reverse transcriptase is considered to be “error-prone” if it has an error rate that is less than one error in about 15,000 nucleotides synthesized.

5 In some embodiments, the reverse transcriptase polypeptide has a mutation or mutations in the RNase H domain. In some embodiments, the reverse transcriptase polypeptide does not comprise an RNase H domain (*e.g.*, the RNase H domain has been removed from the reverse transcriptase polypeptide). In some embodiments, the RNase H domain is truncated in a reverse transcriptase polypeptide. In some embodiments, the reverse transcriptase polypeptide has a mutation or mutations in the RNA-dependent DNA polymerase domain. In some embodiments, 10 the reverse transcriptase polypeptide is a variant that has altered thermostability characteristics. The ability of a reverse transcriptase to withstand high temperatures is an important aspect of cDNA synthesis. Elevated reaction temperatures help denature RNA with strong secondary structures and/or high GC content, allowing reverse transcriptases to read through the sequence. As a result, reverse transcription at higher temperatures enables full-length cDNA synthesis and higher yields. Wild-type M-MLV reverse transcriptase typically has an optimal temperature in 15 the range of 37-48°C; however, mutations may be introduced that allow for the reverse transcription activity at higher temperatures of over 48°C, including 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, and higher.

20 Variant reverse transcriptase polypeptides used herein may be at least about 20% identical, at least about 25% identical, at least about 30% identical, at least about 35% identical, at least about 40% identical, at least about 45% identical, at least about 50% identical, at least about 55% identical, at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% 25 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 reverse transcriptase polypeptide, including any wild-type reverse transcriptase, mutant reverse transcriptase, or fragment of a reverse transcriptase, or other reverse transcriptase variant disclosed or 30 contemplated herein or known in the art. In some embodiments, a reverse transcriptase 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, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or up to 100, or up to 200, or up to 300, or up to 400, or up to 500 or more amino acid changes compared to a reference reverse transcriptase. In some embodiments, the reverse transcriptase

variant comprises a fragment of a reference reverse transcriptase, such that the fragment is at least about 20% identical, at least about 25% identical, at least about 30% identical, at least about 35% identical, at least about 40% identical, at least about 45% identical, at least about 50% identical, at least about 55% identical, at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% 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 the reference reverse transcriptase.

Variant reverse transcriptases, including error-prone reverse transcriptases, thermostable reverse transcriptases, and reverse transcriptases with increased processivity, can be engineered by various routine strategies, including mutagenesis or evolutionary processes. In some cases, the variants can be produced by introducing a single mutation. In other cases, the variants may require more than one mutation. For those mutants comprising more than one mutation, the effect of a given mutation may be evaluated by introduction of the identified mutation to the wild-type gene by site-directed mutagenesis in isolation from the other mutations borne by the particular mutant. Screening assays of the single mutant thus produced will then allow the determination of the effect of that mutation alone.

In some embodiments, the reverse transcriptase polypeptides comprise or is fused to a domain to improve extension rates and/or efficiency of the reverse transcriptase. In some embodiments, the reverse transcriptase polypeptide is fused to an Sso7d polypeptide such as an Sso7d polypeptide from *Sulfolobus solfataricus*. See, e.g., Wang et al., Nucleic Acids Res. 32(3): 1197-207 (2004).

In some embodiments, the reverse transcriptase as in any one of the embodiments described herein interacts with a ligase, an integrase, and/or a recombinase. In some embodiments, the reverse transcriptase as in any one of the embodiments described herein is fused to a ligase, an integrase, and/or a recombinase. In some embodiments, the ligase, integrase, and/or recombinase is fused to the N-terminus or C-terminus of the reverse transcriptase. In some embodiments, the ligase, integrase, and/or recombinase is fused internally to the reverse transcriptase. In some embodiments, the integrase is a serine integrase. In some embodiments, the integrase is a Bxb1, TP901, or PhiBT1 integrase. In some embodiments, the recombinase is a serine recombinase or a tyrosine recombinase. In some embodiments, the recombinase is a CRE recombinase. In some embodiments, a reverse transcriptase that interacts with or is fused to a ligase, integrase, and/or recombinase further interacts with or is fused to the CRISPR nuclease

polypeptide disclosed herein.

In other embodiments, the gene editing system provided herein may comprise a nucleic acid encoding the RT polypeptide. In some examples, the nucleotide sequence encoding the RT polypeptide described herein can be codon-optimized for use in a particular host cell or organism. In some examples, the nucleic acid encoding the RT polypeptides can be an mRNA molecule, which can be codon optimized. Exemplary codon-optimized nucleotide sequences encoding exemplary RT polypeptides can be found in **Table 7** below, which is within the scope of the present disclosure. In some examples, the gene editing system may comprise a vector (*e.g.*, a viral vector such as an AAV vector, an AdV vector, or a retroviral vector) encoding the RT polypeptide.

(iii) Fusion Polypeptides

In some embodiments, the gene editing system provided herein comprise a fusion polypeptide that includes both the CRISPR nuclease polypeptide and the RT polypeptide as also disclosed herein. Alternatively, the gene editing system may comprise a nucleic acid (*e.g.*, a vector such as an expression vector) encoding the fusion polypeptide.

As used herein, the terms “fusion” and “fused” refer to the joining of at least two nucleotide or protein molecules. For example, “fusion” and “fused” can refer to the joining of at least two polypeptide domains that are encoded by separate genes (*e.g.*, the CRISPR nuclease polypeptide and the reverse transcriptase polypeptide provided herein) in nature. The fusion can be an N-terminal fusion, a C-terminal fusion, or an intramolecular fusion.

In some embodiments, the fusion polypeptide may comprise the reverse transcriptase polypeptide at its N-terminus and the CRISPR nuclease polypeptide downstream to the RT polypeptide. In other embodiments, the fusion polypeptide may comprise the CRISPR nuclease polypeptide at its N-terminus and the RT polypeptide downstream to the CRISPR nuclease polypeptide. In some embodiments, the RT polypeptide may be fused with the CRISPR nuclease polypeptide at an intramolecular position within the RT polypeptide, for example, the CRISPR nuclease polypeptide may be within a loop of the reverse transcriptase polypeptide.

Any of the CRISPR nuclease polypeptide disclosed herein and any of the RT polypeptides disclosed herein may be used for constructing the fusion polypeptides. In some instances, the CRISPR nuclease polypeptide may be a CRISPR nuclease such as SEQ ID NO: 1. Alternatively, the CRISPR nuclease polypeptide may be a variant of the reference CRISPR nuclease of SEQ ID NO: 1. For example, the variant may be a nickase, *e.g.*, having any of the mutations noted above in the HNH domain relative to the reference CRISPR nuclease, for

example, the nickase of SEQ ID NO: 32. In other examples, the variant may comprise a combination of mutation(s) in the HNH domain and one or more arginine and/or lysine substitutions as those disclosed herein.

5 In some embodiments, any of the CRISPR nuclease-RT fusion polypeptides disclosed herein may comprise one or more additional functional elements, *e.g.*, those provided herein. In some instances, the additional functional elements may be one or more NLS elements. In some examples, the fusion polypeptide may comprise an NLS at its N-terminus, at its C-terminus, or both. Alternatively or in addition, the additional functional elements may be a flexible peptide linker, which can be located between the CRISPR nuclease polypeptide and the RT polypeptide.
10 Suitable peptide linkers include, but are not limited to, G/S rich peptide linkers and XTEN peptide linkers. Examples of NLS and peptide linkers are provided in **Tables 1, 5, and 15** below. See also **Examples 1 and 7**.

In some examples, the CRISPR nuclease-RT fusion polypeptide provided herein comprises a peptide linker (the first peptide linker) located between the CRISPR nuclease
15 polypeptide and the RT polypeptide. In some instances, the CRISPR nuclease polypeptide is N-terminal to the RT polypeptide. In other instances, CRISPR nuclease polypeptide is C-terminal to the RT polypeptide. In some instances, an additional peptide linker and/or one or more NLS signals may be located between the CRISPR nuclease polypeptide and the RT polypeptide. For example, an additional peptide linker and an NLS may be placed between the CRISPR nuclease
20 polypeptide and the RT polypeptide, in addition to the first peptide linker. In some specific examples, the peptide linker between the CRISPR nuclease polypeptide and the RT polypeptide is at least 20-aa in length, for example, ranging from about 20 amino acids to 100 amino acids.

Alternatively or in addition, the CRISPR nuclease-RT fusion polypeptide provided herein may comprise at least two NLSs (the first NLS and the second NLS), at least one of
25 which is located at the N-terminus or the C-terminus of the fusion polypeptide. In some examples, one of the two NLSs is located at the N-terminus and the other one is located at the C-terminus. In other examples, the two NLSs are located at the N-terminus. In yet other examples, the two NLSs are located at the C-terminus.

In some instances, the CRISPR nuclease-RT fusion polypeptide provided herein may
30 comprise one or more additional NLS(s) (*e.g.*, a third NLS, and optionally a fourth NLS). Such additional NLS(s) may be located between the CRISPR nuclease polypeptide and the RT nuclease. In other examples, the additional NLS(s) may be located between the CRISPR nuclease/RT polypeptide and a terminal NLS, optionally via a peptide linker.

In some examples, the CRISPR nuclease-RT fusion polypeptide disclosed herein may

comprise, from N-terminus to C-terminus, a first NLS, the CRISPR nuclease, a peptide linker, the RT polypeptide, and a second NLS. In other examples, the CRISPR nuclease-RT polypeptide disclosed herein may comprise, from N-terminus to C-terminus, a first NLS, the RT nuclease, a peptide linker, the CRISPR nuclease polypeptide, and a second NLS (which may be identical to the first NLS), the CRISPR nuclease polypeptide, a first peptide linker, a second NLS (which may be identical to the first NLS), a second peptide linker, the RT nuclease, a third peptide linker (which may be identical to the first peptide linker), a third NLS (which may be identical to the first NLS), a fourth peptide linker, and a fourth NLS. Examples of the CRISPR nuclease-RT polypeptides are provided in **Table 8** below, all of which are within the scope of the present disclosure.

In other examples, the CRISPR nuclease-RT fusion polypeptide disclosed herein may have the configuration set forth in **Table 16** below (from N-terminus to C-terminus). In some instances, the CRISPR nuclease-RT fusion polypeptide does not include the FLAG motif in any of the configurations listed in **Table 16**. In one instance, the CRISPR nuclease-RT fusion polypeptide has a configuration corresponding to Config4, except that the FLAG motif is removed. In another instance, the CRISPR nuclease-RT fusion polypeptide has a configuration corresponding to Config5, except that the FLAG motif is removed. In yet another instance, the CRISPR nuclease-RT fusion polypeptide has a configuration corresponding to Config7, except that the FLAG motif is removed. In still another instance, the CRISPR nuclease-RT fusion polypeptide has a configuration corresponding to Config9, except that the FLAG motif is removed. Alternatively, the CRISPR nuclease-RT fusion polypeptide has a configuration corresponding to Config10, except that the FLAG motif is removed. Additional exemplary CRISPR nuclease-RT fusion polypeptides are provided in **Table 17**. Variants of these fusion polypeptides with the FLAG motif removed are also within the scope of the present disclosure.

In some examples, any of the CRISPR nuclease-RT fusion polypeptide may comprise a CRISPR nuclease variant. In some specific examples, the CRISPR variant is a nickase, such as those provided herein (*e.g.*, comprising the mutations as those listed in **Table 5**). In some instances, the nickase comprises one or more mutations in the HNH domain relative to the reference nuclease of SEQ ID NO: 1, for example, at position H845, *e.g.*, H845A.

In other embodiments, the gene editing system provided herein may comprise a nucleic acid encoding the CRISPR nuclease-RT fusion polypeptide. In some examples, the nucleotide sequence encoding the fusion polypeptide described herein can be codon-optimized for use in a particular host cell or organism. In some examples, the nucleic acid encoding the fusion polypeptides as disclosed herein can be an mRNA molecule, which can be codon optimized.

Exemplary codon-optimized nucleotide sequences encoding exemplary CRISPR nuclease-RT fusion polypeptides can be found in **Tables 8** and **17** below, any of which is within the scope of the present disclosure. See also **Example 5** and **Example 7**. Coding sequences for variants of the exemplary CRISPR nuclease-RT fusion polypeptides with the FLAG motif removed are also
5 within the scope of the present disclosure.

In some examples, the gene editing system may comprise a vector (*e.g.*, a viral vector such as an AAV vector, an AdV vector, or a retroviral vector) encoding the CRISPR nuclease-RT fusion polypeptide.

10 (iv) Preparation of Protein Components

The CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide disclosed herein may be prepared by conventional methods or the methods disclosed herein. For example, the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide can be prepared by culturing host cells such as bacteria
15 cells or mammalian cells, capable of producing the nuclease polypeptides, isolating the nuclease polypeptides thus produced, and optionally, purifying the nuclease polypeptides. The CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide can be also prepared by an *in vitro* coupled transcription-translation system.

Host cells that can be used for preparation of the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide are not particularly limited as long as they can produce the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide. Some nonlimiting examples of the host cells include bacteria cells (*e.g.*, *E. coli* cells), yeast cells, insect cells, or mammalian cells.

25 *Vectors*

The present disclosure provides vectors for expressing the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide. In some embodiments, a vector disclosed herein includes a nucleotide sequence encoding CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide. In some embodiments, the
30 vector comprises a Pol II promoter or a Pol III promoter.

Expression of natural or synthetic polynucleotides is typically achieved by operably linking a polynucleotide encoding the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide to a promoter and incorporating the construct into an expression vector. The expression vector is not particularly limited as long as it includes a

polynucleotide encoding the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide and can be suitable for replication and integration in eukaryotic cells.

Typical expression vectors include transcription and translation terminators, initiation
5 sequences, and promoters useful for expression of the desired polynucleotide. For example, plasmid vectors carrying a recognition sequence for RNA polymerase (pSP64, pBluescript, etc.). may be used. Vectors including those derived from retroviruses such as lentivirus are suitable tools to achieve long-term gene transfer since they allow long-term, stable integration of a transgene and its propagation in daughter cells. Examples of vectors include expression vectors,
10 replication vectors, probe generation vectors, and sequencing vectors. The expression vector may be provided to a cell in the form of a viral vector.

Viral vector technology is well known in the art and described in a variety of virology and molecular biology manuals. Viruses useful as vectors include, but are not limited to phage viruses, retroviruses, adenoviruses, adeno-associated viruses, herpes viruses, and lentiviruses. In
15 general, a suitable vector contains an origin of replication functional in at least one organism, a promoter sequence, convenient restriction endonuclease sites, and one or more selectable markers.

The kind of the vector is not particularly limited, and a vector that can be expressed in host cells can be appropriately selected. To be more specific, depending on the kind of the host
20 cell, a promoter sequence to ensure the expression of the polypeptide(s) from the polynucleotide is appropriately selected, and this promoter sequence and the polynucleotide are inserted into any of various plasmids etc. for preparation of the expression vector.

Additional promoter elements, *e.g.*, enhancing sequences, regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the
25 start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. Depending on the promoter, it appears that individual elements can function either cooperatively or independently to activate transcription.

Further, the disclosure should not be limited to the use of constitutive promoters. Inducible promoters are also contemplated as part of the disclosure. The use of an inducible
30 promoter provides a molecular switch capable of turning on expression of the polynucleotide sequence which it is operatively linked when such expression is desired or turning off the expression when expression is not desired. Examples of inducible promoters include, but are not limited to a metallothionein promoter, a glucocorticoid promoter, a progesterone promoter, and a tetracycline promoter.

The expression vector to be introduced can also contain either a selectable marker gene or a reporter gene or both to facilitate identification and selection of expressing cells from the population of cells sought to be transfected or infected through viral vectors. In other aspects, the selectable marker may be carried on a separate piece of DNA and used in a co-transfection procedure. Both selectable markers and reporter genes may be flanked with appropriate transcriptional control sequences to enable expression in the host cells. Examples of such a marker include a dihydrofolate reductase gene and a neomycin resistance gene for eukaryotic cell culture; and a tetracycline resistance gene and an ampicillin resistance gene for culture of *E. coli* and other bacteria. By use of such a selection marker, it can be confirmed whether the polynucleotide encoding the polypeptide(s) of the present invention has been transferred into the host cells and then expressed without fail.

The preparation method using recombinant expression vectors is not particularly limited, and examples thereof include methods using a plasmid, a phage or a cosmid.

15 *Methods of Expression*

The present disclosure includes a method for protein expression, comprising translating the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide described herein.

In some embodiments, a host cell described herein is used to express the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide. The host cell is not particularly limited, and various known cells can be preferably used. Specific examples of the host cell include bacteria such as *E. coli*, yeasts (budding yeast, *Saccharomyces cerevisiae*, and fission yeast, *Schizosaccharomyces pombe*), nematodes (*Caenorhabditis elegans*), *Xenopus laevis* oocytes, and animal cells (for example, CHO cells, COS cells and HEK293 cells). The method for transferring the expression vector described above into host cells, *i.e.*, the transformation method, is not particularly limited, and known methods such as electroporation, the calcium phosphate method, the liposome method and the DEAE dextran method can be used.

After a host is transformed with the expression vector, the host cells may be cultured, cultivated or bred, for production the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide. After expression, the host cells can be collected and the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide purified from the cultures etc. according to conventional methods (for example, filtration, centrifugation, cell disruption, gel filtration chromatography, ion exchange

chromatography, etc.).

A variety of methods can be used to determine the level of production of a mature CRISPR nuclease polypeptide, a mature RT polypeptide, or a mature CRISPR nuclease-RT fusion polypeptide in a host cell. Such methods include, but are not limited to, for example, methods that utilize either polyclonal or monoclonal antibodies specific for the proteins or a labeling tag as described elsewhere herein. Exemplary methods include, but are not limited to, enzyme-linked immunosorbent assays (ELISA), radioimmunoassays (RIA), fluorescent immunoassays (FIA), and fluorescent activated cell sorting (FACS). These and other assays are well known in the art (See, *e.g.*, Maddox et al., J. Exp. Med. 158:1211 [1983]).

The present disclosure provides methods of *in vivo* expression of the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide (and optionally the gRNA and/or the RT donor RNA in the gene editing system disclosed herein). Such a method may comprise providing a polyribonucleotide encoding the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide to a host cell in a subject (*e.g.*, a human subject) wherein the polyribonucleotide encodes the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide and expressing the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide from the cell.

B. RNA Components

The gene editing systems provided herein also involve at least two RNA components, a guide RNA (gRNA), which directs gene editing at a desired genetic site, and an RT donor RNA, which serves as the RNA template for the RT polypeptide in reverse transcription. The RT donor RNA comprises desired nucleotide substitutions to be inserted into the genetic site of interest. In some embodiments, the gene editing system comprises the two RNA molecules. In specific examples, the gene editing system may comprise a single RNA molecule comprising the gRNA and the RT donor RNA. Alternatively, the gene editing system may comprise one or more nucleic acids encoding the two RNA components. For example, the gene editing system may comprise one or more expression vectors (*e.g.*, viral vectors such as retroviral vectors, adenoviral vectors, or adeno-associated viral vectors) capable of producing the gRNA, the RT donor RNA, or the single RNA molecule comprising such.

In some embodiments, the gRNA and the RT donor RNA as disclosed herein may form a complex.

(i) Guide RNAs

The gene editing system disclosed herein further comprises one or more gRNAs or nucleic acid(s) encoding such. As used herein, the terms “RNA guide”, “RNA guide sequence,” or “guide RNA (gRNA)” refer to an RNA molecule or a modified RNA molecule that facilitates the targeting of a CRISPR nuclease described herein to a genomic site of interest. For example, an RNA guide can be a molecule that comprises a spacer sequence and a scaffold sequence. The spacer sequence recognizes (*e.g.*, binds to) a site in a non-PAM strand that is complementary to a target sequence in the PAM strand, *e.g.*, designed to be complementary to a specific nucleic acid sequence. The scaffold sequence contains a nuclease binding sequence for binding to the CRISPR nuclease. In some embodiments, the scaffold is an RNA sequence.

In some instances, the gRNA disclosed herein may further comprise a linker sequence, a 5' end and/or 3' end protection fragment, or a combination thereof.

Spacer Sequences

As used herein, the term “spacer” and “spacer sequence” (*a.k.a.*, a DNA-binding sequence) is a portion in an RNA guide that is the RNA equivalent of the target sequence (a DNA sequence). The spacer contains a sequence capable of binding to the non-PAM strand via base-pairing at the site complementary to the target sequence (which is in the PAM strand). Such a spacer is also known as specific to the target sequence. In some instances, the spacer may be at least 75% identical to the target sequence (*e.g.*, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99%), except for the RNA-DNA sequence difference. In some instances, the spacer may be 100% identical to the target sequence except for the RNA-DNA sequence difference.

The gene editing system disclosed herein comprises one or more gRNAs, each comprising a spacer for targeting a genomic site of interest and a scaffold, which is recognizable by the variant CRISPR nuclease polypeptide contained in the gene editing system. The target sequence can be adjacent to a protospacer adjacent motif (PAM) of 5'-NDR-3', in which N represents any nucleotide, D represents A, G, or T, and R represents A or G. In some examples, the PAM is 5'-NRR-3', in which N represents any nucleotide and R represents A or G. For example, the PAM may be 5'-NRG-3', in which N represents any nucleotide and R represents A or G. In specific examples, the PAM motif is 5'-NGG-3' in which N represents any nucleotide. In some examples, the PAM sequence recognizable by the CRISPR nuclease polypeptide may be 5'-NGN-3', 5'-NRN-3', or 5'-NYN-3', in which N represents any nucleotide, R represents A or G, and Y represents C or T.

The PAM motif is located on the 3' end of the target sequence. As used herein, the term "protospacer adjacent motif" or "PAM sequence" refers to a DNA sequence adjacent to a target sequence. In some embodiments, a PAM sequence is required for binding of the CRISPR nuclease and/or indel activity. In a double-stranded DNA molecule, the strand containing the PAM motif is called the "PAM-strand" and the complementary strand is called the "non-PAM strand." The gRNA binds to a site in the non-PAM strand that is complementary to a target sequence disclosed herein, and the PAM sequence as described herein is present in the PAM-strand. The PAM motif can be located upstream to the target sequence.

As used herein, the term "adjacent to" refers to a nucleotide or amino acid sequence in close proximity to another nucleotide or amino acid sequence. In some embodiments, a nucleotide sequence is adjacent to another nucleotide sequence if no nucleotides separate the two sequences (*i.e.*, immediately adjacent). In some embodiments, a nucleotide sequence is adjacent to another nucleotide sequence if a small number of nucleotides separate the two sequences (*e.g.*, about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides).

A spacer sequence as disclosed herein may have a length of from about 15 nucleotides to about 30 nucleotides. For example, the spacer sequence can have a length of from about 15 nucleotides to about 20 nucleotides, from about 15 nucleotides to about 25 nucleotides, from about 20 nucleotides to about 25 nucleotides, or from about 20 nucleotides to about 30 nucleotides. In some embodiments, the spacer in the gRNA may be generally designed to have a length of between 15 and 25 nucleotides (*e.g.*, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25) and be complementary to a specific target sequence. In some embodiments, the spacer sequence may be designed to have a length of between 18-22 nucleotides (*e.g.*, 20 nucleotides).

In some embodiments, the spacer sequence may have at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 99.5% sequence identity to a target sequence as described herein and is capable of binding to the complementary region of the target sequence via base-pairing.

In some embodiments, the spacer sequence comprises only RNA bases. In some embodiments, the spacer sequence comprises a DNA base (*e.g.*, the spacer comprises at least one thymine). In some embodiments, the spacer sequence comprises RNA bases and DNA bases (*e.g.*, the DNA-binding sequence comprises at least one thymine and at least one uracil).

Scaffold Sequence

The scaffold sequence in the gRNA is recognizable by the variant CRISPR nuclease polypeptide also in the gene editing system. In some instances, the scaffold sequence comprises SEQ ID NO: 2, which is the cognate scaffold for the reference CRISPR nuclease of SEQ ID NO: 1.

5 GUUUUAGAGCUGUGCUGAAAAGCACAGCACGUUAAAAUAAGGCAGUGAUU
GAAAAAUCCAGUCCGUUUUCAGCUUGAAAAAGUGAGCACCGAAUCGGUGC
UU (SEQ ID NO: 2)

In other instances, the scaffold sequence may be a variant derived from SEQ ID NO: 2.

10 Such a variant scaffold sequence may comprise a nucleotide sequence at least 80% (e.g., at least 85%, 90%, 95%, 98%, or greater) identical to SEQ ID NO: 2. Alternatively or in addition, the variant scaffold sequence may comprise deletions, nucleotide substitutions, or a combination thereof. The variant CRISPR nuclease polypeptide may have increased binding to the variant scaffold sequence as compared with the scaffold of SEQ ID NO: 2. In some examples, the

15 variant scaffold may be a fragment of SEQ ID NO: 2 or a variant thereof as disclosed herein. For example, the variant scaffold for use in the gRNAs provided herein may have a length ranging from 100-150 nucleotides.

In a gRNA, the scaffold may be located at the 3' end of the spacer. In some instances, the scaffold and spacer are connected directly. In other instances, the scaffold and spacer may be

20 connected via a nucleotide linker.

(ii) RT Donor RNA

As used herein, the terms "reverse transcription donor RNA" or "RT donor RNA" refer to an RNA molecule comprising a reverse transcription template sequence (RTT sequence) and a

25 primer binding site (PBS). An RT donor RNA may be fused to an RNA guide at either the 5' end or 3' end of the RNA guide.

Any of the RT donor RNAs disclosed herein comprises: (i) a primer binding site (PBS), and (ii) an RTT sequence. In some instances, the RT donor RNA may further comprise: (iii) a nucleotide linker sequence, (iv) a 5' end and/or 3' end protection fragment (see disclosures

30 herein), or a combination thereof. In some examples, the 5' end or 3' end protection fragment (e.g., 3' extension) may comprise a pseudoknot motif to protect against 3' exonuclease activity.

In some embodiments, a RT donor RNA comprises an aptamer. In some embodiments, the aptamer recruits a reverse transcriptase polypeptide.

35 *Primer Binding Site (PBS)*

In some embodiments, the PBS in an RT donor RNA as disclosed herein is an RNA sequence capable of binding to a DNA strand via base-pairing. The DNA strand has been or can be nicked or cleaved by the CRISPR nuclease polypeptide of the gene editing system disclosed herein. In some embodiments, the PBS comprises an RNA sequence capable of binding to a
5 DNA strand (a PBS-targeting site) via base-pairing. The DNA strand may have a free 3' end or a 3' free end can be generated via cleavage by the CRISPR nuclease polypeptide contained in the same gene editing system. In some examples, the PBS-targeting site may be located on the same DNA strand as the PAM sequence (the PAM strand).

In some embodiments, the PBS may be about 5-50 nucleotides in length. For example,
10 the PBS may be about 5-40, 5-30, or 5-20 nucleotides in length. In specific examples, the PBS may be about 5-20 (*e.g.*, 7-17) nucleotides in length. In some examples, the PBS may contain 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, 11 nucleotides, 12 nucleotides, 13 nucleotides, 14 nucleotides, 15 nucleotides, 16 nucleotides, 17 nucleotides, 18 nucleotides, 19 nucleotides, or 20 nucleotides.

As used herein, the term "PBS-targeting site" refers to the region to which a PBS binds.
15 The PBS-targeting site may be adjacent to (*e.g.*, upstream to) the PAM. In a gene editing system comprising a CRISPR nuclease polypeptide that is a nickase variant (*e.g.*, comprises a disrupted HNH nuclease domain as disclosed herein), the PBS in the RT donor RNA may bind to a region (the PBS-targeting site) on the PAM strand. In some embodiments, the PBS-targeting site may
20 partially or completely overlap with the target sequence. In some instances, the PBS-targeting site may be located upstream to the PAM sequence. For example, the PBS-targeting site may be up to 100 nucleotides upstream to the PAM sequence, for example, up to 50 nucleotides, up to 30 nucleotides, up to 25 nucleotides, up to 20 nucleotides, up to 15 nucleotides, up to 10 nucleotides, or up to 5 nucleotides upstream to the PAM sequence. In specific examples, the
25 PBS-targeting site may start about 3 nucleotides to about 10 nucleotides upstream of the PAM sequence (*i.e.*, the 5'-most nucleotide of the PBS may bind about 3 nucleotides, 4, nucleotides, 5, nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, or 10 nucleotides upstream of the PAM.) In specific examples, the PBS-targeting site may start 1 nucleotide, 1-2 nucleotides, 1-3 nucleotides, 1-4 nucleotides, or 1-5 nucleotides, upstream of the PAM sequence.
30 When a free 3' end is generated by the CRISPR nuclease polypeptide in the gene editing system within or nearby the target sequence, the PBS binding to the PAM strand at a site upstream to the PAM sequence could efficiently facilitate DNA synthesis by the RT polypeptide in the gene editing system, starting from the free 3' end generated in the PAM strand.

Reverse Transcription Template (RTT) Sequence

The reverse transcription template sequence (RTT sequence) serves as the template for the reverse transcription mediated by the RT polypeptide in the gene editing system disclosed herein. In some embodiments, the RTT sequence comprises a sequence with at least one encoded edit. In some embodiments, the RTT sequence comprises sequence homology to a target sequence or its complementary region with at least one encoded edit. In some embodiments, the RTT 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. In some embodiments, the RTT sequence is about 10 nucleotides, 20 nucleotides, 30 nucleotides, 40 nucleotides, 50 nucleotides, 60 nucleotides, 70 nucleotides, 80 nucleotides, 90 nucleotides, 100 nucleotides, 110 nucleotides, or 120 nucleotides in length or any length in between.

In some embodiments, the RTT sequence is about 10 nucleotides. In some embodiments, the RTT sequence is about 11 nucleotides. In some embodiments, the RTT sequence is about 12 nucleotides. In some embodiments, the RTT sequence is about 13 nucleotides. In some embodiments, the RTT sequence is about 14 nucleotides. In some embodiments, the RTT sequence is about 15 nucleotides. In some embodiments, the RTT sequence is about 16 nucleotides. In some embodiments, the RTT sequence is about 17 nucleotides. In some embodiments, the RTT sequence is about 18 nucleotides. In some embodiments, the RTT sequence is about 19 nucleotides. In some embodiments, the RTT sequence is about 20 nucleotides. In some embodiments, the RTT sequence is about 21 nucleotides. In some embodiments, the RTT sequence is about 22 nucleotides. In some embodiments, the RTT sequence is about 23 nucleotides. In some embodiments, the RTT sequence is about 24 nucleotides. In some embodiments, the RTT sequence is about 25 nucleotides. In some embodiments, the RTT sequence is about 26 nucleotides. In some embodiments, the RTT sequence is about 27 nucleotides. In some embodiments, the RTT sequence is about 28 nucleotides. In some embodiments, the RTT sequence is about 29 nucleotides. In some embodiments, the RTT sequence is about 30 nucleotides.

In some embodiments, the reverse transcription template sequence comprises at least one

encoded edit (*e.g.*, at least two) relative to a target sequence. In some embodiments, the at least one encoded edit comprises at least one substitution, insertion, and/or deletion. In some embodiments, the edit in the target sequence comprises a substitution, an insertion, and/or a deletion relative to the sequence of a target sequence. In some embodiments, the reverse transcription template sequence comprises at least one LoxP site.

In some embodiments, the edit can be a single or multi-nucleotide substitution, such as a G to T substitution, a G to A substitution, a G to C substitution, a T to G substitution, a T to A substitution, a T to C substitution, a C to G substitution, a C to T substitution, a C to A substitution, an A to T substitution, an A to G substitution, or an A to C substitution. In some embodiments, the change in sequence can convert a G:C base pair to a T:A base pair, a G:C base pair to an A:T base pair, a G:C base pair to C:G base pair, a T:A base pair to a G:C base pair, a T:A base pair to an A:T base pair, a T:A base pair to a C:G base pair, a C:G base pair to a G:C base pair, a C:G base pair to a T:A base pair, a C:G base pair to an A:T base pair, an A:T base pair to a T:A base pair, an A:T base pair to a G:C base pair, or an A:T base pair to a C:G base pair.

In some embodiments, a template sequence described herein may further introduce one or more silent mutations. As used herein, a silent mutation refers to a mutation that does not change the amino acid residue encoded by the codon comprising the mutation. The RTT sequence can be transcribed into DNA by the reverse transcriptase of the gene editing system described herein. In some embodiments, the RTT sequence is transcribed from 5' to 3' into DNA of the PAM strand.

In some embodiments, the RTT sequence is 5' of the PBS. In some embodiments, the RTT sequence is 3' of the PBS. In some instances, the PBS and the RTT sequence in the RT donor RNA provided herein may be connected via a linker sequence. In some embodiments, the RTT and an end protection fragment (*e.g.*, a 3' end protection fragment) may be connected via a linker sequence to avoid steric hindrance between the two RNA components.

(iii) Single RNA Molecule

In some embodiments, the gene editing system provided herein comprises a single RNA molecule, which includes both the gRNA and the RT donor RNA, or a nucleic acid encoding the single RNA molecule. Such a single RNA molecule is capable of mediating cleavage at a target sequence within a genomic site of interest by the CRISPR nuclease polypeptide and synthesis of a DNA fragment from a free 3' end of a free DNA strand generated by the CRISPR nuclease polypeptide cleavage based on the RTT sequence in the single RNA molecule.

In some embodiments, the single RNA molecule may comprise the RNA guide linked to the RT donor RNA, optionally via a linker. In some examples, the single RNA molecule, from 5' to 3' end, comprises a spacer sequence, a scaffold sequence recognizable by the CRISPR nuclease polypeptide, a RTT sequence, and a PBS. In specific examples, the single RNA molecule may comprise, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT sequence, a PBS, and a protection fragment.

Any of the single RNA molecules provided herein may further comprise a linker, which may be located between a scaffold sequence and an RTT or following a PBS. In some examples, the linker may comprise a hairpin structure. In some examples, the linker may comprise an aptamer domain.

In some examples, the 5' end and/or the 3' end of the single RNA molecule, or the gRNA and/or RT donor RNA, may contain a protection fragment, which may enhance resistance of the RNA molecule to exonuclease activity. In some instances, the end protection fragment may comprise a nucleotide sequence capable of forming a secondary structure, such as hairpin, a circularization, a pseudoknot, or a triplex structure. In other instances, the end protection fragment may comprise the sequence of an exoribonuclease-resistant RNA (xrRNA), a transfer RNA (tRNA), or a truncated tRNA. In some embodiments, the modification is a Zika-like pseudoknot, a murine leukemia virus pseudoknot (MLV-PK) sequence, a red clover necrotic mosaic virus (RCNMV) sequence, a sweet clover necrotic mosaic virus (SCNMV) sequence, a carnation ringspot virus (CRSV) sequence, a preQ1 aptamer sequence, a truncated preQ1 aptamer sequence, a boxB RNA sequence, or an RNA bacteriophage MS2 sequence.

In some examples, the 5' end of the single RNA molecule (or the 5' end of the gRNA and/or the RT donor RNA when separate RNA molecules are used) may contain a 5' extension motif, which can be any of the protection fragments disclosed herein. In some examples, the 3' end of the single RNA molecule (or the 3' end of the gRNA and/or the RT donor RNA when separate RNA molecules are used) may contain a 3' extension motif, which can be any of the protection fragments disclosed herein. One specific example of the 3' extension motif is provided in **Example 6** below.

30 (iv) Modifications of Nucleic Acids

Any of the RNA components in a gene editing system as disclosed herein, *e.g.*, the single RNA molecule, the gRNA, and/or the RT donor RNA, may include one or more modifications. Exemplary modifications can include any modification to the sugar, the nucleobase, the internucleoside linkage (*e.g.*, to a linking phosphate/to a phosphodiester linkage/to the

phosphodiester backbone), and any combination thereof. Some of the exemplary modifications provided herein are described in detail below.

Any of the RNA components disclosed herein may include any useful modification, such as to the sugar, the nucleobase, or the internucleoside linkage (*e.g.*, to a linking phosphate/to a phosphodiester linkage/to the phosphodiester backbone). One or more atoms of a pyrimidine nucleobase may be replaced or substituted with optionally substituted amino, optionally substituted thiol, optionally substituted alkyl (*e.g.*, methyl or ethyl), or halo (*e.g.*, chloro or fluoro). One or more atoms of a purine nucleobase may be replaced or substituted with optionally substituted amino, optionally substituted thiol, optionally substituted alkyl (*e.g.*, methyl or ethyl), or halo (*e.g.*, chloro or fluoro). In certain embodiments, modifications (*e.g.*, one or more modifications) are present in each of the sugar and the internucleoside linkage. Modifications may be modifications of ribonucleic acids (RNAs) to deoxyribonucleic acids (DNAs), threose nucleic acids (TNAs), glycol nucleic acids (GNAs), peptide nucleic acids (PNAs), locked nucleic acids (LNAs) or hybrids thereof). Additional modifications are described herein.

In some embodiments, any of the RNA components in a gene editing system as disclosed herein comprises an abasic site (*i.e.*, a location that does not have a purine or a pyrimidine). In some embodiments, the abasic site (also referred to as an apurinic/apyrimidinic site) is present in an editing template RNA. For example, an abasic site can be present in the RTT of an editing template RNA. In some embodiments, activity of a reverse transcriptase is halted at or near an abasic site.

In some embodiments, the modification may include a chemical or cellular induced modification. For example, some nonlimiting examples of intracellular RNA modifications are described by Lewis and Pan in “RNA modifications and structures cooperate to guide RNA-protein interactions” from Nat Reviews Mol Cell Biol, 2017, 18:202-210.

Different sugar modifications, nucleotide modifications, and/or internucleoside linkages (*e.g.*, backbone structures) may exist at various positions in the sequence. One of ordinary skill in the art will appreciate that the nucleotide analogs or other modification(s) may be located at any position(s) of the sequence, such that the function of the sequence is not substantially decreased. The sequence may include from about 1% to about 100% modified nucleotides (either in relation to overall nucleotide content, or in relation to one or more types of nucleotide, *i.e.* any one or more of A, G, U or C) or any intervening percentage (*e.g.*, from 1% to 20%, from 1% to 25%, from 1% to 50%, from 1% to 60%, from 1% to 70%, from 1% to 80%, from 1% to 90%, from 1% to 95%, from 10% to 20%, from 10% to 25%, from 10% to 50%, from

10% to 60%, from 10% to 70%, from 10% to 80%, from 10% to 90%, from 10% to 95%, from 10% to 100%, from 20% to 25%, from 20% to 50%, from 20% to 60%, from 20% to 70%, from 20% to 80%, from 20% to 90%, from 20% to 95%, from 20% to 100%, from 50% to 60%, from 50% to 70%, from 50% to 80%, from 50% to 90%, from 50% to 95%, from 50% to 100%, from 70% to 80%, from 70% to 90%, from 70% to 95%, from 70% to 100%, from 80% to 90%, from 80% to 95%, from 80% to 100%, from 90% to 95%, from 90% to 100%, and from 95% to 100%).

In some embodiments, sugar modifications (*e.g.*, at the 2' position or 4' position) or replacement of the sugar at one or more ribonucleotides of the sequence may, as well as backbone modifications, include modification or replacement of the phosphodiester linkages. Specific examples of a sequence include, but are not limited to, sequences including modified backbones or no natural internucleoside linkages such as internucleoside modifications, including modification or replacement of the phosphodiester linkages. Sequences having modified backbones include, among others, those that do not have a phosphorus atom in the backbone. For the purposes of this application, and as sometimes referenced in the art, modified RNAs that do not have a phosphorus atom in their internucleoside backbone can also be considered to be oligonucleosides. In particular embodiments, a sequence will include ribonucleotides with a phosphorus atom in its internucleoside backbone.

Modified sequence backbones may include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates such as 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates such as 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included. In some embodiments, the sequence may be negatively or positively charged.

The modified nucleotides, which may be incorporated into the sequence, can be modified on the internucleoside linkage (*e.g.*, phosphate backbone). Herein, in the context of the polynucleotide backbone, the phrases "phosphate" and "phosphodiester" are used interchangeably. Backbone phosphate groups can be modified by replacing one or more of the oxygen atoms with a different substituent. Further, the modified nucleosides and nucleotides can include the wholesale replacement of an unmodified phosphate moiety with another internucleoside linkage as described herein. Examples of modified phosphate groups include, but

are not limited to, phosphorothioate, phosphoroselenates, boranophosphates, boranophosphate esters, hydrogen phosphonates, phosphoramidates, phosphorodiamidates, alkyl or aryl phosphonates, and phosphotriesters. Phosphorodithioates have both non-linking oxygens replaced by sulfur. The phosphate linker can also be modified by the replacement of a linking oxygen with nitrogen (bridged phosphoramidates), sulfur (bridged phosphorothioates), and carbon (bridged methylene-phosphonates).

The α -thio substituted phosphate moiety is provided to confer stability to RNA and DNA polymers through the unnatural phosphorothioate backbone linkages. Phosphorothioate DNA and RNA have increased nuclease resistance and subsequently a longer half-life in a cellular environment.

In specific embodiments, a modified nucleoside includes an alpha-thio-nucleoside (*e.g.*, 5'-*O*-(1-thiophosphate)-adenosine, 5'-*O*-(1-thiophosphate)-cytidine (a-thio-cytidine), 5'-*O*-(1-thiophosphate)-guanosine, 5'-*O*-(1-thiophosphate)-uridine, or 5'-*O*-(1-thiophosphate)-pseudouridine).

Other internucleoside linkages that may be employed according to the present invention, including internucleoside linkages which do not contain a phosphorous atom, are described herein.

In some embodiments, the sequence may include one or more cytotoxic nucleosides. For example, cytotoxic nucleosides may be incorporated into sequence, such as bifunctional modification. Cytotoxic nucleoside may include, but are not limited to, adenosine arabinoside, 5-azacytidine, 4'-thio-aracytidine, cyclopentenylcytosine, cladribine, clofarabine, cytarabine, cytosine arabinoside, 1-(2-C-cyano-2-deoxy-beta-D-arabino-pentofuranosyl)-cytosine, decitabine, 5-fluorouracil, fludarabine, floxuridine, gemcitabine, a combination of tegafur and uracil, tegafur ((RS)-5-fluoro-1-(tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione), troxacitabine, tezacitabine, 2'-deoxy-2'-methylidenecytidine (DMDC), and 6-mercaptopurine. Additional examples include fludarabine phosphate, N4-behenoyl-1-beta-D-arabinofuranosylcytosine, N4-octadecyl-1-beta-D-arabinofuranosylcytosine, N4-palmitoyl-1-(2-C-cyano-2-deoxy-beta-D-arabino-pentofuranosyl) cytosine, and P-4055 (cytarabine 5'-elaidic acid ester).

In some embodiments, the sequence includes one or more post-transcriptional modifications (*e.g.*, capping, cleavage, polyadenylation, splicing, poly-A sequence, methylation, acylation, phosphorylation, methylation of lysine and arginine residues, acetylation, and nitrosylation of thiol groups and tyrosine residues, etc.). The one or more post-transcriptional modifications can be any post-transcriptional modification, such as any of the more than one

hundred different nucleoside modifications that have been identified in RNA (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197) In some embodiments, the first isolated nucleic acid comprises messenger RNA (mRNA). In some embodiments, the mRNA comprises at least one nucleoside selected

5 from the group consisting of pyridin-4-one ribonucleoside, 5-aza-uridine, 2-thio-5-aza-uridine, 2-thiouridine, 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxyuridine, 3-methyluridine, 5-carboxymethyl-uridine, 1-carboxymethyl-pseudouridine, 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyluridine, 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine, 1-taurinomethyl-4-thio-uridine, 5-methyl-uridine, 1-methyl-pseudouridine, 4-thio-1-

10 methyl-pseudouridine, 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine, dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxyuridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, and 4-methoxy-2-thio-pseudouridine. In some embodiments, the mRNA comprises at least one nucleoside selected from the group consisting of 5-aza-cytidine,

15 pseudoisocytidine, 3-methyl-cytidine, N4-acetylcytidine, 5-formylcytidine, N4-methylcytidine, 5-hydroxymethylcytidine, 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine, 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-

20 thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, and 4-methoxy-1-methyl-pseudoisocytidine. In some embodiments, the mRNA comprises at least one nucleoside selected from the group consisting of 2-aminopurine, 2, 6-diaminopurine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-aminopurine, 7-deaza-8-aza-2-aminopurine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-

25 methyladenosine, N6-methyladenosine, N6-isopentenyladenosine, N6-(cis-hydroxyisopentenyl)adenosine, 2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine, N6-glycincylcarbamoyladenosine, N6-threonylcarbamoyladenosine, 2-methylthio-N6-threonyl carbamoyladenosine, N6,N6-dimethyladenosine, 7-methyladenine, 2-methylthio-adenine, and 2-methoxy-adenine. In some embodiments, mRNA comprises at least one nucleoside selected

30 from the group consisting of inosine, 1-methyl-inosine, wyosine, wybutosine, 7-deaza-guanosine, 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine, 6-thio-7-methyl-guanosine, 7-methylinosine, 6-methoxy-guanosine, 1-methylguanosine, N2-methylguanosine, N2,N2-dimethylguanosine, 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-

guanosine, and N2,N2-dimethyl-6-thio-guanosine.

The sequence may or may not be uniformly modified along the entire length of the molecule. For example, one or more or all types of nucleotides (*e.g.*, naturally-occurring nucleotides, purine or pyrimidine, or any one or more or all of A, G, U, C, I, pU) may or may not be uniformly modified in the sequence, or in a given predetermined sequence region thereof. In some embodiments, the sequence includes a pseudouridine. In some embodiments, the sequence includes an inosine, which may aid in the immune system characterizing the sequence as endogenous versus viral RNAs. The incorporation of inosine may also mediate improved RNA stability/reduced degradation. See for example, Yu, Z. et al. (2015) RNA editing by ADAR1 marks dsRNA as “self”. Cell Res. 25, 1283-1284, which is incorporated by reference in its entirety.

In some embodiments, any RNA sequence described herein, such as an editing template RNA, may comprise an end modification (*e.g.*, a 5' end modification or a 3' end modification). In some embodiments, the end modification is a chemical modification. In some embodiments, the end modification is a structural modification. See disclosures herein.

When a gene editing system disclosed herein comprises nucleic acids encoding the CRISPR nuclease and/or the RT polypeptide, *e.g.*, mRNA molecules, such nucleic acid molecules may contain any of the modifications disclosed herein, where applicable.

C. Exemplary Gene Editing Systems

Exemplary gene editing systems described herein, meant to be illustrative only, may comprise:

(a) a fusion polypeptide or a nucleic acid encoding such, wherein the fusion polypeptide comprises any of the CRISPR nuclease polypeptides, any of the RT polypeptides, and optionally one or more NLSs, which may be located at the N-terminus and/or the C-terminus, one or more peptide linkers, or a combination thereof; and

(b) a single RNA molecule, comprising a guide RNA, an RT donor RNA, and optionally one or more nucleotide linkers, one or more 5' end or 3' end protection elements, or a combination thereof.

In some embodiments, the CRISPR nuclease polypeptide in the fusion polypeptide may be a nickase variant (*e.g.*, those provided in **Table 5** below, such as the nickase variant containing the H845 mutation, *e.g.*, H845A). Alternatively or in addition, the CRISPR nuclease polypeptide in the fusion polypeptide may comprise one or more arginine and/or lysine substitutions, for example, at the positions disclosed herein (*e.g.*, K736, L784, Q812, N813,

I857, and/or A919 in SEQ ID NO: 1). In some specific examples, the CRISPR nuclease polypeptide in the fusion polypeptide may comprise a combination of arginine and/or lysine substitutions at positions provided herein, *e.g.*, I857, L784, and K736. Alternatively or in addition, the CRISPR nuclease polypeptide in the fusion polypeptide may comprise one or more mutations for reducing PAM recognition stringency, for example, at position D61, A68, H494, L1117, D1144, S1145, G1227, E1228, S1327, A1332, R1343, R1345, and/or T1347 of SEQ ID NO: 1. In some specific examples, such mutations may comprise: (i) one or more arginine and/or lysine substitutions, optionally arginine substitutions, at position D61, A68, H494, L1117, G1227, S1327, A1332, and/or T1347 of SEQ ID NO: 1; (ii) one or more amino acid substitutions at position D1144, S1145, E1228, R1343, and/or R1345, of SEQ ID NO: 1; or (iii) a combination of (i) and (ii).

In some embodiments, the RT polypeptide in the fusion polypeptide may be an MMLV variant, for example, SEQ ID NO: 53 provided in **Example 5** below.

In some examples, the fusion polypeptide provided herein may comprise an N-terminal CRISPR nuclease polypeptide at the N-terminus and a C-terminal RT polypeptide. Optionally, the fusion polypeptide may comprise a peptide linker (*e.g.*, a G/S rich linker or an XTEN peptide linker) between the CRISPR nuclease polypeptide and the RT polypeptide. Alternatively or in addition, the fusion polypeptide may comprise an NLS at the N-terminus and/or the C-terminus. In some examples, the fusion polypeptide may comprise two different NLSs, one at the N-terminus and the other one at the C-terminus.

In some examples, the fusion polypeptide provided herein may have any of the configurations disclosed herein, for example, those set forth in **Table 16** below (*e.g.*, Config4, Config5, Config7, Config9, or Config10), except that the FLAG motif is removed.

The single RNA molecule contained in the exemplary gene editing system may comprise the guide RNA and the RT donor RNA in any orientation. Optionally, the single RNA molecule may contain one or more nucleotide linkers between the gRNA and the RT donor RNA, and/or between the functional domains in the gRNA (*e.g.*, between the spacer and the scaffold sequences) and/or in the RT donor RNA (*e.g.*, between the PBS and the RTT sequences). In some examples, the single RNA molecule may further comprise a protection fragment (*e.g.*, those disclosed herein) at the 5' and/or 3' end.

In specific examples, the single RNA molecule contained in the exemplary gene editing system may comprise, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT, a PBS, and a 3' extension, which may have a pseudoknot motif.

In some examples, an exemplary gene editing system provided herein may comprise any

of the CRISPR-RT fusion polypeptides provided in **Table 8** or **Table 17** below, or a variant thereof with the FLAG motif removed, and a single RNA molecule provided herein (e.g., comprising, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT, a PBS, and a 3' extension, which may have a pseudoknot motif).

5 In other examples, an exemplary gene editing system provided herein may comprise any of the CRISPR-RT fusion polypeptides provided in **Table 8** or **Table 17** below, or a variant thereof with the FLAG motif removed, and a nucleic acid (e.g., a vector such as a viral vector) coding for a single RNA molecule provided herein (e.g., comprising, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT, a PBS, and a 3' extension, which may have a
10 pseudoknot motif).

In yet other examples, an exemplary gene editing system provided herein may comprise a nucleic acid encoding any of the CRISPR-RT fusion polypeptides provided in **Table 8** below, and a single RNA molecule provided herein (e.g., comprising, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT, a PBS, and a 3' extension, which may have a pseudoknot motif).

15 The nucleic acid may comprise an encoding nucleotide sequence that is codon optimized, for example, those provided in **Table 8** or **Table 17**, or a variant thereof with the FLAG motif removed.

In still other examples, an exemplary gene editing system provided herein may comprise a nucleic acid encoding any of the CRISPR-RT fusion polypeptides provided in **Table 8** or
20 **Table 17** below, or a variant thereof with the FLAG motif removed, and a nucleic acid (e.g., a vector such as a viral vector) encoding a single RNA molecule provided herein (e.g., comprising, from 5' to 3', a spacer sequence, a scaffold sequence, an RTT, a PBS, and a 3' extension, which may have a pseudoknot motif). The nucleic acid encoding the fusion polypeptide may be codon optimized, for example, those provided in **Table 8** or **Table 17**, or a variant thereof with the
25 FLAG motif removed. In some instances, the gene editing system may comprise two vectors, one encoding the fusion polypeptide and the other one encoding the single RNA molecule. Alternatively, the gene editing system may comprise one vector encoding both the fusion polypeptide and the single RNA molecule.

30 **III. Genetic Editing Methods**

Any of the gene editing systems can be used to genetically modify (edit) a target nucleic acid, which can be a genetic site of interest, e.g., a genetic site where genetic editing is needed, for example, to fix a genetic mutation, to introduce a protective mutation, to introduce modifications for modulating expression of a gene, etc.

A. Delivery of Gene Editing System to Target Cells

Components of any of the gene editing systems disclosed herein may be formulated, for example, including a carrier, such as a carrier and/or a polymeric carrier, e.g., a liposome, and
5 delivered by known methods to a cell (e.g., a mammalian cell). Such methods include, but not limited to, transfection (e.g., lipid-mediated, cationic polymers, calcium phosphate, dendrimers); electroporation or other methods of membrane disruption (e.g., nucleofection), viral delivery (e.g., lentivirus, retrovirus, adenovirus, adeno-associated virus (AAV)), microinjection, microprojectile bombardment ("gene gun"), fugene, direct sonic loading, cell squeezing, optical
10 transfection, protoplast fusion, impalefection, magnetofection, exosome-mediated transfer, lipid nanoparticle-mediated transfer, and any combination thereof. In some examples, the delivery method involves the use of lipid nanoparticles to mediate delivery of one or more components of the gene editing system disclosed herein.

In some embodiments, the method comprises delivering one or more nucleic acids (e.g.,
15 nucleic acids encoding the CRISPR nuclease polypeptide, the RT polypeptide, or the fusion polypeptide comprising both, the RNA guide, the RT donor RNA, or the single RNA molecule comprising both, etc.), one or more transcripts thereof, and/or a pre-formed RNA guide/CRISPR nuclease polypeptide/RT polypeptide complex to a cell, where a ternary complex is formed. In some embodiments, an RNA guide and/or RT donor RNA, or a fusion thereof, and an RNA
20 encoding a CRISPR nuclease polypeptide or a RT polypeptide, or a fusion polypeptide comprising both, are delivered together in a single composition. In some embodiments, an RNA guide and an RNA encoding a CRISPR nuclease polypeptide are delivered in separate compositions. In some embodiments, an RNA guide/RT donor RNA and an RNA encoding a CRISPR nuclease polypeptide/RT polypeptide delivered in separate compositions are delivered
25 using the same delivery technology. In some embodiments, an RNA guide/RT donor RNA and an RNA encoding a CRISPR nuclease polypeptide/RT polypeptide delivered in separate compositions are delivered using different delivery technologies.

In some embodiments, one or more of the protein components and one or more of the RNA components are delivered together. For example, the CRISPR nuclease and/or RT
30 polypeptide and the RNA guide and/or RT donor RNA are packaged together in a single AAV particle. In another example, the CRISPR nuclease and/or RT polypeptide and the RNA guide and/or RT donor RNA are delivered together via lipid nanoparticles (LNPs). In some embodiments, the CRISPR nuclease and/or RT polypeptides and the RNA guide and/or RT donor RNA are delivered separately. For example, the CRISPR nuclease and/or RT polypeptides

and the RNA guide and/or RT donor RNA are packaged into separate AAV particles. In another example, the CRISPR nuclease and/or RT polypeptides is delivered by a first delivery mechanism and the RNA guide and/or RT donor RNA is delivered by a second delivery mechanism.

- 5 Exemplary intracellular delivery methods, include, but are not limited to: viruses, such as AAV, or virus-like agents; chemical-based transfection methods, such as those using calcium phosphate, dendrimers, liposomes, or cationic polymers (*e.g.*, DEAE-dextran or polyethylenimine); non-chemical methods, such as microinjection, electroporation, cell squeezing, sonoporation, optical transfection, impalefection, protoplast fusion, bacterial
- 10 conjugation, delivery of plasmids or transposons; particle-based methods, such as using a gene gun, magnetofection or magnet assisted transfection, particle bombardment; and hybrid methods, such as nucleofection. In some embodiments, a lipid nanoparticle comprises an mRNA encoding a CRISPR nuclease-RT fusion polypeptide, an editing template RNA, or an mRNA encoding such. In some embodiments, the present application further provides cells produced by
- 15 such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells.

B. Genetically Modified Cells

- Any of the gene editing systems disclosed herein can be delivered to a variety of cells
- 20 (*e.g.*, to mammalian cells such as a mouse cell, a non-human primate cell, or a human cell). In some embodiments, the cell is in cell culture or a co-culture of two or more cell types. In some embodiments, the cell is *ex vivo*. In some embodiments, the cell is obtained from a living organism and maintained in a cell culture.

- In some embodiments, the cell is derived from a cell line. A wide variety of cell lines for
- 25 tissue culture are known in the art. Examples of cell lines include, but are not limited to, 293T, MF7, K562, HeLa, CHO, and transgenic varieties thereof. 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) (Manassas, Va.)). In some embodiments, the cell is an immortal or immortalized cell. In some embodiments, the cell is a primary cell. In some embodiments, the cell is a stem cell
- 30 such as a totipotent stem cell (*e.g.*, omnipotent), a pluripotent stem cell, a multipotent stem cell, an oligopotent stem cell, or an unipotent stem cell. In some embodiments, the cell is an induced pluripotent stem cell (iPSC) or derived from an iPSC. In some embodiments, the cell is a differentiated cell. In some embodiments, the cell is a mammalian cell, *e.g.*, a human cell or a murine cell. In some embodiments, the murine cell is derived from a wild-type mouse, an

immunosuppressed mouse, or a disease-specific mouse model. In some embodiments, the cell is a cell within a living tissue, organ, or organism.

Any of the genetically modified cells produced using any of the gene editing system disclosed herein is also within the scope of the present disclosure. Such modified cells may
5 comprise a disrupted target gene.

Any of the gene editing systems, compositions comprising such, vectors, nucleic acids, RNA guides and cells disclosed herein may be used in therapy. Gene editing systems, compositions, vectors, nucleic acids, RNA guides and cells disclosed herein may be used in methods of treating a disease or condition in a subject. Any suitable delivery or administration
10 method known in the art may be used to deliver compositions, vectors, nucleic acids, RNA guides and cells disclosed herein. Such methods may involve contacting a target sequence with a composition, vector, nucleic acid, or RNA guide disclosed herein. Such methods may involve a method of editing a target sequence as disclosed herein. In some embodiments, a cell engineered using an RNA guide disclosed herein is used for *ex vivo* gene therapy.

15

IV. Therapeutic Applications

Any of the gene editing systems or modified cells generated using such a gene editing system as disclosed herein may be used for treating a disease that is associated with the target gene, for example, a genetic defect in the target gene.

20 In some embodiments, provided herein is a method for treating a target disease as disclosed herein comprising administering to a subject (*e.g.*, a human patient) in need of the treatment any of the gene editing systems disclosed herein. The gene editing system may be delivered to a specific tissue or specific type of cells where the gene edit is needed. The gene editing system may comprise LNPs encompassing one or more of the components, one or more
25 vectors (*e.g.*, viral vectors) encoding one or more of the components, or a combination thereof. Components of the gene editing system may be formulated to form a pharmaceutical composition, which may further comprise one or more pharmaceutically acceptable carriers.

In some embodiments, modified cells produced using any of the gene editing systems disclosed herein may be administered to a subject (*e.g.*, a human patient) in need of the
30 treatment. The modified cells may comprise a substitution, insertion, and/or deletion described herein. In some examples, the modified cells may include a cell line modified by the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide, and the RNA guide and RT donor RNA or the single RNA molecule comprising both. In some instances, the modified cells may be a heterogenous population comprising cells with different

types of gene edits. Alternatively, the modified cells may comprise a substantially homogenous cell population (*e.g.*, at least 80% of the cells in the whole population) comprising one particular gene edit in the target gene. In some examples, the cells can be suspended in a suitable media.

In some embodiments, provided herein is a composition comprising the gene editing
5 system or components thereof. Such a composition can be a pharmaceutical composition. A pharmaceutical composition that is useful may be prepared, packaged, or sold in a formulation suitable for oral, rectal, vaginal, parenteral, topical, pulmonary, intranasal, intra-lesional, buccal, ophthalmic, intravenous, intra-organ or another route of administration. A pharmaceutical composition of the disclosure may be prepared, packaged, or sold in bulk, as a single unit dose,
10 or as a plurality of single unit doses. As used herein, a “unit dose” is discrete amount of the pharmaceutical composition (*e.g.*, the gene editing system or components thereof), which would be administered to a subject or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage.

A formulation of a pharmaceutical composition suitable for parenteral administration
15 may comprise the active agent (*e.g.*, the gene editing system or components thereof or the modified cells) combined with a pharmaceutically acceptable carrier, such as sterile water or sterile isotonic saline. Such a formulation may be prepared, packaged, or sold in a form suitable for bolus administration or for continuous administration. Some injectable formulations may be prepared, packaged, or sold in unit dosage form, such as in ampules or in multi-dose containers
20 containing a preservative. Some formulations for parenteral administration include, but are not limited to, suspensions, solutions, emulsions in oily or aqueous vehicles, pastes, and implantable sustained-release or biodegradable formulations. Some formulations may further comprise one or more additional ingredients including, but not limited to, suspending, stabilizing, or dispersing agents.

25 The pharmaceutical composition may be in the form of a sterile injectable aqueous or oily suspension or solution. This suspension or solution may be formulated according to the known art, and may comprise, in addition to the cells, additional ingredients such as the dispersing agents, wetting agents, or suspending agents described herein. Such sterile injectable formulation may be prepared using a non-toxic parenterally-acceptable diluent or solvent, such
30 as water or saline. Other acceptable diluents and solvents include, but are not limited to, Ringer’s solution, isotonic sodium chloride solution, and fixed oils such as synthetic mono- or di-glycerides. Other parentally-administrable formulations which that are useful include those which may comprise the cells in a packaged form, in a liposomal preparation, or as a component of a biodegradable polymer system. Some compositions for sustained release or implantation

may comprise pharmaceutically acceptable polymeric or hydrophobic materials such as an emulsion, an ion exchange resin, a sparingly soluble polymer, or a sparingly soluble salt.

V. Kits and Uses Thereof

5 The present disclosure also provides kits that can be used, for example, to carry out a method described herein for genetical modification of a target gene. In some embodiments, the kits include an RNA guide and an RT donor RNA, or a single RNA molecule comprising both, a CRISPR nuclease polypeptide, and an RT polypeptide, or a fusion polypeptide thereof. In some
10 embodiments, the kits include the single RNA molecule and the CRISPR nuclease-RT fusion polypeptide. In some embodiments, the kits include a polynucleotide that encodes the CRISPR nuclease polypeptide, the RT polypeptide, or the CRISPR nuclease-RT fusion polypeptide, and optionally the polynucleotide is comprised within a vector, *e.g.*, as described herein. In some
15 embodiments, the kits include a polynucleotide that encodes the RNA components disclosed herein. The CRISPR nuclease polypeptide, the RT polypeptide, or a fusion polypeptide thereof (or polynucleotide encoding such) and the RNA components (*e.g.*, as a ribonucleoprotein) can be
packaged within the same or other vessel within a kit or can be packaged in separate vials or other vessels, the contents of which can be mixed prior to use.

 The CRISPR nuclease polypeptide, the RT polypeptide, and the RNA components can be packaged within the same or other vessel within a kit or can be packaged in separate vials or
20 other vessels, the contents of which can be mixed prior to use. The kits can additionally include, optionally, a buffer and/or instructions for use of the RNA components, the CRISPR nuclease polypeptide, and the RT polypeptide, or the fusion polypeptide thereof.

General techniques

25 The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as *Molecular Cloning: A Laboratory Manual*, second edition (Sambrook, et al., 1989) Cold Spring Harbor Press; *Oligonucleotide*
30 *Synthesis* (M. J. Gait, ed. 1984); *Methods in Molecular Biology*, Humana Press; *Cell Biology: A Laboratory Notebook* (J. E. Cellis, ed., 1989) Academic Press; *Animal Cell Culture* (R. I. Freshney, ed. 1987); *Introduction to Cell and Tissue Culture* (J. P. Mather and P. E. Roberts, 1998) Plenum Press; *Cell and Tissue Culture: Laboratory Procedures* (A. Doyle, J. B. Griffiths, and D. G. Newell, eds. 1993-8) J. Wiley and Sons; *Methods in Enzymology* (Academic Press,
35 Inc.); *Handbook of Experimental Immunology* (D. M. Weir and C. C. Blackwell, eds.): Gene

- Transfer Vectors for Mammalian Cells (J. M. Miller and M. P. Calos, eds., 1987); Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds. 1987); PCR: The Polymerase Chain Reaction, (Mullis, et al., eds. 1994); Current Protocols in Immunology (J. E. Coligan et al., eds., 1991); Short Protocols in Molecular Biology (Wiley and Sons, 1999); Immunobiology (C. A. Janeway and P. Travers, 1997); Antibodies (P. Finch, 1997); Antibodies: a practice approach (D. Catty., ed., IRL Press, 1988-1989); Monoclonal antibodies: a practical approach (P. Shepherd and C. Dean, eds., Oxford University Press, 2000); Using antibodies: a laboratory manual (E. Harlow and D. Lane (Cold Spring Harbor Laboratory Press, 1999); The Antibodies (M. Zanetti and J. D. Capra, eds. Harwood Academic Publishers, 1995); *DNA Cloning: A practical Approach*, Volumes I and II (D.N. Glover ed. 1985); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds.(1985»; *Transcription and Translation* (B.D. Hames & S.J. Higgins, eds. (1984»; *Animal Cell Culture* (R.I. Freshney, ed. (1986»; *Immobilized Cells and Enzymes* (IRL Press, (1986»; and B. Perbal, *A practical Guide To Molecular Cloning* (1984); F.M. Ausubel *et al.* (eds.).
- Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present disclosure to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the present disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

20

EXAMPLES

- The following examples are provided to further illustrate some embodiments of the present disclosure but are not intended to limit the scope of the present disclosure; it will be understood by their exemplary nature that other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

25

Example 1: CRISPR Nuclease-Mediated Editing of Human Target Genes in HEK293T Cells

- This Example describes genomic editing of exemplary target genes, including the AAVS1, EMX1, and VEGFA genes, by the CRISPR nuclease of SEQ ID NO: 1 introduced into cells by lipid-based transient transfection into the HEK293T cell line.

30

- The CRISPR nuclease was tagged with an N-terminal SV40 nuclear localization sequence (NLS) and a C-terminal XTEN linker directly upstream of a nucleoplasmin NLS, and its coding sequence was converted to a human codon-optimized DNA sequence, synthesized, and cloned into a pcDNA3.1 vector (Invitrogen), containing a CMV promoter for expression.

35

The reference and NLS-tagged sequences used are in **Table 1**. Plasmids were purified using a midiprep kit.

Table 1. Sequences of CRISPR Nuclease Constructs

Component	Nucleotide Sequence (Codon optimized for human cells)	Amino acid sequence
Reference	atgaacaagccttacagcatcggcctggatatcgccaccaacacgcgtg ggatggtctatcatcacagatgactacaagtgccctgccaagaaaatg agagtgttaggcaacaccgacaaggagtacatcaagaagaacctgat cggcgccctcctggttcgacggcggaacacagccgccgatcggcgg ctgaaaaaggaccgccagacggagatatacgcgcgagaaatagaa tcctgtacttacaggaaatcttcgctgaggagatgagcaaatggacg actctttttccacagactggaagatagcttttggaggagagacaag cgaggtctaagtaccctatcttcgctacactgcaaggagaaaaggac taccacgagaagtftcccacatctaccacctgaggaaaagagctggcg gacaaacaagagaagggccacctgagactgatctaccttgccctggct catatcatcaagttcagaggccacttctgatcggagacgacagattcg acgtgagaatacagatatccagaagcagttaccaggcctttctggaga cttcgacaccacctttgaaaacaaccacctgctgagccaaaacgtgg atgtggaagccatcttaaccgataagatctccaagagcgccaagaaag atagaatcctcgcccagtatcctaaccagaaaagcaccggtatcttcgc cgaaattcctgaagctgatcgtgggcaaccaggccgacttcaaaaaaca cttaacctggaggacaaaacacctctgcagtttgccaaggactcctac gacgaagatctggaacacctgctggccagatcggcgacgagttcgc cgatctgttcagcgccgctaaagaactgtacgacagcgtgctgctgag cggcatcttgacagtgatcgtgtccacaaaggccccctgagcgc cagcatgatccagagatacagcagcacagagagatctgaagcag ctgaagcagttgtgaaggcctctctgcccagaaagttaccaggagattt ttcgggactccagcaagatggatagcccgatacattgaggggcaag accaaccaggaagctttctacaagtacgtgttaactctgaccaaac tgaagaattccgagaacttctggagaaaatcaagaacgaagatttcc tgagaaagcagaggaccttcgacaacgcgacatccccaccagggt ccacctgaccgagctgaaggctatcatcagacgacagagtgaagtacta ccccttctgaaggaaaatcaagacagaatcgagaagattctgacattc agaatcccatactatacggacctctggcccgtgaaaagagcgacttc gcctggatgaccagaaaagaccgatgattctatccggccctggaacttc gaggacctggtggacaaggaaaagtctgcccaggccctcatccaccg gatgaccaacaacgacttctacctgcctgaagagaaggctgctgcctaa acattccctgatctacgaaaagtacagtgatcaacgagctgaccaag gttagalacaagaacgagcagggcgagacgtacttcttcgacttaala tcaagcaggaaatcttcgatggagtgttaaggagcaccggaaggtct ctaagaagaaactgttagacttctgctaaggaaatagcaggaattccg gatcgtggacgtgatcggactggataaggagaacaaagcctttaacgc cagcctgggaacataccacgacctggagaagatcctcgacaaggact tcctggataaccccacaatgagtcctcctggaagatatcgtgcaga cgctgacactgttcgaggacagagatgatcaagaagagactggaa aactacaaggatctgttcaccgagagccagctgaagaagtgtataga cggcactacaccggatggggccgctgtccgctaactgatcaacgg catcagagataaggaaaagtcagaaaacctcctgactacttaacga cgatggcagaagcaaccggaacttcagcagctgatcaacgatgatg gcttgagcttcaagtctatcatcttaaggcccaggccggaagccaca gcgataacctgaagggaagtgggtgggtgagctgcccggctcctgcc	MNKPYSIGLDIGTNSVGWSIIT DDYKVPAAKKMRVLGNTDKEY IKKNLIGALLFDGGNTAADRR LKRTARRRYTRRRNRILYLQEI FAEEMSKVDDSFHRLSDSFLV EEDKRGSKYPIFATLQEEKDYH EKFPTIYHLRKEADKQEKAD LRLIYLALAHIIKFRGHFLIEDD RFDVRNTDIQKQYQAFLEIFDT TFENNHLSSQNVDVEAILTDKI SKSAKKDRILAQYPNQKSTGIF AEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLENLLG QIGDEFADLFSAAKKLYDSVLL SGILTVIDLSTKAPLSASMIQRY DEHREDLKQLKQFVKASLPEK YQEIFADSSKDG YAGYIEGKTN QEAIFYKLYSKLLTKQEDSENF LEKIKNEDFLRKQRTFDNGSIP HQVHLTELKAIIRRQSEYYPFL KENQDRIEKILTRIPYYIGPLA REKSDFAWMTRKTDDSRPWN FEDLVDKEKSAEAFIHRMTNN DFYLPPEEKVLPKHS LIYEKFTV YNELTKVRYKNEQGETYFFDS NIKQEIFDGVFKEHRKVS KKK LLDFLAKEYEFEFRIVDVIGLDK ENKAFNASLGTYHDLKILDK DFLDNPDNESILEDIVQTLTLFE DREMIKKRLNYKDLFTESQL KKLYRRHYTGWGRLSAKLING IRDKESQKTILDYLIDGGRSNR NFMQLINDDGLSFKSIISKAQA GSHSDNLKEVVGELAGSPAIAK KGILQSLKIVDELVKVMGYEP EQIVVEMARENQTTNQGRNRS RQRYKLDDQVKNLNSDLNG NILKEYPTDNQALQNERLFLY YLQNGRDMYTGEALDIDNLS QYDIDHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYD NLTKAERGGLTSDDKARFIQR QLVETRQITKHVARILDERFNN ELDSKGRRIRKVKIVTLKSNLV SNFRKEFGFYKIREVNNDYHHA

	atcaagaaggcatcctgcaaaagtctaaagattgtcgacgagctggg aagtgatggctatgaacctgagcagatcgtggtgaaatggccag agaaaaccagacaaccaatcaaggcagaagaacagcagacagcg gtacaagctgctggacgacggcgtgaagaatctggccagcgacctga acggcaacatcctgaaagagtacccaaccgacaaccaggccctgca gaacgagcgctgttctgtactacctgcagaacggcagagacatgta caccggcgaagccctggatacgcataacctgagccagtagatacga ccacataatccacaggcccttatcaaggatgattctattgataatcgggt gcttctctccagcgcgaagaacagagaaagagcgacgacgtgccc agcctcgaatcgtgaaagactgcaaggtgtctggaagaagctcctg gacgccaagctgatgagccagagaagtagcacaacctgaccaagg ccgagagaggagccctgaccagcgacgataaggctagattcaccaa aggcagctggtcgaaccagacagatcaccaagcacgtggccaggga tcctcgacgagcggttcaataacgagctggacagcaaaaggtagaaga attagaaagggttaagatcgtcaccctgaagtctaattgtgtccaacttc cggaaaggagttcgcttctacaagatcagagagggtgaacgactacca ccacgcccattgatgcctacctgaacggcgtgtgccaagccatcct gaccaataacccctgaactggagcccaggtcgtgtacggcctgtacag acggaagaagctgacaaaatcgtgcgggaagataaggagaaga gtacagcgaagccacaaggaaaatgtttctatagcaacctgatgaac atgtcaagagagtggtgagactggccgacggtatctatcgtgtcgg cctgtgatcgaaccggcagatacatgggcaaaaccgctgggacaa aaagaagcacttcgccaccgtgcggaaggtagctaccctcagg tgaacatagtgaaaaagacagagatccagaccggcggtcttctaaaag agtccatcctcgctcacggcaatagcgacaagctgatcccagaag acaaaagacatctacctggaccctaagaagtagcgcgcttcgacagc cccatcgtggcttacagcgtcctgtgtgcccgaatfaagaaggca aggctcagaagctgaagacggtgacagagctgctgggcataacaatc atggaacgcagcaaatcgaagaaccaagcgctttctggaagc aagggtacctgaacatccgggacgataagctgatgacctgctaag tacagcctgttcgaactggaatggcagacggcggtgctgaccag cgccggagagctgcaaaagggaatgagctggccctgacctaccag ttcatgaaatcctgtatctgacctctagatataacgagagcaagggcaa acctgaagagatcgaagaagaacaggagttgtgaaccaacacgtga gtacttcgacgacatcctgcagctgatcaatgacttcagcaagggt gatcctggccgacccaacctggaaaaatcaacaaatlatcagga caacaaggagaatctcctggacgaactggctaacaacatcattaat ctgttcacattaccagccttggcgctcctgcccctttaagtcttcgac aaaatcgtggaccgcaagcgttacaccagcacaaggaggtgcttaa cagcacactgatccaccagagcatcactgcctgtatgagacaagaat cgacctgtctaagctggcgaggat (SEQ ID NO: 5)	HDAYLNAVVAKAILTKYPQLE PEFVYGLYRRKKLSKIVREDK EEKYSEATRKMFYSLNLMNMF KRVVRLADGSIVVRPIETGRY MGKTAWDKKKHFAVTRKVL YPQVNIVKKTEIQTGGFSKESI LAHGNSDKLIPRKTCDIYLDPK KYGGFDSPIVAYSVLVADIKK GKAQKLKTVTELLGITIMERS KFEKNPSAFLESKGYLNIRDD KLMILPKYSLFELENGRRLLA SAGELQKGNELALPTQFMKFL YLASRYNESKKGPEEIEKKQEF VNQHVSYFDDILQLINDFSKRV ILADANLEKINKLYQDNKENIS VDELANNIINLFTFTSLGAPAA FKFFDKIVDRKRYTSTKEVLNS TLIHQSITGLYETRIDL SKLGED (SEQ ID NO: 1)
SV40 NLS	aaacggacagccgacggaagcaggttcgagtcaccaaaagaagaag cggaaaagtc (SEQ ID NO: 79)	KRTADGSEFESPKKKRKV (SE Q ID NO: 3)
Nucleoplas min NLS	aaaaggccggcgccacgaaaaaggccggcaggcaaaaagaa aaag (SEQ ID NO: 6)	KRPAATKKAGQAKKKK (SEQ ID NO: 4)
XTEN linker	tctggaggatctagcggaggatcctctggcagcgagacaccaggaa aagcagtcagcaaccagagagcagtgccggcagcagcgcg cagcagc (SEQ ID NO: 7)	SGGSSGGSSGSETPGTSESATP ESSGGSSGGSS (SEQ ID NO: 8)
NLS-tagged Reference	atgaaacggacagccgacggaagcaggttcgagtcaccaaaagaaga agcggaaagtcaacaagccttacagatcgccctggatcggcacc aacagcgtggatggtctatcatcacagatgactacaaggtgcctgcc aagaaaatgagagtgtaggcaacaccgacaaggagtagatcaaga aa	MKRTADGSEFESPKKKRKVNK PYSIGLDIGTNSVGWSITDDY KVPKKMRVLGNTDKEYIKK NLIGALLFDGGNTAADRLKR

gaacctgacggcgccctcctgttcgacggcggaacacagccgccc atcgccgctgaaaaggaccgccagacggagatatacggcgga gaaatagaatcctgtacttacaggaaatcttcgctgaggagatgagcaa agtggacgactctttttccacagactggaagatagctttttgtggaa aggacaagcgaggccttaagtaccctatcttcgctacactgcaagagg aaaaggactaccacgagaagttccaccatctaccactgaggaaa agctggcgggacaaaacagagaaggccgacctgagactgactacctt gcccgtgctatatacaagttcagaggccacttctgacgaggacg acagattcgacgtgagaatacagatatacagaagcagtaggagccct ttctggagatcttgacaccaccttgaacaacacacactgctgagcca aaacgtgagatggaagccatcttaaccgataagatccaagagcgc caagaaagatagaatcctcgccagatcttaaccagaaaagcaccg glatcttcgccgaattcctgaagctgacgtgggcaaccagccgactt caaaaaacactttaacctggaggacaaaacacactctgcagtttgcaa ggactcctacgacgaagatctggaacacactgctggccagatcgccg acgagttcgccgactctgttcagcgccgctaagaaactgtacgacagcg tgctgctgagcggcatcttgacagtgtacgtgctccacaagggcccc cctgagcgccagcatgatccagagatacagcagcagacagagaggat ctgaagcagctgaagcagttgtgaaggcctctctgcccagagaagtac caggagatctttgcgagcctcagcaaggatgatacggcgatacatt gaggggcaagaccaaccaggaagccttctacaagtacgtgtctaaactg ctgaccaaacaggaagattccgagaactctcgagaaaatcaagaa cgaagatttctgagaagcagagacacttcgacaacggcagcatcc cccaccaggtccactgaccgagctgaaggctatcatcagacacag agtgaagtactaccccttctgaaggaaaatcaagacagaatcgagaag attctgacattcagaatccatactatctcgacactctggcccgtaaaa gagcgaacttcgctggaagaccagaagaccgatgattctatccggcc ctggaacttcgaggacctgtgtgacaagaaaagtctgcccaggccct tcattccaccggatgaccaacaacgacttctacctgacctgaagagaagg tgctgcctaaacattcctgatctacgaaaagtacagtgataacga gctgaccaaggttagatacaagaacgagcaggccgagacgtactctt cgactctaataatcaagcaggaatcttcgatggagtgttcaaggagcac cggaaggctctctaagaagaactgttagacttctgctaagggaatac aggaaattccgacgtgacgtgacgtgactgggataaggagaacaaa gcctttaacgccagcctgggaacataccacgacctgagaagatcctc gacaaggacttctggataacccgacaatgagtcacatcctggaagat atcgtgcagacgctgacactgttcgaggacagagagatgatcaagaa gagactggaaaactacaagatctgttcaccgagagccagctgaaga agttgtatagacggcactacaccggatggggccgctgtccgctaaac tgatcaacggcatcagagataaaggaaagtcagaaaacacatcctggact actaatcagcagatggcagaagcaaccggaaactcatgcagctgatca acgatgatggcttgagctcaagttatcatctctaagcccaggccgg aagccacagcgataacctgaaggagtggtggtgagctggccggc tctcctgccatcaagaaggccatcctgcaagtgctaaagattgtcagc agctgggtgaagggtgatggcctatgaacctgagcagatcgtgtggaa atggccagagaaaaccagacaaccaatcaaggcagaagaacagca gacagcgggtacaagctgctggacgacggcgtgaagaatctggccag cgacctgaacggcaacatcctgaaagagtaccacacgacaaccag gcctgcagaacgagcgctgttctgtactacctgcagaacggcaga gacatgtacaccggcgaagccctggatatacgaacctgagccagtag gatatcgaccacataatcccacaggcccttataaggtgattctatgat aatcggggtgctgtctccagcggaagaacagaggaaaagagcgacg acgtgcccagcctgaaatcgtgaagactgcaaggtgttctggaaaga agctcctgacgccaagctgatgagccagagaagtacgacaacctg accaaggccgagagaggagcctgaccagcgacgataaggctagat	TARRRYTRRRNRILYLQEIFAE EMSKVDDSPFHRLEDSFLVEE DKRGSKYPIFATLQEEKDYHE KFPTIYHLRKELADKQEKADL RLIYLALAHIIKFRGHFLIEDDR FDVRNTDIQKQYQAFLEIFDIT FENNHLSSQNVDEAILTDKIS KSAKKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFNLED KTPLQFAKDSYDEDLNLLGQI GDEFADLFSAAKKLYDSVLLS GILTVIDLSTKAPLSASMIQRY DEHREDLKQLKQFVKASLPEK YQEIFADSSKDGAGYIEGKTN QEAFYKYLKLLTKQEDSENF LEKIKNEDFLRKQRTFDNGSIP HQVHLTELKAIIRRQSEYYPFL KENQDRIEKILTRIPYIYGLA REKSDFAWMTRKTDDSIKRPWN FEDLVDKEKSAEAFIHRMTNN DFYLPPEEKVLPKHSIYEKFTV YNELTKVRYKNEQGETYFFDS NIKQEIFDGVFKEHRKVSXXX LLDFLAKEYEFRIVDVIGLDK ENKAFNASLGTYHDLKILDK DFLDNPDNSESILEDIVQTLTLFE DREMIKKRLNYKDLFTESQL KKLYRRHYTGWRSLAKLING IRDKESQKTILDYLIIDGRSNR NFMQLINDDGLSFKSIISKAQA GSHSDNLKEVVGELAGSPAIF KGILQSLKIVDELVKVMGYEP EQIVVEMARENQTTNQGRNRS RQRYKLDDGVKNLASDLNG NILKEYPTDNQALQNERLFLY YLQNGRDMYTGEALDIDNLS QYDIDHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYD NLTKAERGGTSDDKARFIQR QLVETRQITKHVARILDERFNN ELDSKGRIRKVKIVTLKSNLV SNFRKEFGFYKIREVNDYHHA HDAYLNAVVAKAILTKYPQLE PEFVYGLYRRKKLSKIVREDK EEKYSEATRKMFFYSNLMNMF KRVRRLADGSIVVRPIETGRY MGKTAWDKKKHFATVRKVLS YPQVNIKKTEIQTGGFSKESI LAHGNSDKLIPRKTCDIYLDPK KYGGFDSPIVAYSVLVADIKK GKAQKLKTVTELLGITIMERS KFEKNPSAFLESKGYLNIRDD KLMILPKYSLFELENGRRLLA SAGELQKGNELALPTQFMKFL
---	--

tcaccaaaggcagctggtcgaaccagacagatcaccaagcacgtg gccagatcctcgacgagcgttcaataacgagctggacagcaaag gtagaagaattagaaaggtaagatcgtcacctgaagtctaattgtg tccaactfccggaaggagttcgcttctacaagatcagagaggtgaac gactaccaccacgcccattgatgcctacctaagccggtgtggcaa agccatcctgaccaataacctcaactggagcccagttcgtgtacgg cctgtacagacgggaagcgtgagcaaatcgtgcgggaagataag gaagagaagtacagcgaagccacaaggaaaatgttttctatagaac ctgatgaacatgttcaagagagtgtgagactggccgacggatctatc gtcgtgcggcctgtgtatcgaacccggcagatacatgggcaaacgc ctgggacaaaaagaagcacttcgccaccgtgcggaaggtactgagct accctcaggtgaacatagtgaaaaagacagagatccagaccggcgg cttctctaaagagtcacctcctcgtcacggcaatagcacaagctgac ccagaaaagacaaaagacatctacctggaccctaagaagtatggcgg cttcgacagccccatcgtggttacagctcctggtgtggccgatatta agaaggggcaaggctcagaagctgaagacgggtgacagagctgtggg cataacaatcatggaacgcagcaaatcgaagaaccaagcgccttt cttggaagcaagggtacctgaacatccgggacgataagctgatgat cctgcctaagtacagcctgttcgaactggaaaatggcagacggcggct gtagccagcgcggagagctgcaaaagggaatgagctggccctg cctaccagttcatgaaattcctgtatctggccctatagataacgagag caaggcgaacctgaagagatcgaagaagaacaggagtttgaacc aacacgtgagctacttcgacgacatcctgcagctgatcaatgacttcag caagagggtgatcctggccgacgccaacctggaaaaatcaacaac tatacaggacaacaaggagaatatctccgtggacgaactggctaaca acatcattaatctgttcacattcaccagccttggcgctcctgcgccttta agttcttcgacaaaatcgtggaccgaagcggtagaccagcaccaag gaggtgcttaacagcacactgatccaccagagcatcactggcctgtat gagacaagaatcgacctgtctaagctggcgaggatctggaggatct agcggaggatcctctggcagcagacaccaggaacaagcagtcag caacaccagagagcagtggtggcgagcagcggcgagcagcaaaa aggccggcgccacgaaaaaggccggcaggcaaaaagaaaaa g (SEQ ID NO: 9)	YLASRYNESKGKPEEIEKKQEF VNQHVSFYDDILQLINDFSKR VILADANLEKINKLYQDNKENIS VDELANNIINLFTFTSLGAPAA FKFFDKIVDRKRYTSTKEVLNS TLIHQSITGLYETRIDLKSLGED SGGSSGGSSGSETPGTSESATP ESSGGSSGGSSKRPAATKKAG QAKKKK (SEQ ID NO: 10)
---	--

RNA guides were designed and cloned into a pUC19 plasmid following the U6 PolIII promoter and terminated with a 6x polyT sequence. RNA guides were designed to be specific to target sequences within the coding exons of AAVS1, EMX1, and VEGFA with 5'-NGG-3' PAM sequences (the PAM sequence is on the 3' end of the target sequence). The U6 PolIII promoter uses a +1 G at the start of the transcript (i.e., the 5' end of the RNA) for more efficient transcription that is excluded from the sequences described here. See all RNA guide sequences in Table 2. Plasmids were purified using a midiprep kit.

Table 2. Target and RNA Guide Sequences

Guide	PAM	Target	Spacer	gRNA Guide Sequence

AAVS1-T3	AGG	TTGTGAGAA TGGTGCCTC (SEQ ID NO: 11)	UUGUGAGAAU GGUGCGUCCU (SEQ ID NO: 12)	UUGUGAGAAUGGUGCGUCCU guuuu agagcugugcugaaaaagcacagcacguuuuuuuuag gcagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 13)
AAVS1-T6	TGG	CTTTGTAGC CTCTCCCGC TC (SEQ ID NO: 14)	CUUUGUAGCC UCUCCCGCUC (SEQ ID NO: 15)	CUUUGUAGCCUCUCCCGCUC guuuuu gagcugugcugaaaaagcacagcacguuuuuuuuag cagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 16)
EMX1-T2	AGG	TGGATGGCG ACTTCAGGC AC (SEQ ID NO: 17)	UGGAUGGCGA CUUCAGGCAC (SEQ ID NO: 18)	UGGAUGGCGACUUCAGGCAC guuuu agagcugugcugaaaaagcacagcacguuuuuuuuag gcagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 19)
EMX1-T7	GGG	CTTTAGCAA GGGACTATT CA (SEQ ID NO: 20)	CUUUAGCAAG GGACUUAUCA (SEQ ID NO: 21)	CUUUAGCAAGGGACUUAUCA guuuu agagcugugcugaaaaagcacagcacguuuuuuuuag gcagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 22)
VEGFA-T3	TGG	TTGTTGACA TTGTCCACA CC (SEQ ID NO: 23)	UUGUUGACAU UGUCCACACC (SEQ ID NO: 24)	UUGUUGACAUUGUCCACACC guuuu agagcugugcugaaaaagcacagcacguuuuuuuuag gcagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 25)
VEGFA-T6	AGG	TTTGCCCAT ACTGGGGAC CA (SEQ ID NO: 26)	UUUGCCCAUA CUGGGGACCA (SEQ ID NO: 27)	UUUGCCCAUACUGGGGACCA guuuu agagcugugcugaaaaagcacagcacguuuuuuuuag gcagugauugaaaaaauccaguccguuuuacagcuuga aaaagugagcaccgaaucggugcuu (SEQ ID NO: 28)

* Spacer in upper case and scaffold (SEQ ID NO: 2) in lower case

Approximately 16 hours prior to transfection, 25,000 HEK293T cells in DMEM/10%FBS+Pen/Strep (D10 media) were plated into each well of a 96-well plate. On the day of transfection, the cells were 70-90% confluent. For each well to be transfected, a mixture of Lipofectamine 2000™ (ThermoFisher Scientific) and Opti-MEM™ (ThermoFisher Scientific) was prepared and incubated at room temperature for 5 minutes (Solution 1). After incubation, the Lipofectamine 2000™:Opti-MEM™ mixture was added to a separate mixture containing the CRISPR nuclease plasmid (NLS-tagged), RNA guide plasmid, and Opti-MEM™ (Solution 2). In the case of negative controls, the CRISPR nuclease plasmid was excluded. Solutions 1 and 2 were mixed by pipetting up and down, then incubated at room temperature for 25 minutes. Following incubation, the Solution 1 and 2 mixture was added dropwise to each well of a 96-well plate containing the cells. Approximately 72 hours post transfection, cells were trypsinized by adding TrypLE™ (Thermo Fisher Scientific) to the center of each well and incubating at 37°C for approximately 5 minutes. D10 media was then added to each well and mixed to resuspend cells. The resuspended cells were centrifuged for

10 minutes to obtain a pellet, and the supernatant was discarded. The cell pellet was then resuspended in QuickExtract™ buffer (Lucigen®), and cells were incubated at 65°C for 15 minutes, 68°C for 15 minutes, and 98°C for 10 minutes.

Next Generation Sequencing (NGS) samples were prepared by two rounds of PCR.

- 5 Three technical replicates were analyzed per target for the reference and each variant. The first round (PCR1) was used to amplify specific genomic regions depending on the target. Round 2 PCR (PCR2) was performed to add Illumina adapters and indices. Reactions were then pooled and purified by column purification. Sequencing runs were performed using a 150 Cycle NextSeq 500/550 Mid or High Output v2.5 Kit or a 200 Cycle NovaSeq 6000 SP or S1
10 Reagent Kit v1.5.

For NGS analysis, the indel mapping function used a sample's fastq file, the amplicon reference sequence, and the forward primer sequence. For each read, a kmer-scanning algorithm was used to calculate the edit operations (match, mismatch, insertion, deletion) between the read and the reference sequence. In order to remove small amounts of primer
15 dimer present in some samples, the first 30 nt of each read was required to match the reference and reads where over half of the mapping nucleotides are mismatches were filtered out as well. Up to 50,000 reads passing those filters were used for analysis, and reads were counted as an indel read if they contained an insertion or deletion. The QC standard for the minimum number of reads passing filters was 10,000.

- 20 For each target, indel ratios, referring to the fraction of NGS reads containing indels, were calculated for each sample and its cognate no protein control. Targets comprising a higher percentage of indels when the CRISPR nuclease was included in the transfection were indicative of DNA editing outcomes in the cell.

As shown in **FIG. 1**, each of the six targets tested demonstrated a greater level of indels
25 observed when the CRISPR nuclease plasmid was present. This Example thus shows that the CRISPR nuclease of SEQ ID NO: 1 edited human genes.

Example 2: Effectiveness of Variant CRISPR Nucleases for Targeting of Exemplary Mammalian Genes

- 30 This Example describes indel assessment on exemplary mammalian targets using CRISPR nuclease variants transfected into HEK293T cells.

Arginine scanning mutagenesis was performed to individually substitute selected non-arginine residues of the reference CRISPR nuclease (SEQ ID NO: 1) to arginine. SEQ ID NO: 1 is referred to herein as the reference sequence. This resulted in 372 single arginine

substitution variants. Nucleic acids encoding the reference and each CRISPR nuclease variant were then individually cloned into a pcDNA3.1 backbone (Invitrogen™), and the plasmids were prepped and diluted. The plasmids comprised a CMV promoter, a first NLS (MKRTADGSEFESPKKKRKV; SEQ ID NO: 3) upstream of the coding sequence, an XTEN linker (SGGSSGGSSGSETPGTSESATPESSGGSSGGSS; SEQ ID NO: 8), and a second NLS (KRPAATKKAGQAKKKK; SEQ ID NO:4) downstream of the coding sequence. *See also Example 1* above.

Exemplary RNA guides of VEGFA-T6 and EMX1-T7 were used in this study. Details of these gRNAs are provided in **Table 2** above. RNA guides were cloned into a pUC19 backbone (New England Biolabs®). The plasmids were purified using a maxi-prep kit and diluted. Cells were transfected, and samples were prepared for NGS as described in **Example 1**. Indel ratios, referring to the fraction of NGS reads containing indels, were calculated for the reference and for each variant. The indel ratios used for fold change calculations were the average of two technical replicates. To then calculate fold change in indel ratios, the indel ratio for each variant was divided by the indel ratio for the reference. **Table 3** shows fold change in indel ratios for each target tested. Numbering is relative to the reference nuclease of SEQ ID NO: 1 (*i.e.*, without an NLS).

As shown in **Table 3**, 6 of the 372 variants with single arginine substitutions (left column) were characterized as yielding at least a 1.5X increase in indel ratio relative to the reference indel ratio, when averaged across the two targets (right column).

Table 3. Fold Change in Indel Ratios*

Substitution	EMX1	VEGFA	Average Fold Change
I857R	1.70	2.42	2.06
N813R	1.50	1.92	1.71
L784R	1.05	2.17	1.61
K736R	0.95	2.21	1.58
A919R	1.26	1.85	1.55
Q812R	1.31	1.70	1.51

* Variant indel ratio/Reference indel ratio

55 variants with single arginine substitutions were analyzed as having indel ratios 1X-1.5X of the reference indel ratios: G1329R, Q741R, P1238R, A1236R, Q1230R, E1214R, A730R, Q849R, S473R, D985R, M753R, K918R, I1106R, M1205R, Y1015R, Q1360R, K1344R, F501R, K1091R, G731R, N1295R, F1241R, V752R, N1099R, E1179R, D720R, F1285R, I1370R, E1037R, E982R, S1094R, S872R, P1096R, A1226R, Q840R, I1147R,

L1117R, E331R, T1108R, K1298R, Y986R, S898R, A1333R, P1208R, E1105R, Q809R, L1281R, A1292R, K1131R, I581R, I79R, D1284R, T1104R, Y348R, and S1348R. The remaining variants with single arginine substitutions (311 variants) resulted in decreased indel ratios relative to the reference indel ratios (fold change in indel ratios of less than 1.0).

- 5 The following variants, which exhibited at least a 1.5X-fold increase in indel ratio relative to the reference, were selected to engineer combination variants: I857R, N813R, L784R, K736R, A919R, and Q812R.

10 **Example 3: Effectiveness of Combination CRISPR Nuclease Variants for Targeting of Mammalian Genes**

This Example describes indel assessment on mammalian targets using CRISPR nuclease variants comprising two or more substitutions identified as increasing indel activity in Example 2. 35 combination CRISPR variants were tested.

- Each CRISPR nuclease variant and RNA guide was cloned as described in **Example 2**.
- 15 Exemplary RNA guides of VEGFA-T6 and EMX1-T7 were used in this study. Details of these gRNAs are provided in **Table 2** above. HEK293T cells were further transfected, followed by NGS analysis, as described in **Example 2**. For each target, indel ratios, referring to the percentage of NGS reads comprising indels, were calculated for the reference CRISPR nuclease (SEQ ID NO: 1) and for each variant CRISPR nuclease. The indel ratios shown in
- 20 **Table 4** were calculated as the average of two bioreplicates, each of which contained two technical replicates.

Table 4. Indel Ratios for Mammalian Targets

Substitutions	EMX1 Average Indel Ratio	VEGFA Average Indel Ratio	Average indel Ratio (Both Targets)
I857R, L784R, K736R	0.233	0.349	0.291
I857R, A919R, K736R	0.235	0.341	0.288
I857R, N813R, L784R	0.228	0.325	0.276
I857R, L784R, A919R	0.231	0.318	0.275
I857R, N813R, K736R	0.205	0.333	0.269
I857R, N813R	0.203	0.311	0.257
L784R, A919R, K736R	0.200	0.311	0.255
I857R, L784R	0.196	0.308	0.252
I857R, A919R	0.200	0.302	0.251
I857R, N813R, A919R	0.211	0.285	0.248
N813R, L784R, K736R	0.180	0.295	0.238
N813R, L784R, A919R	0.191	0.266	0.229

N813R, A919R, K736R	0.167	0.290	0.229
L784R, K736R, G1329R	0.170	0.276	0.223
I857R, L784R, G1329R	0.168	0.278	0.223
I857R, K736R	0.154	0.291	0.222
L784R, A919R	0.168	0.258	0.213
A919R, K736R	0.141	0.254	0.197
I857R, G1329R	0.139	0.245	0.192
L784R, A919R, G1329R	0.152	0.225	0.188
I857R, K736R, G1329R	0.124	0.252	0.188
N813R, A919R	0.138	0.236	0.187
N813R, K736R	0.118	0.245	0.181
L784R, K736R	0.126	0.230	0.178
K736R, G1329R	0.132	0.222	0.177
N813R, L784R	0.135	0.216	0.176
A919R, K736R, G1329R	0.118	0.212	0.165
I857R, N813R, G1329R	0.120	0.204	0.162
A919R, G1329R	0.137	0.180	0.158
L784R, G1329R	0.115	0.175	0.145
I857R, A919R, G1329R	0.109	0.181	0.145
N813R, L784R, G1329R	0.107	0.172	0.139
N813R, A919R, G1329R	0.106	0.162	0.134
N813R, K736R, G1329R	0.082	0.180	0.131
N813R, G1329R	0.090	0.150	0.120
Reference	0.096	0.139	0.118

As shown in **Table 4**, each of the CRISPR nuclease variants with combinations of amino acid substitutions exhibited higher indel activity than the reference CRISPR nuclease (SEQ ID NO: 1). 9 CRISPR nuclease variants resulted in indel ratios of over 0.25 when averaged across both targets, indicating that over 25% of NGS reads comprised indels. These 9 CRISPR nuclease variants comprised the following substitution combinations: a) I857R, L784R, K736R; b) I857R, A919R, K736R; c) I857R, N813R, L784R; d) I857R, L784R, A919R; e) I857R, N813R, K736R; f) I857R, N813R; g) L784R, A919R, K736R; h) I857R, L784R; and i) I857R, A919R. 8 CRISPR nuclease variants resulted in indel ratios between 0.2 and 0.24 when averaged across both targets, indicating between 20% and 24% of NGS reads comprised indels. 18 CRISPR nuclease variants resulted in indel ratios of 0.1 to 0.19 when averaged across both targets, indicating between 10% and 19% of NGS reads comprised indels. The average indel ratio across both targets exceeded that of the reference for all variants tested.

Based on this experiment, the top-performing CRISPR nuclease variant comprising substitutions I857R, L784R, K736R was selected for further testing. This CRISPR nuclease variant exhibited a 2.5-fold increase in indel activity compared to the reference CRISPR

nuclease.

Example 4: Engineering and Effectiveness of CRISPR Nickase Variants for Targeting Mammalian Genes

5 This Example describes introducing mutations into the CRISPR nuclease of SEQ ID NO: 1 that disrupt either the HNH or RuvC domains to produce a functional nickase. D844, H845, and N868 were identified as putative catalytic residues of the HNH domain. D10, E763, and D991 positions were identified as putative catalytic residues of the RuvC domain. These positions were identified by analyzing models generated with AlphaFold2 (Jumper *et al.*,
10 Nature 596: 583-9 (2021)) for structural regions resembling known HNH and RuvC active sites and/or by performing sequence alignments to other nucleases for which candidate positions had been previously identified. Examples of reference structures used to identify the HNH and RuvC active sites are represented with the following Protein Data Bank (PDB) identifiers: 5h0m, 7eu9, 6ltu, 7odf, 7lys, 8dc2, 4cmp, 4oo8, 7z4j, 5axw, 5b2o, 6kc8, 7utn, 8csz, 8ctl,
15 8dmb.

 The coding sequence of the reference CRISPR nuclease was converted into an *E. coli*-codon optimized DNA sequence, synthesized, and cloned into a pET-28a(+) vector (Novagen) containing lac and T7 RNA polymerase promoters for gene expression. To test for nickase activity, individual alanine mutants were cloned for each of the positions identified as putative
20 active site residues of the HNH and RuvC domains. A leucine mutant was also cloned for position H845. Research grade plasmids were received from GenScript. The engineered nickase sequences are shown in **Table 5**. The codon encoding the substituted residue is capitalized, bold, and underlined in the nucleotide sequence, and the substituted residue is shown in bold and underlined in the amino acid sequence. The putative HNH-knockout
25 nickases were anticipated to cleave the non-target strand but not the target strand. The putative RuvC-knockout nickases were anticipated to cleave the target strand but not the non-target strand.

Table 5. CRISPR Nuclease and Nickase Sequences

30

CRISPR Nucleases	Nucleotide Sequence (Codon optimized for <i>E. coli</i> cells)	Amino acid sequence
Reference	atgaataaacctattcaataggactagatcggcaccaatagcgtgggct ggagcatcatcactgacgattacaaagtcccgcaaaaagatgcgtgttct cggcaataaccgacaaggagtacattaagaaaaatctcatcggtgcaactgct gttcgatgggtgcaataccgcggcgaccgccgcttaaaacgtaccgcc gtcgtcgttacaccggcgccgaaccgcacatctgtatctcaggaaatttt gctgaagagatgtcgaaggtcgacgactcctcttcatcgtctggaggata	SEQ ID NO: 1

gcttctggggaagaggacaaacgtgtagcaagtaccctatcttcgcgac tcttcaagaagaaaaggactatcacgaaaagttccgacgatctatcacctg cgtaaaggagctcgtgataagcaggagaaagccgatccgcttgatctatt tagcgtggcccacatcaatagttccggtcattttctgattgaagacgac cgttttgacgttcgtaacaccgacattcaaaagcaataccaagctttcttga gatctttgataccacctttgagaataaccatctgctgcaaaacgtggacg ttgagccatcctgaccgataagatcagcaaaagcgtataaaaggatcgta tttggcgcagtagcccaaccagaaatccaccggtattttgcagaattttga agctgattgtcgtaatcaagccgactcaaaaaacacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgaggtatctggaga acctgctgggtcagatcgtgatgagttcgcggtattttgagcgcgcaaa aaaattatagactcagttctgttctgttattctgacggtgattgacctgagc actaaggctccgttgagtgctgtctatgatccagcgtacgtatgagcaccgc gaggacctcaaacagctgaaacaattgtgaaagcgtccctgccggagaa gtatcaggagatcttctgactccttaaggacggtacgcaggttatattg aaggtaaacgaaccaggagggcttataagtaacaaactgctaa ccaacaagaagactccgaaaactcctggaaaagatcaagaacgaagac ttctgcgtaagcagcgtacattcgacaacggtagcattccgcatcaggttc atctgaccgaactgaaagctatcatccgtcgtcaaacgagtagtactccgttt ttaaaagaaaaccaggatcgatcgataaaaatcctgacctccgtattccgta ttatattggtccactggcgcgtgaaaagagcgaactcgcgtggatgacctgc aaaaccgatgacagcattcgcctgtggaatttgaagattgtgctgataagg agaagctcgcagaagcgtttatccaccgatgactaacaacgatttctacctg ccggaagagaagtttccaaaacacagcctgatttacgagaagttacc gtttataatgagctaacgaaagttctgtacaaagagcaggggtgaaacgt acttctttgattccaacatcaacaagaattttcgtatgggtttttaagagcat cgtaaagtttccaaaagaagcgttagatttctcgcgaaggagtacgagg agttccgtatcgtcagcgtgacgtgtgacaaaagagaacaaagcgtttaa cgccagcctgggcacatcatcagacctcgagaaaatcctggataaggactt cttagacaatccggataacgaatctattcttgaggacattgtccagacctga cgttgttcgaagaccgtgaaatgattaagaagcgttggagaattataaggat ttattaccgaatctcaactgaagaactgtaccgtcgtcattataccggttgg ggtcgcctgtcgcgaagttaataatggtatccgggataaagaagccag aagaccattttgactacttgattgacgacgggtcgtccaaccgtaactcat gcagctgatcaacgatgacggctgtctttaagagcatatttcaaggcg caagcaggcagccatagcgacaacctgaaagagggtgtgcaattggc gggcagcccggccatcaaaaagggtattctgagcttttgaagatcgttgat gagctgtgaaaagtcattgggttatgagccggaacagattgtgtggaatg gcacgtgaaaaccaaacgaccaatcaggggcgtcgaactctcgtcaaa atataaactcgtgtagcgtgtgaaaaacttgccgagcgacctgaacgg taacatcctgaaggataacccgacagataaccaggctctgcagaatgaacg tctgttttgtactacctgcaaacggcgtgatattgtatctggtgagccct ggacatcgataacttgagccaatcagatagatcatattatcccgagccct tcatcaaggacgactctatcgacaaccgtgttctggttagcagcgccaaaaa tcgtgtaaatccgacgacgtaccaagcttgagatcgtgaagactgcaa agtgttctgaaaaaactgttgacgcaaaactatgagccagcgcaata cgacaatttgacgaaggcggagcgtgtgtgacctgacctggatgacaagg cgcgtttattcaacgtcaactgttgaaaccaggcagattaccaagcacgt ggctcgaatcctgacgaacgttcaataatgagttggattccaaggccgt cgtatccgcaagtgaaaattgtaccctgaagagcaacctggtgtgaatt tcagaaaggagttcgggttttacaaaatagagaagtgaacgattatcatatg cacacgatccctatctcaatgctgtgtgtgcgaaggccatcttgaccaaatat ccgcaactggaaccggaattcgtttatggcctgtatcgtcgaagaagcgtgt cgaaaatcgttagagagataaaggagaaaagtactctgagcgacccgc aaaatgtttttatagcaatttaataatgattttaaactgtcgtgagacttgc	
--	--

	agatgggtcaattgtctgtacgtccggttatcgaaacgggacgctatagggta agaccgcatgggataaaaaaacactttgcgaccgtgaggaagggtctta gctacccgcaagtgaattgtgaaaaagaccgagattcagaccgggtgct tttagcaaggaaagcattctcgcgcacggcaactctgataagttgatccacg taagaccaaggatattctgtgacccgaaaaagtacggcggcttgattca ccgattgttgcataatagcgttctgtgtgagcagatatcaagaaggcgaagg cgcagaaaactaaaaaccgtaccgaactcctcgggtataaccattatggaacg ttccaagtctgagaagaaccggagcgcgttctggagagtaagggtacct gaacatccgtgatgataaattaatgattcttccgaaatactcttattcgaagt gagaacggcgtcgtcgttacttgcctcagcgggtgaactgcaaaaagg caacgaactggcgttgcgacgcaattatgaaattcctgtaccttcaaggcc gctataacgagtcctcaaggcgaaggcggagaaatcgaagaagcaggga gctcgtgaaaccaacagttagctatttcgacgacatcttgaactgattaacg actttagcaaacgtgttatacttgcggacgcgaatctggagaaatcaaca actgtaccaggataacaagagaacattagcgtagacgagctggcaaca atattattaacctgtttacatttacttctactggggctccggcggcgttcaatt tttagacaagattgttaccgtaagcgttacaccagtagcaagaaggtctca atagcaccctgatccaccagagcatcacgggcctgtacgaaccctgatcg atctgtccaaattgggggaagac (SEQ ID NO:278)	
D844A (putative HNH-knockout)	atgaataaacctattcaatagactagatatcggcaccatagcgtgggt ggagcatcatcactgacgattacaagtcgccgcaaaaagatgcgtgtct cggcaataccgacaaggagtacattaagaaaaatctcgcgtgactgct gtctgatgttgcaataccgcggcggaccgcccttaaaacgtaccgcc gtctgtctgtacaccggcggcgaaccgcatcctgtatcttcaggaatttt gctgaagagatgtcgaagtcgacgactccttcttctcatcgtctggaggata gctttctgggtggaaggacaaacgtgtagcaagtaccctatcttcgcgac tctcaagaagaaaaggactatcacgaaaagtctccgacgactatcacctg cgtaaggagctcgtgataagcaggagaaagccgatccgcttgatctatt tagcgtggcccatcattaaagttccggtcattttctgattgaagacgac cgttttgacgttctgaacaccgacattcaaaagcaataccaagctttctgga gatctttgataccacctttgagaataaccatctgctgtcgcgaacgtggagc ttgagggccatcctgaccgataagatcagcaaaagcgtcaaaaaggatcgt tttggcgcgactcccaaccagaaatccaccgggtattttcgcagaattttga agctgattgtcgtgaatcaagccgactcaaaaaacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgagatctggaga acctgctgggtcagatcgtgtagagttcgcgattgttttagcgcgcga aaaaattatcagactcagttctgtgtctgttctgacggtgattgacctgac actaaggtcctggtgagtgctgtatgatccagcgtacgacgacacccgc gagacctcaaacagctgaaacaattgtgaaagcgtccctgccggagaa gtatcaggagatcttctgactccttaaggacggctacgacggttatattg aaggtaaaacgaaccaggagcgttctataagtagcctaagcaaaactgctaa ccaacaagaagactccgaaaactcctggaagaatcaagaacgaagac ttctcgtgaagcagctacattcgacaacgtagcattccgcatcagggtc atctgaccgaactgaagctatcaccgtcgtcaaaagcagtagtactacgttt ttaaaagaaaaccaggatcgcacgaaaaatcctgaccttccgtattccgta ttatattgttccactggcgcgtgaaaagagcgaacttcgctggatgacccgc aaaaccgatgacagcattcggcgtggaattttgaagatttggtcgataagg agaagtctgcagaagcgtttatccaccgcatgactaacaacgattttacctg ccggaagagaagttctgcaaaacacagcctgatttacgagaagttcacc gtttataatgagctaacgaaagtctgtacaagaacgagcagggtgaaact acttctttgattccaacatcaacaagaattttcgtatggtgttttaagagcat cgtaagggtttcaaaaagaagctgttagatttctcgcgaaggagtagcagg agttccgtatcgtcgcgtgacgtggtcgtgacaaagagaacaaagcgtttaa cggcagcctgggcacatcacgacctcgagaaatcctggataaggact cttagacaatccggataacgaatctattctgaggacattgtccagaccctga	MNKPYSIGLDIGTNSVGW SIITDDYKVPAAKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSFHRLSDSFLVEE DKRGSKYPPIFATLQEEKD YHEKFPTIYHLRKEADK QEKADRLIYLALAHIIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDFTFENNHL LSQNVDEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPLQFAKDSYDEDL ENLLGQIGDEFADLFSAA KKL YDSVLLSGILTVIDLS TKAPLSASMIQRYDEHRE DLKQLKQFVKASLPEKYQ EIFADSSKDG YAGYIEGKT NQEA FYKYLSKLLTKQED SENFLEKIKNEDFLRKQRT FDNGSIPHQVHLTELKAI RRQSEYYPFLKENQDRIE KILTRIPY YIGPLAREKSD FAWMTRKTDDSI RPNWFE DLVDKEKSAEAFIHRMTN NDFYLP EEKVLPKHS LIYE KFTVYNELTKVRYKNEQ GETYFFDSNIKQEIFDGVF KEHRKVS KKKLLDFLAKE YEEFRIVDVIGLDKENKAF NASLGT YHDL EKILDKDF LDNPDNESILEDIVQTLTL FEDREMIKKRL ENYKDLF TESQLKKLYRRHYTGWG

	cgtgttcgaagaccgtgaaatgattaagaagcgtttggagaattataaggat tattaccggaatcactgaagaaactgtaccgtcgtcattataccggttgg ggctgcctgtcggcgaaagttaataatggtatccgggataaagaagccag aagaccattttgactacttgattgacgacggctcgtccaaccgtaactcat gcagctgatcaacgatgacggtgtgctttaagagcatcattccaaggcg caagcagcgagccatagcgacaacctgaaaagggtggttggcgaattggc ggcgagcccgccatcaaaaagggtattctgcagctcttgaagatcgttgat gagctgtgaaagtcatgggttatgagccggaaacagattgtggtggaatg gcacgtgaaaaccaaacgaccaatcaggccgctcgaactctcgtcaaaag atataaactgctgatgacggtgtgaaaaacttgccgagcgacctgaacgg taacatcctgaaggaataccgacagataaccagcgtctgcagaatgaacg tctgttttgtactactgcaaaacggccgtgatattgatactggtgagccct gcacatcgataactgagccaatcagatataGCGcatattatcccagcag cctcatcaaggacgactctatcgacaaccgtgttctggttagcagcggcaa aatcgtgtgaaatccgacgacgtaccaagcttgagatcgtgaaagactg caaagtgttctggaaaaaactgttgacgcaaaactcatgagccagcgcaa atacgacaatttgacgaagcgcgagcgtgtggtgacctgacctcgatgaca aggcgcgctttattcaacgtcaactggttgaaccaggcagattaccaagca cgtggctcgaaatcctagacgaacgcttcaataatgattgattccaaggcg cgtcgtatccgcaaaagtgaatgtcacctgaagagcaacctggtgtcga atttcagaaggaggtcgtttttacaaaattagagaagtgaacgattatcatc atgcacacgatgcctatctcaatgctgtgttgcgaaggccatcttgaccaa atatccgcaactggaaaccggaattcgtttatggcctgtatcgtcgaagaag ctgtcgaaaatcgttagagaggataaagaggaaaagtacttgaggcgacc cgcaaaatgtttttatagcaatttaataatgtttaaacgtgtcgtgagact tgcagatggttcaattgtctacgtccgggtatcgaaacgggacgctatatgg gtaagacggcatgggataaaaaaaacactttgcgaccgtgaggaaggtg cttagctacccgcaagtgaattgtgaaaaagaccgagattcagaccggtg gcttttagcaaggaaagcattctcgcgcacggcaactctgataagttgatccc acgtaagaccaaggatattctatctggaatccgaaaaagtacggcgctttgat tcaccgattgttcataatagcgttctggtgtgagcagatatcaagaaggcga ggcgcgaaactaaaaaccgttaccgaactcctcgtgataaccattatggaa cgttccaagtgcgagaagaacccgagcgcgttcttgagagtaaggcgta cctgaacatccgtgatgataaataatgattctccgaaatactcctattcagag ctggagaacggccgtcgtcgttacttgcctcagcgggtgaactgcaaaaa ggcaacgaactggcgttgccgacgcaatttatgaaattcctgtaccttgcaa ggcgtctataacgagtcgcaaaaggcagccggaggaaatcgaaaagaagca ggagttcgtgaaccaacacgttagctatttcgacgacatcttgaactgatta acgactttagcaaacgtgttatacttgcggacgcaatctggagaaaatcaa caactgtaccaggataaacaagaacattagcgttagacgagctggcaa acaatattattaacctgtttacatttacttactggggctccggcggttca aatttttgacaagattgttgaccgtaagcgttacaccagtacgaaagaagt ctcaatagcaccctgatccaccagacatcacggcctgtacgaaaccgt atcgatcgttccaattgggggaagac (SEQ ID NO: 29)	RLSAKLINGIRDKESQKTI LDYLIDDGRSNRNFMLI NDDGLSFKSIISKAQAGSH SDNLKEVVGELAGSPAIAK KGILQSLKIVDELVKVMG YEPEQIVVEMARENQTTN QGRNRNSRQRYKLLDDGV KNLASDLNNGNILEYPTD NQALQNERLFLYYLQNGR DMYTGEALDIDNLSQYDI AHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEI VKDCKVFWKKLLDAKLM SQRKYDNLTKAERGGLTS DDKARFIQRLVETROITK HVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNF RKEFGFYKIREVNDYHHA HDAYLNAVVAKAITLKY QLEPEFVYGLYRRKKLSKI VREDKEEKYSEATRKMMFF YSNLMNMFKRVRRLADG SIVVRPVIETGRYMGKTA WDKKKHFAIVRKVLSYP QVNIVKKTEIQTGGFSKES ILAHGNSDKLIPRKTKDIY LDPKKYGGFDSPIVAYS LVVADIKKGKAQKLKTVT ELLGITIMERSKFEKNPSA FLESKGYLNIRDDKLMILP KYSLFELENGRRRLASA GELQKGNELALPTQFMKF LYLASRYNESKGKPEEIEK KQEFVNQHVSYFDDILQLI NDFSQRVILADANLEKIN KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSIITGLYETRIDLKSLG ED (SEQ ID NO: 30)
H845A (putative HNH-knockout)	atgaataaacctattcaatagactagatcggcaccatagcgtgggct ggagcatcatcactgacgattacaagtcggcgaataaagatgcgtgtct cggcaataccgacaaggagtacattaaagaaaaatctatcgtgactgct gttcgatgttggaataccgcggcgaccgccgttaaaacgtaccgccc gtcgtcttaccaccggcgccgcaaccgatcctgtatcttcaggaaatttt gctgaagagatgtcgaaggtcgcgactccttcttcatcgtctggaggata gctttctgtggaagaggacaaactgttagcaagtaccctatcttcgcgac tcttcaagaagaaaaggactatcacgaaaagtctccgacgatctatcacctg cgtaaggagctcgtgataagcaggagaaaagccgatccgctgtgatctatt tagcgctggccacatcatlaagttccgggtcattttctgattgaagacgac cgttttgacgttcgtaacaccgacattcaaaagcaataccaagctttcttga	MNKPYSIGLDIGTNSVGW SIITDDYKVPAAKKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSSFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKLADK QEKADRLRIYALAHIIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDTTFENNHL

gatctttgataccacctttgagaataaccatctgctgctgcaaaacgtggacg	LSQNV DVEAILTDKISKSA
ttgaggccatcctgaccgataagatcagcaaaagcgctaaaaagatcgta	KKDRILA QYPNQKSTGIFA
ttttggcgagtagcccaaccagaaatccaccggtattttcgagaattttga	EFLKLIVGNQADFKKHFN
agctgattgtcgtaatcaagccgactcaaaaaacacttcaacctggaga	LEDKTPLQFAKDSYDEDL
caaaacaccgctgcagttcgcgaagatagctacgacgaggatctggaga	ENLLGQIGDEFADLFSA
acctgctgggtcagatcgtgatggtcgcgattgttttagcgcgcgcaaa	KKLYDSVLLSGILTVIDLS
aaaattatagactcagttctgttctggttattctgacggtgattgacctgagc	TKAPLSASMIQRYDEHRE
actaaggctccgttgagtcgcttatgatccagcgtacgatgagcaccgc	DLKQLKQFVKASLPEKYQ
gaggacctcaaacagctgaacaatttgaagcgtccctgccggagaa	EIFADSSKDG YAGYIEGKT
gtatcaggagatctttgctgactccttaaggacggtacgaggttatattg	NQEA FYK YLSKLLTKQED
aagggtaaaacgaaccagaggcggttctataagctaaacgaactgctaa	SENFLEKIKNEDFLRKQRT
ccaaacaagaagactccgaaaactcctggaagatcaagaacgaagac	FDNGSIPHQVHLTELKAI
ttcctcgtaagcagcgtacattcgacaacggtagcattccgcatcaggttc	RRQSEYYPFLKENQDRIE
atctgaccgaactgaagctatcctcgtcgtcaagcaggtactatccgtt	KILTFRIPYYIGPLAREKSD
ftaaaagaaaaccggatcgatcgcaaaaaatcctgacctccgtattccgt	FAWMTRKTDDSI RPNWFE
ttatatgttccactggcgcgtgaaaagagcgtctcgtggtgacccgc	DLVDKEKSAEAFIHRMTN
aaaaccgatgacagcattcccccgtggaattttgaagatttggcgataag	NDFYLP EEKVLPKHS LIYE
agaagtctgcagaagcgtttatccaccgcatgactaacaacgatttctacctg	KFTVYNELTKVRYKNEQ
ccggaagagaaaagttctgcaaaacacagcctgatttacgagaagttcacc	GETYFFDSNIK EYFDGVF
gtttataatgagctaacgaaagttcgttacaagaacgagcagggtaaacgt	KEHRKVS KKKLLD FLAKE
acttctttgattccaacatcaaacagaatcttcgatggtgttttaagagcat	YEEFRIVDVIGLDKENKAF
cgttaaggtttccaaaaagaagctgttagatttccctcgcgaaggagtacgagg	NASLGTYHDLEKILDKDF
agttccgtatcgtcgcgtgatcggtctgacaaagagaacaaagcgtttaa	LDNPDNESILEDIVQTLTL
cgccagcctggcgacctatcagcaccctcgagaaaatcctggataaggact	FEDREM IKKRL ENYKDLF
cttagacaatccggataacgaatctattctgaggacattgccagaccctga	TESQLKKLYRRHYTGWG
cgttcttcgaagaccgtgaaatgattaagaagcgtttggagaattataaggat	RLSAKLINGIRDKESQKTI
ttattaccgaaatcgaactgaagaactgtaccgtcgtcattataccgggtgg	LDYLIDDGRSNRNF MQLI
ggtcgcctgtcggcgaaagttaataatgggtatccgggataaagaagccag	NDDGLSFKSIISKAQAGSH
aagaccattttgactacttgattgacgacggctgctccaaccgtaactcat	SDNLKEVV GELAGSPA IK
gcagctgatcaacgatgacggctgttctttagagcatcattccaagcgc	KGILQSLKIVDELVKVMG
caagcagcgacgcatagcgacaacctgaagagggtggttggcgaattggc	YEPEQIVVEMARENQTTN
gggcagcccgccatcaaaaagggtattctgcagcttctgaagatcgttgat	QGRNRSRQRYKLLDDGV
gagctgtgtaaaatcatgggttatgagccggaacagattgtgtggaatg	KNLASDLNGN ILKEYPTD
gcacgtgaaaaccaaagcacaatcagggccgtcgcaactctgtcaaa	NQALQNERLFLYYLQNGR
atataaactctggtgatgacgtgtgaaaaacttgccgagcgacctgaacgg	DMYTGEALDIDNLSQYDI
taacatcctgaaggaaataccgacagataaccaggtctcgcagaatgaacg	DAIIPQAFIKDDSIDNRVL
tcgtttttgactaccgcaaaacggccgtgatagtatactggtgagccct	VSSAKNRGKSDDVPSLEI
ggacatcgataacttgagccaatacgaatagatGCGattatccgcagg	VKDCKVFWKKLLDAKLM
ccctcatcaaggacgactctatcgacaaccgtgttctgttagcagcgccaa	SQRKYDNLTKAERGGLTS
aaatcgtgtgtaaatccgacgacgtaccaagcttgagatcgtgaaagactg	DDKARFIQRQLVETRQITK
caaagtgttctgaaaaaactgttgacgcaaaactcatgagccagcgcaa	HVARILDERFNNELDSKG
atacgacaatttgacgaagcgcgagcgtgtgtgacctgacctggatgaca	RRIRKVIVTLKSNLVS NF
agggcgcgtttattcaacgtcaactgttgaaaccagcgagattaccaagca	RKEFGFYKIREVNDYHHA
cgtggctcgaatcctagacgaacgcttaataatgagttgattccaaggcg	HDAYLNAVVA KAILTKYP
cgtcgtatccgcaagtgaaaattgtcacctgaagagcaacctggtgtcga	QLEPEFVYGLYRRKKLSKI
atttcagaaaggagttcgtgttttcaaaaattagagaagtgaacgattatcatc	VREDKEEKYSEATRK MFF
atgcacacgatgcctatctcaatgctgtgttgccgaaggccatcttgaccaa	YSNLMNMFKRVVRLADG
atatccgcaactggaaccggaattcgtttatggcctgtatcgtcgaaagaag	SIVVRPVIETGRYMGKTA
ctgtcgaaaatcgttagagaggataaagaggaaaagtactctgagcgacc	WDKKKH FATVRKVLSYP
cgcaaaatgtttttatagcaatttaataatgttttaaacgtgtcgtgagact	QVNIVKKTEIQTGGFSKES
tgcatgatgttcaattgtcgtacgtccgttatcgaaacgggacgctatatgg	ILAHGNSDKLIPRKTKDIY
gtaagaccgcatgggataaaaaaaacacttgcgaccgtgaggaagggtg	LDPKKYGGFDSPIVAYS V
cttagctaccgcaagtgaattgtgaaaagaccgagattcagaccgggtg	LVVADIKKGKAQKLKTVT
gcttttagcaaggaaagcattctcgcgcacggcaactctgataagtgatccc	ELLGITIMERSKFEKNPSA
acgtaagaccaaggataatctatctgattccgaaaaagctacggcgcttggat	FLESKGYLNIRDDKLMILP

	tcaccgattgttgcatafagcgttctggttggtgcagatatcaagaaggcgaa ggcgagaaactaaaaaccgttaccgaactcctcggtataaccattatggaa cgttccaagttcgagaagaaccgagcgcgttctggagagtaaggcgta cctgaacatccgtgatgataaaffaatgattctccgaaatctcttattcgag ctggagaacggccgtcgtcgttactgcctcagcgggtgaactgcaaaaa ggcaacgaactggcgttgccgacgcaatttatgaaattcctgtacctgcaa gccgctataacgagtcgaaggcaagccggaggaaatcgaaaagaagca ggagttcgtgaaccaacacgttagctatttcgacgacatcttgaactgatta acgacttttagcaaacgtgttatactgcggacgcaatctggagaaaatcaa caactgtaccaggataacaaagagaacattagcgtagacgagctggcaa acaattatttaacctgtttacatttacttactggggctccggcggttca aatttttgacaagattgtgaccgtaagcgttacaccgtagcaaaagaagt gcaatagcaccctgatccacagacatcagggcgtgtacgaaaccgt atcgatctgtccaaattgggggaagac (SEQ ID NO: 31)	KYSLFELENGRRRLASA GELQKGNELALPTQFMKF LYLASRYNESKGKPEEIEK KQEFVNQHVSYFDDILQLI NDFSKRVLADANLEKIN KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSTGLYETRIDLKSLG ED (SEQ ID NO: 32)
H845L (putative HNH-knockout)	atgaataaacctattcaataggactagatcggcaccaatagcgtgggt ggagcatcatcactgacgattacaaagtcggcaaaaagatgcgtgttct cggcaataccgacaaggagtagcattaagaaaaatctcgcgtgactgct gttcgatggtgcaataccgcccggcggaccgccgttaaacgtaccgccc gtcgtcgttacaccggcggcggcgaaccgcalcctgtatcttcaggaaatttt gtcgaagagatgtcgaaggctgacgactccttcttcatcgtctggaggata gcttctggtggaaaggacaaacgtggtagcaagtacctatcttcgacgac tctcaagaagaaaaggactatcacgaaagtcttcgacgatctatcacctg cgtaaggagctcgtgataagcaggagaagccgatcccgcttgatctatt tagcgtggcccacatcattaagttccggtcgtatcttctgattgaagacgac cgttttgacgttcgtaacaccgacattcaaaagcaatacaagctttctgga gactttgataccacctttgagaataaccatctgctgctgcaaaacgtggacg ttgagccatccctgaccgataagatcagcaaaagcgtcaaaaaggatcgta ttttggcgcagtagcccaaccagaatccaccggtattttgcgagaattttga agctgattgtcgtaatcaagccgactcaaaaaacacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgaggatctggaga acctgctgggtcagatcgtgatgattcgcggtttgttagcgcggcga aaaattatagactcagttctgttctgttattctgacgggtgattgacctgac actaaggctccgtgagtgctgtatgatccagcgtcatgagcaccgc gaggacctcaaacagctgaaacaattgtgaaagcgtccctgccggagaa gtatcaggagatcttctgactccttaaggacggctacgaggttatattg aaggtaaaacgaaccaggagcgttctataagctacctaagcaaacgtctaa ccaacaagaagactccgaaaacttctggaaaagatcaagaacgaagac ttcctgctgaagcagcgtacattcgacaacggtagcattccgcatcaggttc atctgaccgaaactgaaagctatccgtcgtcaaacgagtagtactatccgtt ttaaaaaaaaccaggatcgatcgatcgaataatcctgaccttccgtattccgta ttatattgtccactggcgcgtgaaaagagcgacttcggtggtgacccgc aaaaccgatgacagcattcggcgtggaattttgaagatttggtcgataagg agaagctcgcagaagcgtttatccaccgcatgactaacaacgatttctacgt ccggaagagaagtttcccaaacacagcctgatttacgagaagttcacc gtttataatgagctaacgaaagttcgttacaagaacgagcagggtgaaacgt acttctttgattccaacatcaacaagaattttcgtggtgtttttaagagcat cgtaagggtttccaaaaagaagctgttagatttctcgcgaaggagtacgagg agttccgtatcgtcagctgacgtggtcgtgacaaaagagaacaaagcgtttaa cgccagcctgggcacatcacgacctcgagaaaatcctggataaggactt cttagacaatccggataacgaatctattctgaggacattgacagaccctga cgtgttctgaagaccgtgaaatgattgaagcgtttggagaattataaggat tattcaccgaaatcgaactgaagaactgtaccgtcgtcattataccggttg ggtcgcctgtcgcgaagttaataatggtatccgggataaagaagccag aagaccattttgactacttgattgacgacggtcgtccaaccgtaactcat gcagctgatcaacgatgacggtgtcctttaagagcatcattccaagcgc	MNKPYSIGLDIGTNSVGW SIITDDYKVPKMKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRRLYLQEIFAEM SKVDDSSFFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKEADK QEKADRLRLIYALAHIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDFTTFENHNL LSQNVVDVEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPLOQFAKDSYDEDL ENLLGQIGDEFADLFSA KKLYDSVLLSGILTVIDLS TKAPLSASMIQRYDEHRE DLKQLKQFVKASLPEKYQ EIFADSSKDGAGYIEGKT NQEAFYKYLSKLLTKQED SENFLEKIKNEDFLRKQRT FDNGSIPHQVHLTELKAI RRQSEYYPFLKENQDRIE KILTFRIPYYIGPLAREKSD FAWMTRKTDDSIIRPWNFE DLVDKEKSAEAFIHRMTN NDFYLPPEKVLPKHSIYE KFTVYNELTKVRYKNEQ GETYFFDSNIKQEIFDGVF KEHRKVSKKLLDFLAKE YEEFRIVDVIGLDKENKAF NASLGTYHDLEKILDKDF LDNPDNESILEDIVQTLTL FEDREMIKKRLNYKDLF TESQLKKLYRRHYTGWG RLSAKLINGIRDKESQKTI LDYLIDDGRSNRNFMQLI NDDGLSFKSIISKAQAGSH SDNLKEVVGELAGSPAIK KGILQSLKIVDELVKVMG

	caagcagcgagccatagcgacaacctgaaagagggtggtgagcgaattgccc ggcagcccgccatcaaaaagggtattctgcagctttgaagatcgttgat gagctggtgaaagtcaggggttatgagccggaacagattgtgtggaatg gcacgtgaaaacaaacgaccaatcagggccgtcgcaactctcgcaaaag atataaactgctgagtgacggtgtgaaaaactggcgagcgacctgaacgg taacatcctgaaggatacccgacagataaaccaggctctgcagaatgaacg tctgtttttgactacctgcaaaacggccgtgatatgtactggtgagggccct ggacatcgataactgagccaatcagatatagat CTG attatcccgagg ccttcacaaaggacgactctatcgacaaccgtgttctggttagcagcgcaa aaatcgtggttaaatccgacgacgtaccaagcttgagatcgtgaaagactg caaaagtgttctgaaaaaactgttgacgcaaaaactcatgagccagcgcaa atagacaatttgacgaagcgagcggtgtgtggcctgacctggatgaca cgtggcgctgttaactaacgtcaactggtggaaccaggcagattaccaagca cgtggctcgaatcctgacgaacgcttaataatgagttgattccaaggcg cgtcgtatccgcaaaagtgaattgtcacctgaagagcaacctggtgtcga atttcagaagaggtcgttttacaaaattagagaagtgaacgattatcatc atgcacacgatgcctatctcaatcgtgtgtggtcgaaaggccatctgacca atatccgcaactggaaccggaaltcgttatggcctgtatcgtcgaagaag ctgtcgaaaactggttagagagataaagagaaaaagtactctgaggcgacc cgcaaaatgtttttatagcaatttaataatgattgttaacgtgtcgtgagact tgcatggttcaattgtcgtacgtccggttatcgaacgggacgctatatgg gtaagaccgcatgggataaaaaaaacactttgcgaccgtgaggaaggtg cttagctacccgcaagtgaataattgtgaaaagaccgagattcagaccggtg gcttagcaaggaaaagcattctcgcgacggcaactctgataagttgatccc acgtaagaccaaggataatctatctgacccgaaaaagtacggcggtttgat tcaccgattgttcataatagcgttctggtgtgagagatcaagaaggcgcaa ggcgagaaactaaaaaccgttaccgaactcctcggtatataaccattatgaa cgttccaagttcgagaagaaccggagcgcgttctgagagtaagggtcga cctgaacatccgtgatgataaataatgattcttcgaaatactccttattcga ctggagaaacggccgtcgtcgttacttgcctcagcgggtgaactgcaaaaa ggcaacgaactggcgttgcgacgcaatttatgaattcctgtacattgcaa ggcgtatataacgagtcgaaggcaagccggaggaatcgaagaagca ggagttcgtgaaccaacacgttagctatttcgacgacatcttgaactgatta acgacttagcaacgtgttatacttgcggacgcaatctggagaaaatcaa caactgtaccagataacaagagaacattagcgtagacgagctggcaa acaataattataacctgtttacatttacttactggtgggtccggcggttca aatttttgacaagattgtgaccgtaagcgttacaccagtagaagaagtg ctcaatagcaccctgatccaccagacatcagggcctgtacgaaaccgt atcgatctgtccaaattgggggaagac (SEQ ID NO: 33)	YEPEQIVVEMARENQTTN QGRRNSRQRKYLDDGV KNLASDLNGNILEYPTD NQALQNERLFLYYLQNGR DMYTGEALDIDNLSQYDI DLIIPQAFIKDDSIDNRVLV SSAKNRGKSDDVPSLEIV KDCKVFWKKLLDAKLMS QRKYDNLTKAERGGLTSD DKARFIQRQLVETRQITKH VARILDERFNNELDSKGR RIRKVKIVTLKSNLYSNFR KEFGFYKIKDDSIDNRVLV DAYLNAVVAKAILTKYPQ LEPEFVYGLYRRKKLSKIV REDKEEKYSEATRKMPFFY SNLMNMFKRVRRLADGSI VVRPVIETGRYMGKTAW DKKKHFATVRKVLVSYQP VNIVKKTEIQTGGFSKESI LAHGNSDKLIPRKTCDIYL DPKKYGGFDSPIVAYSVL VVADIKKGKAQKLKTVTE LLGITIMERSKFEKNPSAF LESKGYNLRDDKLMILPK YSLFELENGRRRLASAG ELQKGNELALPTQFMKFL YLASRYNESKGKPEEIEK KQEFVNQHVSYFDDILQLI NDFSKRVLADANLEKIN KL YQDNKENSISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSTGLYETRIDLKSLG ED (SEQ ID NO: 34)
N868A (putative HNH-knockout)	atgaataaacctattcaatagactagatatcgccaccaatagcgtgggt ggagcatcatcactgacgattacaagtcgccgcaaaaagatcgtgttct cggcaataaccgacaaggagtagacattaagaaaaatctatcgtgactgct gttcgatggtggcaataaccgcgccgacggccgcttaaaacgtaccgccc gtcgtgttacaccggcgccgcaacgcacatcctgtatcttcaggaaatttt gctgaagagatgtcgaaggtcgcgactccttctttcatcgtctggaggata gctttctggtggaaggagacaacgtgtagcaagtaccctatcttcgcgac tcttcaagaagaaaaggactatcacgaaaagtttccgacgatctatcacctg cgtaaaggagctcgtgataagcaggagaaagccgatcctcgtgtatctatt tagcgtggcccacatcattagttccggtcgtcattttctgattgaagacgac cgttttgacgttctgaacaccgacatcaaaagcaataccaagctttcttga gatcttgafaccaccttggagaataaccatctgctgtcgcaaaacgtggagc ttgagggccatcctgaccgataagatcagcaaaagcgctaaaaagatcgtga ttttggcgagtagcccaaccagaataccaccggtatttttcgagaattttga agctgattgtcgtgaatcaagccgacttcaaaaacacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgaggtatctggaga	MKNPYSIGLDIGTNSVGW SIITDDYKVPKMKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKEADK QEKADLRILIYALAHIIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDTTFENNHL LSQNVDVEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPLQFAKDSYDEDL ENLLGQIGDEFADLFSA

accctgctgggtcagatcggtgatgagttcgcggattgttttagcgcggcaaa	KKLYDSVLLSGILTVIDLS
aaaattatacactcagttctgttctgtgtattctgacgggtgattgacctgac	TKAPLSASMIQRYDEHRE
actaaggtcctcgtttagtgcgtctatgatccagcgtacgatgagcaccgc	DLKQLKQFVKASLPEKYQ
gaggacctcaaacagctgaacaattgtgaaagcgtccctgccggagaa	EIFADSSKDGYYAGYIEGKT
gtatcaggagatcttctgactcctctaaggacggctacgcaggttatattg	NQEAIFYKYLSKLLTKQED
aaggtaaaaacgaaccaggagggcttctataagtacctaagcaaaactgctaa	SENFLEKIKNEDFLRKQRT
ccaaacaagaagactccgaaaactcctggaaaagatcaagaacgaagac	FDNGSIPHQVHLTEKAI
ttctgcgttaagcagcgtacattcgacaacggtagcattccgcacaggttc	RRQSEYYPFLKENQDRIE
atctgaccgaactgaaagctatcatccgtcgtcaaaagcagttactatccgtt	KILTFRIPPYYIGPLAREKSD
ttaaaagaaaaccaggatcgcatcgaaaaatcctgaccttcctgattccgta	FAWMTRKTDDSIKPNWFE
ttatattgtccactggcgcgtgaaaagagcgaactcgcgtggtgacccgc	DLVDKEKSAEAFIHRMTN
aaaaccgatgacagcattcggcgtggaattttgaagatttggcgtgataagg	NDFYLPPEEKVLPKHSIYE
agaagctcgcagaagcgtttatccaccgcatgactaacaacgatttctacgtg	KFTVYNELTKVRYKNE
ccggaagagaagttctccaaaacacagcctgatttacgagaagttcacc	GETYFFDSNIKQEIFDGVF
gtttataatgagctaacgaaagtctgttacaagaacgagcagggtaaacgt	KEHRKVSCKKLLDFLAKE
acttctttagtccaacatcaacaagaatttctgattggtgttttaagagcat	YEEFRIVDVIGLDKENKAF
cgtaaggtttccaaaaagaagctgttagatttctcgcgaagagtagcagg	NASLGTYHDLKILDKDF
agttccgtatcgtcagcgtgatcggtcgtgacaaaagagaacaaagcgttaa	LDNPDNESILEDIVQTLTL
cggcagcctgggcacatcatcagacctcgagaaaatcctggataaggactt	FEDREMIKKRLRYKDLF
cttagacaatccggataacgaatctattcttgaggacattgtccagacctga	TESQLKKLYRRHYTGWG
cgttcttgaagaccgtgaaatgattaagaagcgttggagaattataaggat	RLSAKLINGIRDKESQKTI
ttattcaccgaatctcaactgaagaactgtaccgtcgtcattataccggttgg	LDYLIDDGRSNRNFMLI
ggctgcctgtcgcggaagttaataatggtatccgggataaagaagccag	NDDGLSFKSIISKAQAGSH
aagaccattttgactacttgattgacgacggctcgtcccaaccgtaactcat	SDNLKEVVGELAGSPAIK
ccagctgatcaacgatgacgctgtcctttaagagcatcattccaagcgc	KGILQSLKIVDELVKVMG
gaagcagggcagccatagcgacaacctgaagagggtggttggcgaattggc	YEPEQIVVEMARENQTTN
gggcagcccgccatcaaaaagggtattctgcagctcttgaagatcgttgat	QGRNRSRQRYKLLDDGV
gagctggtgaaagtcatgggttatgagccggaacagattgtggtggaatg	KNLASDLNGLNKEYPTD
gcacgtgaaaaccaaacgaccaatcaggggcgtcgcaactctcgtcaag	NQALQNERLFLYYLQNGR
atataaactgctgatgacggtgtgaaaaactggcgagcgcacctgaacgg	DMYTGEALDIDNLSQYDI
taacatcctgaaggaataaccgacagataaccaggctcgtcagaatgaacg	DHIIQAFIKDDSIDNRVL
ctgtttttgtactactgcaaaacggccgtgatatgtactggtgagccct	VSSAKARGKSDDVPSLEI
tgacatcgataacttgagccaatcagatatagatcatattatcccgaggcct	VKDKCVFWKKLLDAKLM
tcataaggacgactctatcgacaaccgtgtctgttagcagcgcgcaaaG	SQRKYDNLTKAERGGLTS
CGcgtgtaaatccgacgacgtaccaagcttgagatcgtgaaagactg	DDKARFIQRLVETRQITK
caaaagtgttctgaaaaaactgttgacgcgaaaactcatgagccagcgcaa	HVARILDERFNNELDSKG
atacgacaatttgacgaagcgggagcgtggtggcctgacctggatgaca	RRIRKVKIVTLKSNLVSNF
aggcgcgctttattcaacgtcaactggttgaaccaggcagattaccaagca	RKEFGFYKIREVNDYHHA
cgtggctcgaatccttagacgaacgcttcaataatgagttgattccaagggc	HDAYLNAVVAKAILTKYP
cgtcgtatccgcaaaagtgaattgtcacccctgaagagcaacctggtgtcga	QLEPEFVYGLYRRKKLSKI
atttcagaagagggttcgtttttacaaaattagagaagtgaacgattatcatc	VREDKEEKYSEATRKMMFF
atgcacacgatgcctatctcaatgctgtgtgtgcgaaggccatcttgacaa	YSNLMNMFKRVRRLADG
ataccgcaactggaaccggaatcgtttatggcctgtatcgtcgaaagaag	SIVVRPVIETGRYMGKTA
ctgtcgaaaactgtagagaggataaagaggaaaagtactctgagggcacc	WDKKKHFAIVRKVLSYP
cgcaaaatgttttttatagcaatttaataatgtttaaacgtgtcgtgagact	QVNIVKKTEIQTGGFSKES
tgacagatggttcaattgtctacgtccgggtatcgaaacgggacgctatatgg	ILAHGNSDKLIPRKTCDIY
gtaagaccgcatgggataaaaaaacactttgcgaccgtgaggaagggtg	LDPKKYGGFDSPIVAYS
cttagctaccgcaagtgaattgtgaaaaagaccgagattcagaccggtg	LVVADIKKGKAQKLKTVT
gctttagcaaggaaagcattctcgcgcacggcaactctgataagttgatcc	ELLGITIMERSKFEKNPSA
acgtaagaccaaggatattcttgatccgaaaaagtacggcggctttgat	FLESKGYLNIRDDKLMILP
tcaccgattgttcataatgctgtgtgtgtgcagatatcaagaaggcgaa	KYSLFELENGRRRLLASA
ggcgcagaaactaaaaaccgttaccgaactcctcggtataaccattatggaa	GELQKGNELALPTQFMKF
cgttccaagtctgagaagaacccgagcgcgttctggagagtaaggccta	LYLASRYNESKGPKEIEK
cctgaacatcctgtatgataaataatgattcttccgaataactcttattcgag	KQEFVNQHVSYFDDILQLI
ctggagaacggcgtcgtcgttacttgcctcagcgggtgaactgcaaaaa	NDFSKRVILADANLEKIN

	ggcaacgaactggcgtgccgacgcaattatgaaattcgtaccttgcaa gccgctataacgagtcacaaaggcaagccggaggaaatcgaagaagca ggagttcgtgaaccaacacgttagctatttcgacgacatcttgaactgatta acgacttagcaaacgtgtatcttgcggacgcgaatctggagaaatcaa caaactgtaccaggataacaagagaacattagcgtagacgagctggcaa acaattattataacctgtttacatttacttactggggcctccggcggcgttca aatgtttgacaagattgtgaccgtaagcgttacaccagtacgaagaagt ctcaatagcaccctgatccaccagagcatcacggcctgtacgaaacctgt atcgatctgtccaaattgggggaagac (SEQ ID NO: 35)	KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSIITGLYETRIDLKSLG ED (SEQ ID NO: 36)
D10A (putative RuvC-knockout)	atgaataaacctattcaataggacta CGG atcggcaccaatagcgtgg gctggagcatcatcactgacgattacaagtcgccgcaaaaagatgcgtg ttctcggcaataccgacaaggagtagcattgaagaaatctcatcggtgcact gtctgtcgtatggtggcaataccgcgccggaccgcccgttaaacgtaccg cccgctcgtctgtacacccggcgccgcaaccgcatcctgtatcttcaggaaat ttttgctgaagagatgctgaaggtcgacgactccttcttcatcgtctggagga tagctttctggtggaagaggacaacgtggtagcaagtaccctatcttcgcg actcttcaagagaaaaaggactatcacgaaaagtccgacgatctatcact gctgaaggagctcgtcgtgalaagcaggagaaagccgatccgcttgatcta tttagcgtgcccacatcattaaagtccggtcattttctgattgaagacga ccgtttgacgttcgtaacaccgacattcaaaagcaataccaagctttctgga gatcttgataccacctttgagaataacctatcgtctgcgcaaacgtggacg ttgagccatcctgaccgataagatcagcaaaagcgctaaaaaggatcgtga ttttggcgcatgaccccaaccagaatccaccggtattttcgcagaattttga agctgattgtcgtgaatcaagccgacttcaaaaacacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgaggtctggaga acctgctgggtcagatcggtgatgagttcgcggattgttttagcgcggcaaa aaaattatacactcagttctgttctggtattctgacggtgattgacctgagc actaaggtcctgtgagtgctctatgatccagcgctacgatgagcaccgc gaggacctcaaacagctgaacaattgtgaaagcgtccctgccggagaa gtatcaggagatctttgctgactcctctaaaggacgctacgaggttatattg aaggtaaaaacgaaccaggagcgttctataagtacctaagcaaaactgctaa ccaaacaagaagactcgaaaaacttctggaagaagatcaagaacgaagac ttctgcgtgaagcagcgtacattcgacaacggtagcattccgcatcaggttc atctgaccgaactgaaagctatcaccgtcgaagcaggtactatccgttt tlaaaagaaaaccaggatcgatcgcaaaaatcctgaccttccgtattccgta ttatattggtccactggcgcgtgaaaagagcgcacttgcgtggtgacccgc aaaaccgatgacgacattcccccgtggaattttgaagatttggcgataagg agaagctgcagaagcgtttatccaccgatgactaacaacgatttctacctg ccggaagagaaagtcttccaaaacacagcctgatttaccgagaagttcacc gtttataatgagctaacgaaagtctgtacaagaacgagcaggtgaaacgt acttctttgattccaacatcaacaagaaatttctgatggtgtttttaaagagcat cgtaagggttccaaaaagaagctgttagatttctcgcgaaggagtacgagg agttccgtatcgtcagcgtatcggtctggacaagagaaacaaagcgtttaa cgccagcctgggacacatcacgacctcgagaaaatcctggataaggactt cttagacaatccggataacgaatctatttctgaggacattgtccagaccctga cgtgttctgaagaccgtgaaatgattaagaagcgtttggagaattataaggat ttattcaccgaatctcaactgaagaaactgtaccgtcgtcattataccggttg ggtcgcctgtcggcgaagttaataatggtatccgggataaagaaagccag aagaccattttgactacttgattgacgacggtcgtccaaccgtaacttcat gcagctgatcaacgatgacgctgtctttaaagagcatcattccaagcgc caagcaggcagccatagcgacaacctgaaaagaggtggttgccgaattggc ggggcagcccgccatcaaaaagggtattctgcagcttgaagatcgttgat gagctggtgaaagtcatgggttatgagccggaacagattgtggtgaaatg gcacgtgaaaaccaaacgaccaatcaggccgctcgcaactctcgtcaaa	MNKPYISGLAIGTNSVGW SIITDDYKVPAAKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKLADK QEKADRLRIYLALAHIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDFTTFENHLL LSQNVDEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPQLQAKDSYDEDL ENLLGQIGDEFADLFSAA KKLYDSVLLSGILTVIDLS TKAPLSASMIQRYDEHRE DLKQLKQFVKASLPEKYQ EIFADSSKDGYYAGYIEGKT NQEAFYKYLSKLLTKQED SENFLEKIKNEDFLRKQRT FDNGSIPHQVHLTELKAI RRQSEYYPFLKENQDRIE KILTFRIPPYYIGPLAREKSD FAWMTRKTDDSIIRPNWFE DLVDKEKSAEAFIHRMTN NDFYLPPEKVLPKHSIYE KFTVYNELTKVRYKNEQ GETYFFDSNIKQEIFDGVF KEHRKVSKKLLDFLAKE YEEFRIVDVIGLDKENKAF NASLGTYHLDLEKILDKDF LDNPDNESILEDIVQTLTL FEDREMIKKRLNENYKDLF TESQLKKLYRRHYTGWG RLSAKLINGIRDKESQKTI LDYLIDDGRSNRNFMQLI NDDGLSFKSIISKAQAGSH SDNLKEVVGELAGSPAIA KGILQSLKIVDELVKVMG YEPEQIVVEMARENQTTN QGRRNSRQRYKLLDDGV KNLASDLNNGNILKEYPTD NQALQNERLFLYYLQNGR

	atataaactgctgtagacggtgtgaaaaactggcgagcgacctgaacgg taacatcctgaaggaaatccccacagataaccaggctctgcagaatgaacg tctgttttgtactacctgcaaaacggcgtgatagtatactggtgagccct ggacatcgataactggagcaafacgataatgataatattatcccgagccct tcatcaaggagactctatcgacaaccgtgttctggttagcagcgccaaaaa tcgtggtaaatccgacgacgtaccaagcttgagatcgtgaagactgcaa agtgttctggaaaaaactgttgacgcaaaactcatgagccagcgcaata cgacaatttgacgaaggcgagcgtgtggtgcctgacctggatgacaagg cgcgcttattcaacgtcaactggtgaaccaggcgagattaccaagcacgt ggctcgaaatcctagacgaacgttcaataatgagttggattccaaggcgct cgtatccgcaaaagtgaattgtcacctgaagagcaacctggtgtcgaatt tcagaaaggagttcgggttttacaataatgagaagtgaacgattatcatg cacacgatgacctatctcaatgctgtgtgtgcgaaggccatctgaccaaata ccgcaactggaaccggaattcgtttatgacctgtatcgtcgaagaagctgt cgaaaatcgttagagagataaaggaaaaagtactctgagcgaccgcg aaaatgtttttatagcaatttaataatgttttaaacgtgctgtgagactgc agatggtcaattgtctgtacgtccggtatcgaaacgggacgctatatggta agaccgcagggataaaaaaacactttgcgacgtgaggaaggtgctta gttaccgcgaagtgaattgtgaaaaagaccgagattcagaccggtgct ttagcaaggaaagcatttctcgcacgcaactctgataagtgtatccacg taagaccaaggataatctatctggtacgcaaaaagtacggcggttgaattca cggattgtcatatagcgttctgtgtgtgcagatatcaagaaggcgcaagg cgcaaaaactaaaaaccgttaccgaactcctcgggtataaccattatgaacg ttccaagtgcgagaagaaccggagcggttctggagagtaagggtacct gaacatccgtgatgataaataatgattctccgaataactcttattcagctg gtagaacggccgtcgtcgttacttccagcggtgaactgcaaaaagg caacgaactggcggttgcgacgcaattatgaaatcctgtaccttgaagcc gctataacgagtcgcaaggcaagccggaggaatcgaaaagaagcagga gttcgtgaaccaacacgttagctatttcgacgacatcttgaactgattaacg acttagcaaacgtgttatacttgcgacgcaatctggagaaaatcaaca actgtaccagatacaaaagagaacattagcgtagacgagctggcaaca ataattattaacctgtttacatttacttactgggggtccggcggttcaattt ttgacaagattgttgaccgtaagcgttacaccagtacgaagaagtgcctca atagcacctgatccaccagagcatcacggcctgtacgaaccgtatcg atctgtccaaattgggggaagac (SEQ ID NO: 37)	DMYTGEALDIDNLSQYDI DHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEI VKDCKVFWKKLLDAKLM SQRKYDNLTKAERGGLTS DDKARFIQRLVETRQITK HVARILDERFNNELDSKG RRIRKVKIVTLKSNLVS NF RKEFGFYKIREVNDYHHA HDAYLNAVVAKAILTKYP QLEPEFVYGLYRRKKLSKI VREDKEEKYSEATRKMMF YSNLMNMFKRVRRLADG SIVVRPVIETGRYMGKTA WDKKKHFAVRKVLSTYP QVNIVKKTEIQTGGFSKES ILAHGNSDKLIPRKTIDY LDPKKYGGFDSPIVAYS LVVADIKKGKAKLKTVT ELLGITIMERSKFEKNPSA FLESKGYNLRDDKLMILP KYSLFELENGRRRLASA GELQKGNELALPTQFMKF LYLASRYNESKGKPEEIEK KQEFVNQHVSFYDDILQLI NDFSQRVILADANLEKIN KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSIITGLYETRIDLSKLG ED (SEQ ID NO: 38)
E763A (putative RuvC-knockout)	atgaataaacctattcaataggactagatcggcaccaatagcgtggct ggagcatcatcactgacgattacaagtcggcgaaaaagatgcgtgttct cggcaataccgacaaggagtagacattaagaaaaatctatcgtgacgtgct gttcgatgtgtgcaataccgcgcgacggccgcttaaaacgtaccgccc gtcgtgttacacccggcggcgcaaccgcatcctgtatcttcaggaaatttt gctgaagagatgtcgaaggtcgcgactcctctttcatcgtctggaggata gctttctgggtgaaggagacaacgtgtagcaagtaccctatcttcgcgac tcttcaagaagaaaaggactatcacgaaaagtgtccgacgatctatcacctg cgtgaaggagctcgtgataagcaggagaagccggtatccgcttgcattatt tagcgtgtgcccacatcattagttccggtgcatctttctgattgaagacgac cgttttgacgttcgaacaccgacattcaaaagcaataccaagctttcttga gatctttgataccacctttgagaataaccatctgctgtcgaacacgtggacg ttgaggccatcctgaccgataagatcagcaaaagcgctaaaaagatcgtga ttttggcgcatcctcaaccagaatccaccggtattttcgcagaatttttga agctgattgtcgtgaatcaagccgacttcaaaaaacacttcaacctggaaga caaaacaccgctgcagttcgcgaagatagctacgacgaggtctggaga acctgctgggtcagatcgtgatgagttcgcggtttgttagcgcggcaaaa aaaattatagactcagttctgtgtctgtattctgacggtgattgacctgacg actaaggctccgttgagtcgtctatgatccagcgctacgatgacaccgc	MNKPYSIGLDIGTNSVGW SIITDDYKVPKMKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKLADK QEKADRLRIYLALAHIIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDFTTFENHNL LSQNVDEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPLQFAKDSYDEDL ENLLGQIGDEFADLFSAA KKLYDSVLLSGILTVIDLS TKAPLSASMIQRYDEHRE DLKQLKQFVKASLPEKYQ

gaggacctcaaacagctgaacaatttftgaaagcgctccctgccggagaa gtatcagagatcttctgactccttaaggacggctacgaggttatattg aaggtaaacgaaccaggagggcttctataagtaacaaactgctaa ccaacaagaaagactccgaaaactcctggaaaagatcaagaacgaagac ttctgctgaagcagcgtacattcgacaacggtagcattccgcatcaggttc atctgaccgaactgaaagctatccgtcgtcaaaaggagtagtactccgtt ttaaaagaaaaccaggatcgatcgaaaaaatcctgaccttcgattccgta ttatattgtccactggcgcgtgaaaagagcgaactcgcgtgagtagccgc aaaaccgatgacagcattcggcgtggaattttgaagatttggcgtataagg agaagctcgcagaagcgtttatccaccgcatgactaacaacgatttctacgt ccggagaagaaagtttccaaaacacagcctgatttacgagaagttcacc gtttataatgagctaagaaaagttgttacaagaacgagcagggtagaacgt actcttctgattccaacatcaacaagaattttcgtggtgttttaagagcat cgtaaggtttccaaaagaagctgttagatttcctcgcgaaggagtacgagg agttccgtatcgtcgacgtgacgtggtcgtgacaaagagaacaaagcgtttaa cgccagcctgggcacgtatcagcactcgagaaaatcctggataaggactt cttagacaatccggataacgaatctattctgaggacattgtccagaccctga cgttgttcgaagaccgtgaaatgattaagaagcgttggagaattataagga ttattaccggaatctcaactgaagaactgtaccgtcgtcattataccggttg ggctgcctgtcgcggaagttaataatggtatccgggataaagaagccag aagaccattttgactactgtattgacgacggcgtcgtccaaaccgtaactcat gcagctgatcaacgatgacggctgtcctttaagagcatcattccaaggcg caagcagcgagccatagcgacaacctgaagagggtggtgagcgaattggc ggcgagcccgccatcaaaaagggtattctgacgtcttgaagatcgttgat gagctggtgaaaagtcattgggttatgagccggaacgattgtggtcGCGa tggcacgtgaaaaccaaacgaccaatcaggccgctcgcaactcgtctcaa agataaaactgctgtagacggtgtgaaaacttggcgagcgacctgaac gtaacatcctgaaggatacccgacagataaccaggtcgtgcagaatgaa cgtctgtttttgactacgtcaaaacggcgtgatatgtactggtgagggc ctggacatcgataactgagccaatcgatagatcatattatcccgagcg cttcatcaagcagcactctatcgacaaccgtgttctggttagcagcgccaaa aatcgtggttaaatccgacgacgtaccaagcttgagatcgtgaaagactgc aaagtgttctggaataaactgttgacgcaaaactcatgagccagcgcaaa tacgacaatttgacgaaggcggagcgtggtgacctgacctggatgacaa ggcgcgctttattcaacgtcaactggttgaaccaggcagattaccaagcac gtggctcgaatcctagacgaacgttcaataatgagttgattccaagggcc gtcgtatccgcaaagtgaattgtcacctgaaagagcaacctggtgtcga tttcagaaaaggagttcgggtttttacaaaattgagaagtgaaacgattatcat ttcacacagatgcctatctcaatgctgtggtgcaaggccatcttgaccaaat atccgcaactggaaccggaattcgttatggcgtgtatcgtcgaagaagct gtcgaatacgttagagaggataaagaggaaggtactctgagcgaccc gcaaaatgtttttatagcaatttaataatgaatatgtttaaacgtcgtgagact gcagatggttcaattgtcgtacgtccggtatcgaaacgggacgctatatgg gtaagaccgcatgggataaaaaaacatttgcgaccgtgaggaaggtg cttagctaccgcaagtgaatatgtgaaaaagaccgagattcagaccggtg gctttagcaaggaaaagcattctcgcgcacggcaactctgataagttgatcc acgtaagaccaaggatattatctggatccgaaaagtagcgcgcttggat tcaccgattgttcataatagcgttctggtgtgagatataaagaaggcga ggcgagaaaactaaaaaccgttaccgaactcctcggtataaccattatgga cgttccaagttcgagaagaacccgagcgcttctgagagtaagggtc cctgaacatccgtgatgataaattaatatttccgaataactccttattcga ctggagaacggcgctcgtcgttacttgcctcagcgggtgaactgcaaaaa ggcaacgaactggcggttccgacgaatttatgaaattcctgtaccttgcaa ggcgctataacgagtcgcaaggcgaagcggaggaatcgaagaagga ggaggtcgtgaaccaacacgttagctatttcgacgacatcttgaactgatta	EIFADSSKDGYAGYIEGKT NQEAFYKYLSKLLTKQED SENFLEKIKNEDFLRKQRT FDNGSIPHQVHLELKAII RRQSEYYPFLKENQDRIE KILTFRIPYIIGPLAREKSD FAWMTRKTDDSIIRPNWFE DLVDKEKSAEAFIHRMTN NDFYLPPEEKVLPKHSIYE KFTVYNELTKVRYKNEQ GETYFFDSNIKQEIFDGVF KEHRKVSCKLLDFLAKE YEEFRIVDVIGLDKENKAF NASLGTYHDLEKILDKDF LDNPDNESILEDIVQTLTL FEDREMIKKRLENYKDLF TESQLKKLYRRHYTGWG RLSAKLINGIRDKESQKTI LDYLIDDGRSNRNFMQLI NDDGLSFKSIISKAQAGSH SDNLKEVVGELAGSPAIK KGILQSLKIVDELVKVMG YEPEQIVVAMARENQTTN QGRNRNRQRYKLLDDGV KNLASDLNGNILKEYPTD NQALQNERLFLYYLQNGR DMYTGEALDIDNLSQYDI DHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEI VKDCKVFWKKLLDAKLM SQRKYDNLTKAERGGLTS DDKARFIQRQLVETRQITK HVARILDERFNNELDSKG RRIRKVIVTLKSNLVS NF RKEFGFYKIREVNDYHHA HDAYLNAVVAKAILTKYP QLEPEFVYGLYRRKKLSKI VREDKEEKYSEATRKMF YSNLMNMFKRVRRLADG SIVVRPVIETGRYMGKTA WDKKKHFAIVRKVLSYP QVNIVKKTEIQTGGFSKES ILAHGNSDKLIPRKTKDIY LDPKKYGGFDSPIVAYS LVVADIKKGKAQKLKTVT ELLGITIMERSKFEKNPSA FLESKGYLNIRDDKLMILP KYSLFELENGRRRLASA GELQKGNELALPTQFMKF LYLASRYNESKGKPEEIEK KQEFVNQHVSYFDDILQLI NDFSQRVILADANLEKIN KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST
--	---

	acgactttagcaaacgtgtatacttgcggacgcgaatctggagaaatcaa caaactgtaccaggataacaagagaacattagcgtagacgagctggcaa acaataattataacctgtttacatttacttctactggggctccggcgcttca aatttttgacaagattgtgaccgtaagcgttacaccagctacgaaagaagt ctcaatagcaccctgatccaccagacatcacggcgctgtacgaaaccgt atcgatctgtccaaattgggggaagac (SEQ ID NO: 39)	LIHQSIITGLYETRIDLKSLG ED (SEQ ID NO: 40)
D991A (putative RuvC-knockout)	atgaataaacctattcaataggactagatcggcaccatagcgtggcgt ggagcatcatcactgacgattacaagtcgccgcaaaaaagatgcgtgtct cggcaataccgacaaggagtagtaagaaaaatctcctcggtagcactgct gttcgtaggtgcgaataccgcgccgaccgccgttaaaacgtaccgcc gtcgtgcttacaccggcgccgcaaccgcatcctgtatcttcagaaatttt gtcgaagagatgtcgaaggctcgcgactccttcttcatcgtctggaggata gctttctggtggaaggagacaacgtggtagcaagtaccctatcttcgcgac tctcaagaagaaaaggactatcacgaaagtcttcgacgatctatcacctg cgtaaggagctcgtgataagcaggagaaagccgatctccgcttgatctatt tagcgctggcccacatcattaaagtccggtcattttctgattgaagacgac cgttttgacgttctgaacaccgacattcaaaagcaatcaaacgtttcttga gatctttgataccactttgagaataaccatctgctgcaaaacgtggacg ttgagccatcctgaccgataagatcagcaaaagcgtcaaaaaggatcgta ttttggcgtagtaccgaacccaacgaatccaccggtattttcgcagaattttga agctgattgtcggtaataagccgactcaaaaaacacttcaacctggaga caaaacaccgctgcagttcgcgaagatagctacgacgagatctggaga acctgctgggtcagatcgtgatgagttcgcgattttgtagcggcgcaaa aaaattatagactcagttctgttctgttattctgacgggtgattgacctgac actaaggctccgttgagtgctgtatgatccagcgtacgatgagcaccgc gaggaacctcaaacagctgaacaattgtgaaagcgtccctgccggagaa gtatcaggagatcttctgactccttaaggacggctacgaggttatattg aaggtaaaacgaaccaggagcgttctataagctaaacaaactgctaa ccaaacaagaagactccgaaaactcctggaagaatcaagaacgaagac ttctgcgtgaagcagcgtacattcgacaacgtagcattccgcatcaggttc atctgaccgaactgaaagctatcatcgtcgtcaaaagcgtactatccgttt ttaaaagaaaaccaggatcgcacgcaaaaaatcctgaccttccgtattccgta ttatattggtccactggcgctgaaaagagcgcactcgcgtggtgtagccgc aaaaccgatgacagcattcggccgtggaattttgaagatttggcgtgataag agaagctgtagaagcgtttatccaccgcatgactaacaacgatttctacctg ccggaagagaagttctgcaaaacacagcctgatttacgagaagttcacc gtttataatgagctaacgaaagttcgttacaagaacgagcaggggtgaacgt acttctttgattccaacatcaacaagaatttctgatggtgttttaagagcat cgtaagggttccaaaaagaagctgttagatttctcgcgaaggagtacgagg agttccgtatcgtcagcgtgatcgtgtgacaaaagagaacaaagcgtttaa cgccagcctgggcacatcatcagacctcgagaaaatcctggataaggactt cttagacaatccggataacgaatctattcttgaggacattgtccagaccctga cgttgttgaagaccgtgaatgattgaagcgtttggagaattataaggat ttattcaccgaatctcaactgaagaactgtaccgtcgtcattataccggttg ggtcgcctgtcggcgaaagttaataatggtatccgggataagaagccag aagaccattttgactacttgattgacgacggtcgtccaaccgtaacttcat gcagctgatcaacgatgacgctgtctttaaagagcatattccaagcgc caagcaggcagccatagcgacaacctgaagagggtggttggcgaattggc gggcagcccggccatcaaaagggtattctgagctcttgaagatcgttgat gagctgtgaaagtcaggttgattgagccggaacagattgtggtgaaatg gcacgtgaaaaccaaacgaccaatcaggccgctcgcactctcgtcaaaag atataaactctggatgacggtgtgaaaaacttggcgagcgacctgaacgg taacatctgaaggaaatcccgacagataaccaggctcgtcagaatgaacg tctgtttttgtactacctgcaaaacggccgtgatattgtactggtgagccct	MNKPYSIGLDIGTNSVGW SIITDDYKVPKMKMRVLG NTDKEYIKKNLIGALLFD GGNTAADRRLKRTARRR YTRRRNRILYLQEIFAEEM SKVDDSFHRLSDSFLVEE DKRGSKYPIFATLQEEKD YHEKFPTIYHLRKEADK QEKADLRILIYALAHIKF RGHFLIEDDRFDVRNTDIQ KQYQAFLEIFDFTTFENHNL LSQNVDEAILTDKISKSA KKDRILAQYPNQKSTGIFA EFLKLIVGNQADFKKHFN LEDKTPQLQFAKDSYDEDL ENLLGQIGDEFADLFSAA KKLYDSVLLSGILTVIDLS TKAPLSASMIQRYDEHRE DLKQLKQFVKASLPEKYQ EIFADSSKDGAGYIEGKT NQEAFYKYLSKLLTKQED SENFLEKIKNEDFLRKQRT FDNGSIPHQVHTELKAI RRQSEYYPFLKENQDRIE KILTFRIPYYIGPLAREKSD FAWMTRKTTDDSIKPNWFE DLVDKEKSAEAFIHRMTN NDFYLPPEEKVLPKHSIYE KFTVYNELTKVRYKNEQ GETYFFDSNIKQEIFDGVF KEHRKVSCKLLDFLAKE YEEFRIVDVIGLDKENKAF NASLGTYHDLEKILDKDF LDNPDNESILEDIVQTLTL FEDREMIKKRLNYKDLF TESQLKKLYRRHYTGWG RLSAKLINGIRDKESQKTI LDYLIDDGRSNRNFMLI NDDGLSFKSIISKAQAGSH SDNLKEVVGELAGSPAIK KGILQSLKIVDELVKVMG YEPEQIVVEMARENQTTN QGRNRNRQRYKLLDDGV KNLASDLNGNILKEYPTD NQALQNERLFLYYLQNGR DMYTGEALDIDNLSQYDI DHIIPQAFIKDDSIDNRVL VSSAKNRGKSDDVPSLEI

ggacatcgataactgagccaafacgatagatcatattatcccgaggcct tcacaaaggagactctatcgacaaccgtgttctgttagcagcgcaaaaa tcgtggtaaatccgacgacgtaccaagcttgagatcgtgaagactgcaa agtgttctggaaaaactgttggacgcaaaactcatgagccagcgcaata cgacaatttgacgaaggcggagcgtggtggcctgacctggatgacaagg cgcgctttattcaacgtcaactggtgaaccaggcgacattaccaagcacgt ggctcgaaatcctagacgaacgcttcaataatgagttggattccaaggccgt cgtatccgcaaaagtgaattgtcacctgaagagcaacctggtgtcgaatt tcagaaaggagttcggttttacaaaatagagaagtgaacgattatcatatg cacacGCGgcctatctcaatgctgtgtgcaaggccatcttgaccaa tatccgcaactggaaccggaattcgtttatggcctgtatcgtcgaagaagct gcgcaaatgcttagagaggataaaggaaaaagtactctgagggcagcc gcgcaaatgcttttatagcaatttaataatgtttaaacgtgtcgtgagact gcagatggtcaattgtcgtacgtccggtatcgaaaacgggacgctatatgg gtaagaccgcatgggataaaaaaacatttgcgaccgtgaggaaggtg cttagtaccgcaagtgaattgtgaaaagaccgagattcagaccggtg gcttagcaaggaaaagcattctcgcgacggcaactctgataagttgatcc acgtaagaccaaggataatctatctggtacggaaaaagtacggcgcttgat tcaccgattgttcataatagcgttctggtgtgtgcagatatcaagaaggcaa tcgacgagaaactaaaaaccgttaccgaactcctcggtataaccattatgga cgttccaagtgcgagaagaacccgagcgcgttctggagagtaagggcta cctgaacatccgtgatgataaataatgattcttcgaaataccttattcag ctggagaacggcgtcgtcgttactgcctcagcgggtgaactgcaaaaa ggcaacgaactggcgttgcgacgaatttatgaaattcctgtaccttgcaa ggcgtataacgagtcgaaaggcaagccggaggaaatcgaagaagca ggagttcgtgaaccaacacgttagctatttcgacgacatcttgcaactgatta acgacttttagcaaacgtgttatacttgcggacgcaatctggagaaatcaa caaactgtaccaggataacaagagaacattagcgtagacgagctggcaa acaataattataacctgtttacatttacttctcgtgggctccggcgcttca aatttttgacaagattgttgaccgtaagcgttacaccagtacgaaagaagt ctcaatagcacccgtgatccaggagcatcacggcgctgtacgaaaccgt atcgatctgtccaaattgggggaagac (SEQ ID NO: 41)	VKDCKVFWKKLLDAKLM SQRKYDNLTKAERGGLTS DDKARFIQRQLVETRQITK HVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNF RKEFGFYKIREVNDYHHA HAYLNAVVAKAILTKYP QLEPEFVYGLYRRKKLSKI VREDKEEKYSEATRKMMFF YSNLMNMFKRVRRLADG SIVVRPVIETGRYMGKTA WDKKKHFATVRKVLSTYP QVNIVKKTETQGGFSKES ILAHGNSDKLIPRKTIDY LDPKKYGGFDSPIVAYS LVVADIKKGKAQKLKTVT ELLGITIMERSKFEKNPSA FLESKGYLNIRDDKLMILP KYSLEFENGRRRLASA GELQKGNELALPTQFMKF LYLASRYNESKGKPEEIEK KQEFVNQHVSYFDDILQLI NDFSQRVILADANLEKIN KLYQDNKENISVDELANN IINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNST LIHQSTGLYETRIDLKLG ED (SEQ ID NO: 42)
--	--

A linear DNA template encoding an RNA guide was designed with a T7 promoter upstream and a T7Te terminator sequence downstream. The RNA guide was designed to be specific to a previously tested target sequence, described in **Example 1** and **Table 2** above, within the coding exon of EMX1 with a 5'-NGG-3' PAM sequence (the PAM is 3' of the target sequence). The T7 promoter uses a +1 G at the start of the transcript (*i.e.*, the 5' end of the RNA) for more efficient transcription that is shown for SEQ ID NO: 44. The sequence of the encoded RNA guide and its individual components are shown in **Table 6**.

Table 6. RNA Sequences

Description	RNA sequence	SEQ ID NO
Reference sgRNA scaffold	guuuuagagcugugcugaaaagcacagcacguuuuuuaggcgaguuugaaaaauc caguccguuuucagcuuagaaaagugagcaccgaaucgugcuu	2

EMX1-T2 spacer	UGGAUGGCGACUUCAGGCAC	18
T7Te Terminator	<i>ggcucaccuucgggugggcccuuucugcg</i>	43
T7 transcribed EMX1-T2 sgRNA	<i>gUGGAUGGCGACUUCAGGCACguuuagagcugugcugaaaagcaca</i> <i>gcacguuuuuuagggcagugauuuuuuagggcagucgcuuucagcuuuuuuagggcagucgcuuucg</i> <i>agcaccgaaucggugcuuggcucaccuucgggugggcccuuucugcg</i>	44

A DNA target was designed and ordered as a synthesized linear DNA fragment. The target sequence from EMX1 and 10 bases upstream and downstream within the exon was flanked by 200 bases of unrelated sequence upstream and 100 bases of unrelated sequence downstream. The extra sequence was added so that the cleaved and uncleaved products would separate well on a gel. The target and non-target strands were labelled with 5' IR700 and 5' IR800 labels, respectively, through PCR amplification using labelled primers. The sequences of the DNA target, the individual components of the DNA target, and the labelled PCR primers are in **Table 7**.

Table 7. Target gBlock and Primer Sequences

	DNA sequence	SEQ ID
EMX1-T2 target	TGGATGGCGACTTCAGGCAC	17
EMX1-T2 upstream 10 nt	<i>AGGAAAGCTT</i>	45
EMX1-T2 downstream 10 nt	<i>AGGATAGGCC</i>	46
200 nt upstream sequence	<i>cagcaaaaccgctacttttgacgttccggcggtttttgtggcgagtggtttcggcggtgcgcgcaagatccatctagattatgttaaaggcgagtttacatgaattccgagctccaaaaccgcccgttaaggtagggaatgcgagtcctatgtctgttatacgtgtgttcgccaaaccggtccgatactact</i>	47
100 nt downstream sequence	<i>actactgtccgataatcgaagatagagacgaacacccaaatcacctcagcgcgattagctcatagtactacggtacctccagctctaaagctttgagc</i>	48
Full target gblock	<i>cagcaaaaccgctacttttgacgttccggcggtttttgtggcgagtggtttcggcggtgcgcgcaagatccatctagattatgttaaaggcgagtttacatgaattccgagctccaaaaccgcccgttaaggtagggaatgcgagtcctatgtctgttatacgtgtgttcgccaaaccggtccgatactactAGGAAAGCTTTGGATGGCGACTTCAGGCACAGGATAGGCCactactgtccgataatcgaagatagagacgaacacccaaatcacctcagcgcgattagctcatagtactacggtacctccagctctaaagctttgagc</i>	49
Forward primer	/5IRD800/CAGCAAAACCCGCTACTTTTGGACG	50
Reverse primer	/5IRD700/GCTCAAAGCTTTAGACTGGAGGTACCG	51

Cleavage activity of the reference CRISPR nuclease (SEQ ID NO: 1) and each of the putative nickases was assessed using *in vitro* cleavage assays. Each polypeptide was individually co-expressed with the RNA guide *in vitro* by incubating the plasmid encoding the protein of interest from **Table 5** and linear DNA template for the T7 transcribed EMX1-T2 sgRNA from **Table 6** in a PURExpress® solution (NEB) containing SUPERase•In™ RNase Inhibitor (Invitrogen) for 2 hours at 37°C. The unpurified polypeptide/RNA solution was then

diluted into a solution of 1X NEB Buffer 2 (NEB) containing approximately 1 ng/μl of the labelled DNA target amplicon. The solution was then incubated for 1 hour at 37°C. Reactions were stopped by incubating with RNase Cocktail™ (Invitrogen; approximately 1 U/μl final concentration) at 37°C for 15 minutes, followed by incubating with Proteinase K (NEB; approximately 0.04 U/μl final concentration) at 55°C for 30 minutes. The DNA was then purified using CleanNGS DNA & RNA Clean-Up Magnetic Beads (Bulldog Bio).

The cleaved and uncleaved products of the target and non-target strands were separated by running the samples on a 10% TBE-Urea PAGE gel. The gel was imaged using a LI-COR Odyssey M imaging system using the 700 nm and 800 nm channels to visualize the 5' IR700 and 5' IR800 labels on the target and non-target strands of the target DNA substrate. Band intensities were quantified using ImageJ software.

Gel images are shown in **FIGs. 2A-2C**, and quantification of the percent of cleaved target and non-target strands are shown in **FIG. 2D**. The uncleaved, HNH-cleaved, and RuvC-cleaved strands are indicated. **FIG. 2A** is a gel image captured using the 700 nm channel showing cleavage of the target strand. **FIG. 2B** is a gel image captured using the 800 nm channel showing cleavage of the non-target strand. **FIG. 2C** is an overlay of the gel images from **FIG. 2A** and **FIG. 2B**. As shown in **FIG. 2A-2D**, the reference CRISPR nuclease (SEQ ID NO: 1) cleaved both the target strand and the non-target strand, as expected. Three of the four HNH-knockout nickase constructs (H845A, H845L, and N868A) showed significantly decreased activity on the target strand while retaining activity on the non-target strand. Each of the three RuvC-knockout nickase constructs (D10A, E763A, and D991A) showed significantly decreased activity on the non-target strand while retaining activity on the target strand (**FIGs. 2A-2D**).

This Example thus shows that HNH-knockout nickases and RuvC-knockout nickases were successfully engineered. The H845A variant was chosen to install edits into human gene targets, as described in **Example 5** below.

Example 5: Fusion of CRISPR Nuclease and CRISPR Nickase to a Reverse Transcriptase

In this Example, a reverse transcriptase polypeptide was fused to the C-terminus of the CRISPR nuclease of SEQ ID NO: 1 or an H845A nickase variant of the CRISPR nuclease (SEQ ID NO: 32). See **Tables 1** and **5** above.

A sequence encoding the CRISPR nuclease-reverse transcriptase fusion polypeptide was cloned into a pcDNA3.1 vector (Invitrogen) comprising a CMV promoter. The fusion comprised the following components arranged from N- to C-termini: 1) SV40 NLS, 2) the

CRISPR nuclease of SEQ ID NO: 1, 3) XTEN linker, 4) a variant Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, and 5) the nucleoplasmin NLS. The nucleotide and amino acid sequences of the SV40 NLS, the CRISPR nuclease of SEQ ID NO: 1, the XTEN linker, and nucleoplasmin NLS are shown in **Table 1** in **Example 1**, and the nucleotide and amino acid sequences of the variant MMLV are shown in **Table 9** below. The variant MMLV reverse transcriptase was a human codon-optimized DNA sequence. Research grade plasmids were received from GenScript.

Table 9. Reverse Transcriptase (RT) Sequences

Component	Nucleotide Sequence	Amino Acid Sequence
Variant MMLV	atgaccctaaatagagatgagatcgggtacatgagacctcaaaaga gccagatgtttcttaggggtccacatggctgtctgattttctcaggcctgg gcggaaaccgggggcatgggactggcagttcggcaagctcctctgatc atacctctgaaagcaacctctacccccgtgtccataaaacaatacccat gtcacaagaagccagactggggatcaagccccacatacagagactgtt ggaccagggaatactgttaccctgccagtcacctggaacacgcccct gtacccgttaagaaaccagggaactaatgattataggcctgtccaggatc tgagagaagtcaacaagcgggtggaagacatccaccaccgtgccc aacctttacaacctcttgagcgggctccccacgtccaccagtgtaca ctgtgcttgattaaaggatgaccttttctgacctgagactccaccaccag tcagcctctcttcgcttggagtgagagatccagagatgggaatctcag gacaattgacctggaccagactccacagggtttcaaaacagtccac cctgtttaatgaggcactgcacagagacctagcagacttccggatccag caccagactgtatcctgctacagtacgtggtgacttactgctggccgc cacttctgagctagactccaacaaggctactcggccctgttacaaccc tagggaacctcgggtatcggcctcggccaagaaagcccaatttgc agaaacagggtcaagtatctgggtatcttctaaagagggtcagagatg gctgactgagccagaaaagagactgtgatgggcagcctactccgaa gaccctcgacaactaaggagttcctagggaaggcaggttctgtcgc ctcttcacctgggttgcagaaatggcagccccctgtacctctcacc aaaccgggactctgttaattggggcccagaccaaaaaaggcctatc aagaaatcaagcaagctcttctaactgccccagccctggggttgcaga tttgactaagccctttgaactctttgtcagcagaagcagggtacgcca aagggtgtcctaacgcataaactgggaccttggcgtcggccggtggcct acctgtccaaaaagctagaccagtagcagctgggtgcccccttgct acggatggtagcagccattgccgtactgacaaaggatgcaggcaagct aaccatgggacagccactagtcattctggcccccatgcagtagaggca ctagtcaaacaccccccgaccgtgctttccaacgcccgatgactc actatcaggccttgcctttggacacggaccgggtccagttcggaccgtg gtagccctgaaccggctacgtgctctccactgcctgagggaagggtg caacacaactgccttgatacttggccgaagcccacggaacccgaccc gacctaacggaccagccgtccagacggcaccacacctgttacac ggatggaagcagctctttacaagaggacagcgttaaggcgggagctg cggtgaccaccgagaccgaggttaattctgggctaaagccctgccagcc gggacatccgctcagcgggtgaactgatagcactcaccaggcccta aagatggcagaaggtgaagaagctaaatgtttatactgataccgttatgct tttgctactgcccataccatggagaaatatacagaaggcgtgggtggt cacatcagaaggcaagagatcaaaaaataagacgagatcttggcct actaaaagccctcttctgccccaaagacttagcataatcattgtccagg	MTLNIEDEYRLHETSKEPDV SLGSTWLSDFPQAWAETGG MGLAVRQAPLIPLKATSTPV SIKQYPMSQEARLGKPHIQR LLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVN KRVEDIHPTVNPYNLLSGLP PSHQWYTVLCLKDAFFCLR LHPTSQPLFAFEWRDPEMGI SGQLTWTRLPQGFKNSPTLF NEALHRDLADFRIQHPDLILL QYVDDLLLAATSELDCCQG TRALLQTLGNLGYRASAKK AQICQKQVKYLYLLKEGQ RWLTEARKETVMGQPTPKTP RQLREFLGKAGFCRLFIPGFA EMAAPLYPLTKPGTLFNWGP DQKAYQEIKQALLTAPALG LPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARM THYQALLLDTRVQFGPVVA LNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDAD HTWYTDGSLLQEGQKAG AAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGK KLNVTYDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEI LALLKALFLPKRLSIHCPGH QKGHSAEARGNRMADQAA RKAATETPDTSTLLIENSSP (SEQ ID NO: 53)

	acatcaaaaggacacagcgccgaggtagaggcaaccggatggctg accaagcgccccgaaaggcagccatcacagagactccagacacctct accctctcatagaaaattcatcccc (SEQ ID NO: 52)	
--	---	--

The CRISPR nuclease-reverse transcriptase fusion polypeptide plasmid DNA was then used to install an H845A nickase mutation (*see* **Example 4**) using a site-directed mutagenesis kit (New England Biolabs®). The nucleotide and amino acid sequences of the CRISPR

nuclease-reverse transcriptase fusion polypeptide and the CRISPR nickase-reverse transcriptase fusion polypeptide are shown in **Table 8**. The codon encoding the substituted H845A residue is capitalized, bold, and underlined in the nucleotide sequence, and the substituted residue is shown in bold and underlined in the amino acid sequence. The sequence-verified plasmids were then purified using a Qiagen Maxiprep kit.

Table 8. Fusion Polypeptides

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcacaaaga agaagcggaaagtcaacaagccttacagcatcggcctggatac ggcaccaacagcgtggatggctatcatcacagatgactacaag gtgcctgccaaagaaatgagagtggttaggaacaccgacaagga gtacatcaagaagaacctgacggcgccctctgttcgacggcgg aaacacagccgccgacggcggtgaaaaggaccgccagacg gagatatacgcgcgaggagaaatagaatcctgtacttacaggaaat cttcgtgaggagatgagcaaagtgacgactctttttccacaga ctggaagatagcttttggtggaaggacaaagcagggtctaag tacctatcttcgtactactgcaaggagaaaaggactaccacgag aagttcccaccatctaccacctgaggaaagagctggcggacaaa caagagaaggccgacctgagactgatctaccttgccctggctcat atcatcaagttcagaggccacttcctgacgaggacacagattcg acgtgagaataacagatatccagaagcagttaccaggcctttctgg agatcttcgacaccacctttgaaaacaaccacctgctgagccaaaa cgtggatgtggaagccatcttaaccgataagatctccaagagcgc caagaaagatagaatcctcgccagatcttaaccagaaaagcac cggatcttcgccgaattctgaagctgacgtgggcaaccaggc cgacttcaaaaaacactttaacctggaggacaaaacacctctgca gtttgccaaaggactcctacgacgaagatctgaaaacctgctggg ccagatcggcgacgagttcgccgactctgttcagcgccgctaagaa actgtacgacagcgtgctgctgagcggcatcttgacagtatcgat ctgtccacaaaggccccctgagcggcagcatgatccagagata cgacgagcacagagaggtctgaagcagctgaagcagttgtga aggcctctctgccgagaagtaccaggagatttttgcgactcca gcaaggatggatagccggatattgagggaagaccaaacca ggaagctttctacaagtacgtgtctaaactgctgaccaaacaggaa gattccgagaacttctggagaaaatcaagaacgaagatttctga gaaagcagaggaccttcgacaacggcagcatccccaccaggt ccacctgaccgagctgaaggctatcatcagacgacagagtga actacccttctgaaggaaaatcaagacagaatcgagaagattct gacattcagaatccatactatctcgacctctgccccgtgaaaag	MKRTADGSEFESPKKKRKVN KPYSIGLDIGTNSVGWSIITD DYKVPKKMRVLGNTDKE YIKKNLIGALLFDGGNTAA DRRLKRTARRRYTRRRNRIL YLQEIFAEMSKVDDSFHR LEDSFLVEEDKRGSKYPIFA TLQEEKDYHEKFPTIYHLRK ELADKQEKADLRILIYALA HIIKFRGHFLIEDDRFDVRN TDIQKQYQAFLEIFDTTFEN NHLLSQNVDEAILTDKISK SAKKDRILAQYPNQKSTGIF AEFLKLIVGNQADFKKHFN LEDKTPLQFAKDSYDEdle NLLGQIGDEFADLFSAKK LYDSVLLSGILTVIDLSTKAP LSASMIQRYDEHREDLKQL KQFVKASLPEKYQEIFADSS KDGYAGYIEGKTNQEAfYK YLSKLLTKQEDSENfLEKIK NEDFLRKQRTFDNGSIPHQV HLTELKAIIRRQSEYYPFLK ENQDRIEKILTRIPYYIGPL AREKSDFAWMTRKTDDSiR PWNFEDLVDKEKSAEAFIH RMTNNDfYLPPEKVLPHKS LIYEKFTVYNELTKVRYKN EQGETYFFDSNIKQEIfDGV FKEHRKVSKKKLLDfLAKE YEEFRIVDVIGLDKENKAFN ASLGTYHDLEKILDKDfLD

agcgacttcgctggaaccagaaagaccgatgattctatccgg ccctggaaactcagagacctggtggacaagaaaagtctgccga ggccttcacccaccggatgaccaacaacgacttctacctgcctgaa gagaaggtgctgcctaacaattccctgatctacgaaaagtccacag tgtacaacgagctgaccaaggttagatacaagaacgagcagggc gagacgtacttctcacttaatacaagcaggaattctcctgaggg agtgttcaaggagcaccggagggtctcttaagaagaaactgtaga cttctggttaaggatacaggaattccggtatcgtggacgtgatc ggactggataaggagaacaaagccttaacgccagcctgggaac ataccacgacctggagaagatcctgacaaggacttctgataaa ccccgacaatgagtcacctggaagatatactgagacgctgac actgttcgagagacagagatgatacaagaagactgaaaact acaaggatctgttaccgagagccagctgaagaagtgtatagac ggcactacaccggatggggccgctgtccgctaactgatcaac ggcatcagagataaggaaagtcagaaaaccatcctggactactta atcgacgatggcagaagcaaccggaaactcatgacgtgatcaac gatgatgcttgagcttcaagtctatctcttaagggccagggc gaagccacagcgataacctgaaggaggtggtgggtgagctggc cggctctcctccatcaagaaggcactctgcaagctctaaagatt gtcgacgagctgtgaaggtgagggctatgaacctgagcagatc gtgtgtgaaatggcagagaaaaccagacaaccaatcaaggca gaagaacagcagacagcggtaacgtgctggacgacggcgt gaagaatctggccagcgacctgaacggcaacatcctgaaagagt acccaaccgacaaccagcccctgcagaacgagcgcctgtttctgt actacctgcagaacggcagagacatgtacaccggcgaagccctg gatatcgataacctgagccagtagcagatcaccacataatccac aggcctttatcaaggatgattctatgataatcgggtgctgtctcca gcgcgaagaacagaggaagagcgcacgtgccagcctcg aaatctgaaagactgcaaggtgttctggaagagctcctggacg ccaagctgatgagccagagaaagtacgacaacctgaccaaggc cgagagaggagcctgaccagcgacgataaggctagattcatcc aaaggcagctgtcgaaccagacagatcacaagcacgtggc caggtatcctcgacgagcgggtcaataacgagctgacagcaaaag gtagaagaattagaagggttaagatcgtcacctgaagtctaatctt gtgtccaactccggaaggagttcggcttctacaagatcagagag gtgaacgactaccaccagcccctgatgctacctgaacggcctg gtggccaaagccatcctgaccaataccctcaactggaagccgga gttctgtacggcctgtacagacggaagaagctgagcaaaatcgt gcgggaagataaggagaagtagacgcaagccacaagga aatgttttctatagcaacctgatgaacatgttcaagagagtggtga gactggccgacggatctatcgtcgtcgccctgtgatcgaaccg gcagatacatgggcaaaaccgctgggacaaaagaaagcacttc gccaccgtcggaaggtactgagctacccctcaggtgaacatagt aaaaagacagagatccagaccggcgcttctctaaagagtcctatc ctcgtcacggcaatagcgacaagctgaccccgaaagacaaa agacatctacctggaccctaagaagtatggcgcttcgacagccc catcgtggcttacagcgtcctgtgtggccgatattaagaagggc aaggctcagaagctgaagacggtgacagagctgctgggcataac aatcatggaacgcagcaaatcagagaagaaccaagcgctttct ggaaagcaagggtacctgaacatccgggacgataagctgatga tctgcttaagtagacgctgttcgaactgaaaatggcagacggc ggctgctagccagcggcgagagctgcaaaagggaatgagct ggccctgectaccagttcatgaattcctgtatcgtgcctctagat ataacgagagcaagggcaaacctgaagagatcgaagaagaaca ggagttgtgaaccaacacgtgagctacttcgacgacatcctgcag	NPDNESILEDIVQTLTLFEDR EMIKKRLNENYKDLFTESQL KKLYRRHYTGWGRLSAKLI NGIRDKESQKTILDYLLDDG RSNRNFMQLINDDGLSFKSI ISKAQAGSHSDNLKEVVGE LAGSPAIKKGILQSLKIVDEL VKVMGYEPEQIVVEMARE NQTTNQGRNRNSRQRYKLLD DGVKNLASDLNGNILKEYP TDNQALQNERLFLYYLQNG RDMYTGEALDIDNLSQYDI DHIIQAFIKDDSIDNRVLVS SAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQK YDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILD ERFNNELDSKGRIRKVKIV TLKSNLVSNFKEFGFYKIR EVNDYHHAHDAYLNAVVA KAILTKYPQLEPEFVYGLYR RKKLKIVREDKEEKYSEAT RKMFFYSNLMNMFKRVR LADGSIVVRPIETGRYMG KTAWDKKKHFATVRKVLSY PQVNIIVKSNFIEQGGFSKESI LAHGNSDKLIPRKTCDIYLD PKKYGGFDSPIVAYSVLVVA DIKKGKAQKLKTVTELLGIT IMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELE NGRRRLASAGELQKGNEL ALPTQFMKFLYLASRYNES KGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVLAD ANLEKINKLYQDNKENISV DELANNIINLFTFTSLGAPA AFKFFDKIVDRKRYTSTKEV LNSTLIHQSTIGLYETRIDL KLGEDSGGSSGGSSGSETPG TSESATPESSGGSSGSSSTLNI EDEYRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLA VRQAPLIPLKATSTPVSIIQ YPMSQEARLGKPHIQRLLD QGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKR EDIHPTVPNPYNLLSGLPPS HQWYTVLDDLKDAFFCLRL HPTSQPLFAFEWRDPEMGIS GQLTWTRLPQGFKNSTPLF NEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCCQ GTRALLQTLGNLGYRASAK KAQICQKQVKYLYLLKEG
---	--

ctgafcaatgacttcagcaagagggtgatcctggccgacgccaac ctggaaaaaatcaaaaactatatcaggacaacaaggagaataatc tccgtggacgaactggctaacaacatcattatctgttcacattcac cagccttggcgctcctgccgcctttaagtcttcgacaaaatcgtgg accgcaagcgggtacaccagcaccgaaggagtgcttaacagcac actgatccaccagagcatcactggcctgtatgagacaagaatcga cctgtctaagctggcgaggattctggaggatctagcggaggatc ctctggcagcgagacaccaggaacaagcgagtcagcaacacca gagagcagtggcgagcagcgccggcagcagcaccctaaat atagaagatgagtatcggtacatgagacctcaaaagagccagat gtttcttaggtccacatggctgtctgatttctcagggcctggcg gaaaccggggcatgggactggcagttcgccaagctcctctgat catacctctgaaagcaacctctacccccgtgtccataaaacaatc cccatgtcacaagaagcagactgggagatcaagccccacataca gagactgttgaccagggaatactgttaccctgccagtcacctg gaacacgcccctgtacccgttaagaaaccaggactaatgattat aggcctgtccagatctgagagaagtcaacaagcgggtggaaga catccacccccaccgtgccaaccttacaacctcttgagcgggct cccaccgtccaccagtggtacactgtgcttgatttaaaggatgcct ttttctgctgagactccacccccaccagtcagcctctcttcgccttg agtggagagatccagagatgggaatctcaggacaattgacctgg accagactcccacagggtttcaaaaacagtcaccctgtttaatg aggcactgcacagagacctagcagacttccggtatccagcacca gacttgatcctgctacagtacgttgatgacttactgtgcccga cttctgagctagactgccaacaagggtactcggccctgttacaac cctagggaacctcgggtatcggcgctcgccaagaagcccaaa tttgccagaacaagggtcaagtatctgggtatcttcaaaagggt cagagatggctgactgagccagaaaagagactgtgatggggc agcctactccgaagaccctcgcacaactaaggagttcctaggga aggcagcgttctgtcctcttcatcctgggttgcagaatggca gccccctgtaccctctcaccaaaaccgggactctgttaattggg gcccagaccaaaaaaggcctatcaagaatcaagcaagctcttc taactgccccagccctggggtgccaagtgtgactaagcccttga actcttctgacgagaagcagggtctacgccaagggtgtcctaac gcaaaaactgggaccttggcgtcgccgggtgacctactgtccaa aaagctagaccagtagcagctgggtggccccctgtcctacggat ggtagcagccattgccgtactgacaaggatgcagcgaagctaa ccatgggacagccactagtcttctggcccccatgcagtagagg cactagtcaaacaccccccgaccgtggcttccaacgcccga tgactcactatcaggcccttgccttggacacggaccgggtccagttc ggaccggtgtagccctgaacccggctacgtgctccactgcct gagggaagggtgcaacacaactgccttgatctctggccgaagc ccacggaacccgaccgacctaacgaccagccgctccagac gccgaccacacctgtacacggatggaagcagctcttacaaga gggacagcgtgaaggcgggagctgcgggtgaccaccgagaccga ggtaatctgggctaaagccctgccagccgggacatccgctcagc gggctgaactgatagcactcaccagggccctaaagatggcagaa ggtaagaagctaaatgtttatactgatagccgttatgcttctgctactg cccatatccatggagaaatatacagaagcgtgggtggtcctacat cagaaggcaagagatcaaaaataaagacgagatcttggcccta ctaaaagccctcttctgccccaaaagacttagcataatccattgtcca ggacatcaaaaggacacagcggcaggctagaggcaaccgg atggctgaccaagcggcccgaaggcagccatcacagagactc cagacacctctaccctctcatagaaaattcatcaccaaaaggcc ggcgccacgaaaaaggccggccaggcaaaaaagaaaaag	QRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFN WGPDQQKAYQEIKQALLTA PALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGWRRPVA YLSKKLDPVAAGWPPCLRM VAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDR WLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPE EGLQHNCILDILAEAHGTRP DLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETE VIWAKALPAGTSAQRAELIA LTQALKMAEGKKLVNYTD SRYAFATAHIHGEIYRRRGW LTSEGKEIKNKDEILALLKA LFLPKRLSIHCPGHQKGHS AEARGNRMADQAARKAAI TETPDTSTLLIENSSPKRPA TKKAGQAKKKK (SEQ ID NO: 55)
--	---

	(SEQ ID NO: 54)	
CRISPR nickase (H845A)-reverse transcriptase fusion polypeptide	atgaacggacagccgacggaagcgagttcagtcaccaaga agaagcggaaagtcaacaagccttacgacatcgccctggatc ggcaccacagcgtggatggctatcatcacagatgactacaag gtgcctccaagaaatgagagtgttaggaacaccgacaagga gtacatcaagaagaacctgacggcgccctctgttcgacggcg aaacacagccgacgacgaggaatagaaatcgtactacaggaat gagatatcgcgcggagaaatagaaatcgtactacaggaat cttcgctgaggagatgagcaagtgacgactctttttccacaga ctggaagatagcttttggaggaggaacgagcgtcttaag taccctatcttgctacactgcaagaggaaaaggactaccacgag aagttccaccatctaccacgtgaggaagagctggcgacaaa caagagaaggccgacctgagactgatctaccttgcctggctcat atcatcaagttcagaggccacttctgacgagcagacagattcg acgtgagaatacagataccagaagcagtagcagcctttctgg agatcttcgacaccactttgaaaacaaccacgtctgagccaaa cgtggatgtggaagccatcttaaccgataagatctccaagagcgc caagaagatagaatcctcgccagtagcttaaccagaaaagcac cggtatcttcgcccgaattcctgaagctgacgtgggcaaccaggc cgacttcaaaaacactttaacctggaggacaaaacacctctgca gtttccaaggactcctacgacgaagatctgaaaacctgctggg ccagatcgccgacgagttcgccgatctgttcagcgcgctaagaa actgtacgacagcgtgctgctgagcggcatcttgacagtgatcgt ctgtccacaaaggccccctgagcgcagcatgatccagagata cgacgagcacagagaggtctgaagcagctgaagcagttgtga aggcctctctgcccagaaagtagcagagatttttgcgactcca gcaaggatggatagccggatatactgagggcaagaccaacca ggaagcttttacaagtacgtgtctaaactgctaccaaacaggaa gattccgagaactcctggagaaaatcaagaacgaagatttctga gaaagcagaggaccttcgacaacggcagcatccccaccaggt ccacctgaccgagctgaagcctatcatcagacgacagagtgagt actacccttctgaaggaaaatcaagacagaatcgagaaagtct gacattcagaatccatactatcggaccttgcctggaag agcgacttcgctgtagcaccagaagaccgatgatttatccgg ccctggaacttcgagacctgtggaagaaaagtctgccga ggccttcatccaccggatgaccaacaacgacttctacctgcctgaa gagaaggtgctgcctaaacattcctgatctacgaaaagtacag gtacaacgagctgaccaaggttagatacaagaacgagcagggc gagacgtacttcttcgactctaatacaagcaggaatcttcgatgg agtgtcaaggagcaccggaggtctctaaagaagaactgttaga cttctggtgaagaaatcaggaattccgagcgtggacgtgac ggactggataaggagaacaaagccttaacgccagcctgggaac ataccacgacctggagaagatcctgacaaggacttctgataaa ccccgacaatgagtcacctggaagatatcgtgcagacgctgac actgttcgaggacagagagatgaagaagagactgaaaact acaaggatctgttaccgagagccagctgaagaagtgtatagac ggcactacaccggatggggccgctgtccgctaactgatcaac ggcatcagagataaggaagtcagaaaaccatcctggactactta atcgacgatggcagaagcaaccggaaactcatgacgtgatcaac gatgatgcttgagcttaagttatcatcttaagcccagggc gaagccacagcgataacctgaaggaggtgtgggtgagctggc cgctctcctgcatcaagaaggcactcctgcaaggtctaaagatt gtcgacgagctgtgaaggtgagtgctatgaacctgagcagatc gtgtgtgaaatggcagagaaaaccagacaaccaatcaaggca gaagaacagcagacagcggtagaagctgctggacgacggcgt	MKRTADGSEFESPKKKRKVN KPYSIGLDIGTNSVGWSIITD DYKVPKKMRVLGNTDKE YIKKNLIGALLFDGGNTAA DRRLKRTARRRYTRRRNRIL YLQEIFAEMSKVDDSFHR LEDSFLVEEDKRGSKYPIFA TLQEEKDYHEKFPTIYHLRK ELADKQEKADLRILIYALA HIIKFRGHFLIEDDRFDVRN TDIQKQYQAFLEIFDFTFEN NHLSQNVVDVEAILTDKISK SAKKDRILAQYPNQKSTGIF AEFLKLIVGNQADFKKHFN LEDKTPLFQAKDSYDEDLE NLLGQIGDEFADLFSAAKK LYDSVLLSGILTVIDLSTKAP LSASMIQRYDEHREDLKQL KQFVKASLPEKYQEIFADSS KDG YAGYIEGKTNQEAFYK YLSKLLTKQEDSENFLKIK NEDFLRKQRTFDNGSIPHQV HLELKAIIRRQSEYYPFLK ENQDRIEKILFRIPYYIGPL AREKSDFAWMTRKTDD SIR PWNFEDLVDKEKSAEAFIH RMTNND FYLP EKVLPKHS LIYEKFTVYNELTKVRYKN EQGETYFFDSNIQEIFDGV FKEHRKVSKKLLD FLAKE YEEFRIVDVIGLDKENAFN ASLGTYHDLEKILDKDFLD NPDNESILEDIVQTLTLFEDR EMIKKRLENYKDLFTESQL KKLYRRHYTGWGRLSAKLI NGIRDKESQKTILDYLI DDG RSNRNFMQLINDDGLSFKSI ISKAQAGSHSDNLKEV VGE LAGSPAIKKGILQSLKIVDEL VKVMGYEPEQIVVEMARE NQTTNQGRNRNRQRYKLLD DGVKNLASDLNGNILKEYP TDNQALQNERLFLYYLQNG RDMYTGEALDIDNLSQYDI DAIIPQAFIKDDSIDNRVLVS SAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRK YDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILD ERFNNELDSKGRIRKVKIV TLKSNLVSNFRKEFGFYKIR EVNDYHHAHDAYLNAVVA KAILTKYPQLEPEFVYGLYR RKKLSKIVREDKEEKYSEAT

gaagaatctggccagcgacctgaacggcaacatcctgaaagagt acccaaccgacaaccagccctgcagaacgagcgcctgtttctgt actacctgcagaacggcagagacatgtacaccggcgaagccctg gatatcgataacctgagccagctacgatatcgacGCCataatccc acaggccctttatcaaggtatgtattatgataatcgggtgctgtctc cagcggcgaagaacagaggaaagagcgacgacgtgcccagcct cgaaatcgtgaaagactgcaaggtgttcggaagaagctcctgga cgccaagctgatgagccagagaaagtacgacaacctgaccaag gcccagagagaggagccctgaccagcgacgataaggctagattcat ccaaaggcagctgtgtcgaaccagacagatcaccaagcacgtg gccaggtatcctcagcagcggttcataacgagctggacagcaa aggtagaagaattagaaggttaagatcgtcacctgaagtcta ctgtgtccaaacttccggaaggagttcggcttctacaagatcagag aggtgaacgactaccaccagcccatgatgctacctaagcgc gtgtgtgccaagccatcctgaccaataacctcaactggagccc gagttcgtgtacggcctgtacagcgaagaagctgacaaaatc gtcggggaagataagggaagagaggtacagcgaagccacaagg aaaatgttttctatagcaacctgatgaacatgttcaagagagtggtg agactggccgacggatctatctgtcgtcggcctgtgatcgaacc ggcagatcatgtggcaaacggcctgggacaaaagaagcactt cgccaccgtgcggaagggtactgagctaccctcaggtgaacatagt gaaaaagacagagatccagaccggcggcttctctaaagagtccat cctcgtcacggcaatagcgacaagctgatcccagaaagacaa aagacatctacgtgaccctaaagatgtgcccgttcgacagcc ccatcgtggttacagcgtcctggtgtgcccgatattaagaagg caaggtcagaaagctgaagacgggtgacagagctgctggcataa caatcatggaacgagcaaatcgaagaagaaccaagcgcttttct ggaaagcaagggtacgtgaacatccgggacgataagctgatga tctgcctaagtacagcctgttcgaactggaatggcagacggc ggctgctagccagcggcggagagctgcaaaagggaatgagct ggccctgacctaccagttcatgaattcctgtatctgaccttagat ataacgagagcaagggcaaacctgaagagatcgaagaagaaca ggagttgtgaaccaacacgtgagctacttcgacgacatcctgacg ctgatcaatgacttcagcaagaggtgatcctggccgacgccaac ctggaaaaaatcaacaactatatacaggacaacaaggagaatc tccgtgacgaactggctaacacatcattatctgttcacattcac cagcccttggcgtcctgcccgcctttaagtcttcgacaaaatcgtg accgcaagcgggtacaccagcacaaggaggtgcttaacagcac actgatccaccagagcatcactggcctgtatgagacaagaatcga cctgtctaagctggcgaggatctgagagctatagcggagatc ctctggcagcgagacaccaggaacaagcagtcagcaacacca gagagcagtgccggcagcagcggcggcagcagcacccctaaat atagaagatgagtatcgctacatgagacctcaaaagagccagat gtttcttaggtccacatggctgtctgatttctcagcctggcg gaaaccggggcagtgaggactgagtcgccaagctcctctgat catacctctgaaagcaacctctacccccgtgtccataaaacaatc cccatgtcacaagaagccagactgggatcaagccccacataca gagactgttgaccagggaatactggtaccctgcccagtcacctg gaacacgcccctgctacccgttaagaaaccaggactaatgattat aggcctgtccagatctgagagaagtcaacaagcgggtggaaga catccacccaccgtgccaaccttacaacctcttgagcgggct cccaccgtcccaccagtggtacactgtgcttgatttaaagatgcct ttttctgctgagactcccccaccagtcagcctctcttcgctttg agtggagagatccagagatgggaatctcaggacaattgacctgg accagactcccacagggtttcaaaaacagctcccacctgtttaatg	RKMFFYSNLMNMFKRVR LADGSIVVRPIETGRYMG KTAWDKKKHFATVRKVL SY PQVNIVKKTEIQTGGFSKESI LAHGNSDKLIPRKTCDIYLD PKKYGGFDSPIVAYSVLVVA DIKKGKAQKLKTVTELLGIT IMERSKFEKNPSAFLESKY LNIRDDKLMILPKYSLFELE NGRRRLASAGELQKGNEL ALPTQFMKFLYLASRYNES KGKPEEIEKKQEFVNQHV SY YFDDILQLINDFSKRVILAD ANLEKINKLYQDNKENISV DELANNIINLFTFTSLGAPA AFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLS KLGEDSGSGSGSGSSETPG TSESATPSSGSGSGSSTLN IEDEYRLHETSKEPDVSLGS TWLSDFPQAWAETGGMGL AVRQAPLIPLKATSTPVS SIK QYPMSQEARLGIKPHIQRLL DQGILVPCQSPWNTPLLPVK KPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPP SHQWYTVLCLKDAFFCLRL HPTSQPLFAFEWRDPEMGIS GQLTWTPLPQGFKNSTPLF NEALHRDLADFRIOHPDLIL LQYVDDLLLAATSELDCQQ GTRALLQTLGNLGYRASAK KAQICQKQVKYLYGILLKEG QRWLTEARKETVMGOPTPK TPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFN WGPDQQKAYQEIQALLTA PALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRM VAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDR WLSNARMTHYQALLDLD RVQFGPVVALNPATLLPLPE EGLQHNCLDILAEAHGTRP DLTDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTETE VIWAKALPAGTSAQRAELIA LTQALKMAEGKKLVNYTD SRYAFATAHIHGEIYRRRGW LTSEGKEIKNKDEILALLKA LFLPKRLSIHCPGHQKGHS AEARGNRMADQAARKAAI TETPDTSTLLIENSSPKRPAA TKKAGQAKKKK (SEQ ID
--	--

	aggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtgtagtacttctgctggcccca cttctgagctagactgccaacaaggtagtctggccctgttacaac cctagggaacctcgggtatcgggcctcggccaagaaagcccaaa tttgccagaacaggtcaagtatctgggtatcttctaaaagagggt cagagatggctgactgagccagaaaagagactgtgatggggc agcctactccgaagaccctcgacaactaaggagttcctaggga aggcaggcttctgtcctcttcatcctgggtttgcagaaatggca gccccctgtacctctcaccaaaccgggactctgtttaattggg gccagaccaacaaaaggcctatcaagaatcaagcaagctcttc taactgccccagccctgggttgccagattgactaagcccttga actcttctgtcagagaagcagggtacgccaaaggtgtcctaac gcaaaaactgggaccttggtcggtcgccgtggtgctacctgtcaa aaagctagaccagtagcagctgggtggccccctgctacggat ggtagcagccattgccgtactgacaaggatgcaggcaagctaa ccatgggacagccactagtattctggcccccatgcagtagagg cactagtcaaacaccccccgaccgtggctttccaacggccgga tgactcactatcagcccttgcctttggacacggaccgggtccagtgc ggaccggtgtagccctgaaccgggtacgtgctcccactgcct gaggaaaggctgcaacacaaactgccttgatctctggccgaagc ccacggaacccgaccgacctaacggaccagccgtcccagac gccgaccacacctgtacacggatggaagcagtccttacaaga gggacagcgttaaggcgggagctgctgggtgaccaccgagaccga ggtaacttgggctaaagccctgccagccgggacatccgctcagc gggctgaactgatagcactcaccagccctaaagatgacagaa ggtaagaagctaaatgtttatactgataccgttatgctttgctactg cccatatccatggagaataatacagaaggcgtgggtggtcacat cagaaggcaagagatcaaaaataaagacgagatcttggcccta ctaaaagccctcttctgccccaaaagacttagcataatcattgtcca ggacatcaaaaaggacacagcggcaggttagaggcaaccgg atggctgaccaagcggccgaaaggcagccatcacagagactc cagacacctctaccctctcatagaaaattcatcccaaaaggcc ggcgccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 56)	NO: 57)
--	---	---------

This Example describes how the CRISPR nuclease-reverse transcriptase fusion polypeptide and CRISPR nickase (H845A)-reverse transcriptase fusion polypeptide constructs were cloned. These constructs were used in **Example 6** to install edits in human target genes.

5

Example 6: RNA-Templated Editing of Human Genes in HEK293T Cells Using CRISPR Nuclease-Reverse Transcriptase and CRISPR Nickase-Reverse Transcriptase Fusion Polypeptides

10

This Example shows genetic modification of human genes utilizing the CRISPR nuclease-reverse transcriptase and CRISPR nickase-reverse transcriptase fusion polypeptides constructed in **Example 5**. Specifically, the fusion polypeptides were used to install sequence substitutions of 6 nucleotides into the human target genes.

Editing template RNAs were designed to be specific to the target sequences shown in

Table 11 and cloned into a pUC19 plasmid comprising a U6 PolIII promoter and a 6x polyT terminator sequence. The editing templates were synthesized by GenScript and comprised the following five components, from 5' to 3': 1) spacer sequence, 2) scaffold motif, 3) reverse transcription template (RTT) encoding 6 nucleotide substitutions, 4) primer binding site (PBS), and 5) a 3' extension motif. The RTT was 23-nucleotides in length. Two different PBS sequences, varied in length, were tested: 7-nucleotides (editing templates 1-5) and 9-nucleotides (editing templates 6-10). The 3' extension motif contains a short linker sequence and a pseudoknot. The linker sequence was added to prevent steric clashes between the PBS and the pseudoknot motif. The pseudoknot was added to protect against 3' exonuclease activity of the editing template RNA.

Guides AAVS1-T3, EMX1-T2, EMX1-T7, VEGFA-T3, and VEGFA-T6 were used in this example. The PAM, target, spacer, and gRNA sequences are provided in **Table 2** above. The sequences of each component and the full-length editing template RNA sequences are shown in **Tables 11** and **12** below. The U6 PolIII promoter uses a +1 G at the start of the transcript (*i.e.*, the 5' end of the RNA) for more efficient transcription that is excluded from the sequences described in **Table 12**.

Table 11. Sequences of Editing Template RNA Components

Editing template RNA	Target	Spacer	Scaffold	RTT	PBS	3' Extension
1	AAVS1-T3	UUGUGAG AAUGGUG CGUCCU (SEQ ID NO: 12)	GUUUUAGAGCUGUGC UGAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	CCACGACCU GGUGAACuggaucG (SEQ ID NO: 58)	ACGCACC	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
2	EMX1-T2	UGGAUGG CGACUUC AGGCAC (SEQ ID NO: 18)	GUUUUAGAGCUGUGC UGAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	AGCUGCCUC UGGCCUAaggacaG (SEQ ID NO: 59)	CCUGAAG	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
3	EMX1-T7	CUUUAGC AAGGGAC UAUUCA (SEQ ID NO: 21)	GUUUUAGAGCUGUGC UGAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU	CUUCAUCCCU GCUUCAaggga cA (SEQ ID NO: 61)	AUAGUCC	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU

Editing template RNA	Target	Spacer	Scaffold	RTT	PBS	3' Extension
			GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)			AGAAA (SEQ ID NO: 60)
4	VEGFA-T3	UUGUUGA CAUUGUC CACACC (SEQ ID NO: 24)	GUUUUAGAGCUGUGC UGAAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	GAAGCUGAA AGCCGAUagguccU (SEQ ID NO: 62)	GUGGACA	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
5	VEGFA-T6	UUUGCCC AUACUGG GGACCA (SEQ ID NO: 27)	GUUUUAGAGCUGUGC UGAAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	UCAUGCACU UGACACUaggaacG (SEQ ID NO: 63)	UCCCCAG	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
6	AAVS1-T3	UUGUGAG AAUGGUG CGUCCU (SEQ ID NO: 12)	GUUUUAGAGCUGUGC UGAAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	CCACGACCUG GUGAACuggauAUcG (SEQ ID NO: 64)	ACGCACC	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
7	EMX1-T2	UGGAUGG CGACUUC AGGCAC (SEQ ID NO: 18)	GUUUUAGAGCUGUGC UGAAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	AGCUGCCUC UGGCCUAaggaUcaG (SEQ ID NO: 65)	CCUGAAG	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
8	EMX1-T7	CUUUAGC AAGGGAC UAUUCA (SEQ ID NO: 21)	GUUUUAGAGCUGUGC UGAAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	CUUCAUCCCU GCUUCAaggaUA (SEQ ID NO: 66)	AUAGUCCC UUGCU (SEQ ID NO: 280)	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)

Editing template RNA	Target	Spacer	Scaffold	RTT	PBS	3' Extension
9	VEGFA-T3	UUGUUGA CAUUGUC CACACC (SEQ ID NO: 24)	GUUUUAGAGCUGUGC UGAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	GAAGCUGAA AGCCGAUagguAU ccU (SEQ ID NO: 67)	GUGGACA	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)
10	VEGFA-T6	UUUGCCC AUACUGG GGACCA (SEQ ID NO: 27)	GUUUUAGAGCUGUGC UGAAAAGCACAGCAC GUUAAAAUAAGGCAG UGAUUGAAAAAUCCA GUCCGUAUUCAGCUU GAAAAAGUGAGCACC GAAUCGGUGCUU (SEQ ID NO: 2)	UCAUGCACU UGACACUaggaA acG (SEQ ID NO: 68)	UCCCCAGU	UCUCUCUCU UGACGCGGU UCUAUCUAG UUACGCGUU AAACCAACU AGAAA (SEQ ID NO: 60)

*6-nucleotide substitution in lowercase

Table 12. Full-length Editing Template RNA Sequences

Editing template RNA	Target	Sequence
1	AAVS1-T3	UUGUGAGAAUGGUGCGUCCUGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUCCACGACCUGGUGAACuggau cGACGCACCUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUAAAA CCAACUAGAAA (SEQ ID NO: 69)
2	EMX1-T2	UGGAUGGCGACUUCAGGCACGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUAGCUGCCUCUGGCCUAaggac aGCCUGAAGUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUAAAA CCAACUAGAAA (SEQ ID NO: 70)
3	EMX1-T7	CUUUAGCAAGGGACUAUUCAGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUCUUAUCCUGCUUCAaggga cAAUAGUCCUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUAAAA CCAACUAGAAA (SEQ ID NO: 71)
4	VEGFA-T3	UUGUUGACAUUGUCCACACCGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUGAAGCUGAAAGCCGAUagguc cUGUGGACAUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUAAAA CCAACUAGAAA (SEQ ID NO: 72)

5	VEGFA-T6	UUUGCCCAUACUGGGGACCAGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUUCAUGCACUUGACACUaggaa cGUCCCCAGUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUAAA CCAACUAGAAA (SEQ ID NO: 73)
6	AAVS-T3	UUGUGAGAAUGGUGCGUCCUGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUCCACGACCUGGUGAACuggau cGACGCACCAUUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 74)
7	EMX1-T2	UGGAUGGCGACUUCAGGCACGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUAGCUGCCUCUGGCCUAaggac aGCCUGAAGUCUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 75)
8	EMX1-T7	CUUUAGCAAGGGACUAUUCAGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUUCUUAUCCUGCUUCAaggga cAAUAGUCCCUUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 76)
9	VEGFA-T3	UUGUUGACAUUGUCCACACCGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUGAAGCUGAAAGCCGAUagguc cUGUGGACAAUUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 77)
10	VEGFA-T6	UUUGCCCAUACUGGGGACCAGUUUUAGAGCUGUGCUGAAAAGCACAG CACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCCGUUUUCAGCU UGAAAAAGUGAGCACCGAAUCGGUGCUUUCAUGCACUUGACACUaggaa cGUCCCCAGUAUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 78)

**6-nucleotide substitution in lowercase*

- Approximately 16 hours prior to transfection, 25,000 HEK293T cells in
- 5 DMEM/10%FBS+Pen/Strep (D10 media) were plated into each well of a 96-well plate. On the day of transfection, the cells were 50-70% confluent. For each well to be transfected, a mixture of Lipofectamine 2000™ (ThermoFisher Scientific) and Opti-MEM™ (ThermoFisher Scientific) was prepared and incubated at room temperature for 5 minutes (Solution 1). After incubation, the Lipofectamine 2000™: Opti-MEM™ mixture was added to a separate mixture
 - 10 containing the CRISPR nuclease-reverse transcriptase fusion polypeptide, editing template RNA, and Opti-MEM™ (Solution 2) or CRISPR nickase-reverse transcriptase fusion polypeptide, editing template RNA, and Opti-MEM™ (Solution 2). Solutions 1 and 2 were mixed by pipetting up and down, then incubated at room temperature for 25 minutes. Following incubation, the Solution 1 and 2 mixture was added dropwise to each well of a 96-
 - 15 well plate containing the cells. Approximately 72 hours post transfection, cells were

trypsinized by adding TrypLE™ (Thermo Fisher Scientific) to the center of each well and incubating at 37°C for approximately 5 minutes. D10 media was then added to each well and mixed to resuspend cells. The resuspended cells were centrifuged for 10 minutes to obtain a pellet, and the supernatant was discarded. The cell pellet was then resuspended in Quick
5 Extract™ buffer (Lucigen®), and cells were incubated at 65°C for 15 minutes, 68°C for 15 minutes, and 98°C for 10 minutes.

Samples were prepared for NGS and analyzed as described in **Example 1**. For each target, the fraction of NGS reads containing indels were calculated for each sample and its cognate no protein control. To determine the percentage of edits installed in the target genes,
10 sequencing reads comprising the 6-nucleotide substitution encoded by the editing template RNAs were analyzed and quantified. The percentage of NGS reads comprising indels and the 6-nucleotide edits are shown in **Table 13** and **Table 14** and further depicted in **FIG. 3A** and **FIG. 3B**. In **FIG. 3A** and **FIG. 3B**, the percentage of NGS reads is shown on the y-axis, total edits are shown as in black bars, and 6 nucleotide edits are shown as grey bars. The data in **Table 13**,
15 **Table 14**, **FIG. 3A**, and **FIG. 3B** is the average of three technical replicates.

Table 13. Editing Efficiencies with CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptide

Editing Template RNAs	Average % Indel	Average % 6-nt substitution installation
1	11.67	2.86
2	41.32	12.80
3	23.30	0.29
4	14.47	3.24
5	20.28	6.03
6	10.89	3.06
7	39.67	11.63
8	25.15	0.10
9	17.54	5.28
10	18.78	6.21

Table 14. Editing Efficiencies with CRISPR Nickase-Reverse Transcriptase Fusion Polypeptide

Editing Template RNAs	Average % Indel	Average % 6-nt substitution installation
1	0.44	0.94
2	0.73	10.66
3	0.43	0.06

4	0.32	0.72
5	0.29	2.01
6	0.44	0.65
7	0.74	8.4
8	0.44	0.05
9	0.32	1.30
10	0.29	1.83

As shown in **FIG. 3A** and **FIG. 3B**, the CRISPR nuclease-reverse transcriptase fusion polypeptide and the CRISPR nickase-reverse transcriptase fusion polypeptide, respectively, introduced substitutions encoded by the tested editing template RNAs at AAVS1, EMX1 and VEGFA target loci. For the CRISPR nuclease-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 11.67% to 41.32%, while the average percentage of NGS reads comprising encoding edits ranged from 0.29% to 12.80% (**Table 13** and **FIG. 3A**). Editing template RNA 3 exhibited the lowest incorporation of the encoded edit, while editing template RNA 2 displayed the highest indel and encoded edit installation. For the CRISPR nickase-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 0.29% to 0.74%, while the average percentage of NGS reads comprising encoding edits ranged from 0.05% to 10.66%. (**Table 14** and **FIG. 3B**). For the CRISPR nickase-reverse transcriptase fusion polypeptide, low indel incorporation (under 1%) confirmed that the H845A substitution converted the CRISPR nuclease to a CRISPR nickase. For both the CRISPR nuclease-reverse transcriptase fusion polypeptide and the CRISPR nickase-reverse transcriptase fusion polypeptide, editing template RNA 2 displayed the greatest encoded edit installation.

Conversely, controls consisting of the CRISPR nuclease of SEQ ID NO: 1 with the RNA guides of **Table 2** or the editing templates of **Table 13** and **Table 14** induced indel formation but did not incorporate the 6-nucleotide substitution encoded by the editing template RNAs. Furthermore, controls consisting of the CRISPR nuclease-reverse transcriptase fusion polypeptide or CRISPR nickase-reverse transcriptase fusion polypeptide with the RNA guides of **Table 2** did not result in incorporation of the 6-nucleotide substitution encoded by the editing template RNAs.

Editing efficiency mediated by the CRISPR nuclease-reverse transcriptase and CRISPR nickase-reverse transcriptase fusion polypeptides was further tested with editing template RNAs with PBS lengths of 11, 13, 15, and 17 nucleotides as well as RTT lengths of 18 and 23 nucleotides. These editing template RNAs were found to behave similarly to editing template

RNAs comprising a PBS with a length of 7 or 9 nucleotides and an RTT with a length of 23 nucleotides.

Overall, this Example shows that the CRISPR nuclease-reverse transcriptase fusion polypeptide and the CRISPR nickase-reverse transcriptase fusion polypeptide incorporated
5 substitutions encoded by editing template RNAs into human genes.

Example 7: Optimization of CRISPR Nuclease and CRISPR Nickase Fusions to a Reverse Transcriptase

This Example describes the engineering of additional fusions of the CRISPR nuclease
10 of SEQ ID NO: 1 or an H845A nickase variant of the CRISPR nuclease (SEQ ID NO: 32) to a reverse transcriptase polypeptide.

A plasmid library was designed comprising various combinations and orientations of NLS tags, flexible linkers, the CRISPR nuclease of SEQ ID NO: 1 or CRISPR nickase of SEQ
ID NO: 32, the variant reverse transcriptase of SEQ ID NO: 53, and a FLAG tag and
15 synthesized by GenScript. The sequences of the individual NLS, linker, and FLAG tag components are shown in **Table 15**. The resulting configurations are shown in **Table 16**.

Table 15. Sequences of Library Components

Component	Nucleotide Sequence	Amino Acid Sequence
NLS01	CCCAAGAAGAAGAGGAAGGTA (SEQ ID NO: 80)	PKKKRKV (SEQ ID NO: 81)
NLS02	CCTGCCGCTAAGAGAGTGAAGCTGGAC (SEQ ID NO: 82)	PAAKRVKLD (SEQ ID NO: 83)
NLS03	AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ ID NO: 6)	KRPAATKKAGQAKKKK (SEQ ID NO: 4)
NLS04	AAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTC (SEQ ID NO: 79)	KRTADGSEFESPKKKRKV (SEQ ID NO: 3)
NLS05	TGCTACGGCAGAAAGAAGCGGAGACAGAGAAGACGG (SEQ ID NO: 84)	CYGRKKRRQRRR (SEQ ID NO: 85)
NLS06	GACACCTGGACAGGCGTGGAGGCCC TGATCCGGATCCTGCAGCAGCTGCT GTTCATCCACTTCCGGATTGGATGCA GACACAGCAGAATCGGCATCATCCA GCAAAGACGCACCAGAAACGGCGC C (SEQ ID NO: 86)	DTWTGVEALIRILQQLFIHFRIGCRHSRIGIIQRRTRNGA (SEQ ID NO: 87)
NLS07	CGGAGAAAGGAAGAGGCCTTCAGACACAAG (SEQ ID NO: 88)	RRKEEAFRHK (SEQ ID NO: 89)
NLS08	CAGGAGGACATCATCAACAGATGCGAACAGTTCGTGGGCCCTCTGACCGTGAACGAGAAGCGGAGACTGAAGCTGATCATG (SEQ ID NO: 90)	QEDIINRCEQFVGPLTVNEKRRLKLIM (SEQ ID NO: 91)

Component	Nucleotide Sequence	Amino Acid Sequence
NLS09	GAGGGCGCCCCTCCCGCTAAGCGGG CCAGA (SEQ ID NO: 92)	EGAPPAKRAR (SEQ ID NO: 93)
NLS10	CCGGCAGCTAAGAAAAAGAACTG GAT (SEQ ID NO: 94)	PAAKKKKLD (SEQ ID NO: 95)
NLS11	TCCGGCGGAAGCTCTGGTGGCAGCA AGCGGACCGCCGACGGCTCTGAATT CGAGAGCCCTAAGAAGAAAAGAAA GGTGAGCGGAGGCTCTAGCGGCGGA AGC (SEQ ID NO: 96)	SGGSSGGSKRTADGSEFESPKKK RKVSGGSSGS (SEQ ID NO: 97)
NLS12	TCTGGCGGCTCAAAAAGAACCGCCG ACGGCAGCGAATTCGAGCCCAAGAA GAAGAGGAAAGTC (SEQ ID NO: 98)	SGGSKRTADGSEFEPKKKRKV (SEQ ID NO: 99)
Linker01	TCTGGAGGATCTAGCGGAGGATCCT CTGGCAGCGAGACACCAGGAACAA GCGAGTCAGCAACACCAGAGAGCA GTGGCGGCAGCAGCGGCGGCAGCA GC (SEQ ID NO: 7)	SGGSSGGSSGSETPGTSESATPES GGSSGGSS (SEQ ID NO: 8)
Linker02	AGCGCTGGAGGAGGTGGAAGCGGA GGAGGAGGAAGCGGAGGAGGAGGT AGCGGA (SEQ ID NO: 100)	SAGGGGSGGGGSGGGGSG (SEQ ID NO: 101)
Linker03	GGCACCGGCGGG (SEQ ID NO: 102)	GTGG (SEQ ID NO: 103)
Linker04	TCTGGCGGCTCA (SEQ ID NO: 104)	SGGS (SEQ ID NO: 105)
Linker05	TCTGGCGGCTCAAAAAGAACCGCCG ACGGCAGCGAATTCGAGTCTCCCAA GAAGAAGAGGAAAGTCGGCTCTGG CCCTGCCGCTAAGAGAGTGAAGCTG GAC (SEQ ID NO: 106)	SGGSKRTADGSEFESPKKKRKVG SGPAAKRVKLD (SEQ ID NO: 107)
Linker_1xGGG GS	GGCGGCGGCGGATCT (SEQ ID NO: 108)	GGGGS (SEQ ID NO: 109)
Linker_16xGG GGS	GGCGGAGGCGGCAGCGGCGGCGGC GGCAGCGGAGGCGGdoGCGGCAGCGG CGGCGGAGGCGGCGGCGGCGGAGG CAGCGGCGGCGGCGGCTCTGGCGGA GGCGGTTCCGGCGGCGGAGGCTCTG GCGGCGGCGGATCCGGCGGCGGAG GCAGCGGCGGCGGAGGCAGCGGCG GCGGCGGAAGCGGCGGCGGAGGCT CCGGCGGCGGCGGATCTGGCGGCGG CGGCAGCGGCGGAGGAGGCAGC (SEQ ID NO: 110)	GGGGSGGGSGGGGSGGGGSGG GGSGGGSGGGSGGGSGGGG SGGGSGGGSGGGSGGGSGG GGGSGGGSGGGGS (SEQ ID NO: 111)
Linker_8xGGG GS	GGTGGCGGAGGCTCTGGCGGCGGA GGCAGCGGAGGCGGAGGCAGCGGC GGCGGAGGCTCCGGCGGCGGCGGC AGCGGCGGCGGCGGCAGCGGCGGC GGAGGCAGCGGCGGCGGCGGATCT (SEQ ID NO: 112)	GGGGSGGGSGGGGSGGGGSGG GGSGGGSGGGSGGGGS (SEQ ID NO: 113)
Linker_4xSGGS	TCCGGCGGATCTAGCGGCGGCAGCA GCGGCGGATCTAGCGGCGGCAGC (SEQ ID NO: 114)	SGGSSGGSSGGSSGS (SEQ ID NO: 115)
Linker_4xGGG GS	GGCGGAGGCGGAAGCGGCGGCGGA GGCTCTGGCGGCGGAGGCTCCGGCG GCGGCGGCAGC (SEQ ID NO: 116)	GGGGSGGGSGGGGSGGGGS (SEQ ID NO: 117)

Component	Nucleotide Sequence	Amino Acid Sequence
Linker_4xEAA AK	GCCGAGGCCGCGCTAAGGAAGCCG CTGCCAAGGAAGCCGCGCTAAAGA GGCCGCGCCAAGGCC (SEQ ID NO: 118)	AEAAAKEAAAKEAAAKEAAAK A (SEQ ID NO: 119)
Linker_2xGGG GS_flank_T7P1 7	GGCGGGCGGCTCCGGCGGCGGC GGCAGCAAGAACGAGAGCAGCACC AACGCCACAAACACCAAGCAGTGG CGGGACGAGACAAAGGGCTTCAGA GATGAAGCTAAAAGATTCAAGAATA CCGCCGGCGGCGGAGGAGGCTCTGG CGGAGGAGGCAGC (SEQ ID NO: 120)	GGGSGGGGSKNESSTNATNTK QWRDETKGFRDEAKRFKNTAGG GGGSGGGGS (SEQ ID NO: 121)
Linker_2xGGG GS_flank_BPN LS	GGCGGAGGCGGCTCCGGCGGCGGA GGCAGCAAGCGGACCGCCGACGGC TCTGAATTCGAGAGCCCTAAGAAGA AAAGAAAGGTGGGCGGCGGAGGCA GCGGCGGCGGCGGAAGC (SEQ ID NO: 122)	GGGSGGGGSKRTADGSEFESPK KKRKVGGGSGGGGS (SEQ ID NO: 123)
Linker_2xGGG GS_flank_XTE N	GGCGGGCGGCGGAGCGGCGGCGGA GGCAGCAGCGGCTCTGAAACCCCTG GCACCAGCGAGTCCGCCACACCCGA GAGCGGCGGCGGAGGATCTGGCGG AGGCGGCAGC (SEQ ID NO: 124)	GGGSGGGGSSGSETPGTSESATP ESGGGSGGGGS (SEQ ID NO: 125)
Linker_2xEAA AK_flank_XTE N	GCCGAGGCCGCTGCTAAAGAAGCCG CCGCCAAGTCCGGCAGCGAGACACC CGGCACCAGCGAGAGCGCCACCCCT GAGTCTGAAGCCGCGCCAAGGAA GCCGCTGCCAAGGCC (SEQ ID NO: 126)	AEAAAKEAAAKSGSETPGTSESA TPESEAAAKEAAKA (SEQ ID NO: 127)
Linker_2xSGGS _flank_BPNLS	TCCGGCGGATCTAGCGGCGGCAGCA AGCGGACCGCCGACGGCTCTGAATT CGAGAGCCCTAAGAAGAAAAGAAA GGTGTCCGGCGGAAGCAGCGGCGG CAGC (SEQ ID NO: 128)	SGGSSGGSKRTADGSEFESPKKK RKVSGGSSGGS (SEQ ID NO: 129)
Linker_2xSGGS _flank_T7P17	AGCGGAGGCTCTAGCGGCGGCAGCA AGAACGAGAGCTCCACCAACGCCAC AAACACCAAGCAGTGGCGGGACGA GACAAAGGGCTTCAGAGATGAAGCT AAAAGATTCAAGAATACCGCCGGCA GCGGCGGATCTAGCGGCGGCTCC (SEQ ID NO:279)	SGGSSGGSKNESSTNATNTKQWR DETKGFRDEAKRFKNTAGSGGSS GGS (SEQ ID NO: 131)
Linker_2xPAPA P_flank_XTEN	GCCCCGCCCCCTGCCCCCTCTGCCCC TGCTCCCAGCGGCTCTGAGACACCT GGCACCAGCGAGTCCGCCACCCAG AAAGCCCCGCCCCGCTCCACCTGC CCCTGCCCCCTGCC (SEQ ID NO: 132)	APAPAPPAPAPSGSETPGTSESATP ESPAPAPPAPAPA (SEQ ID NO: 133)
Linker_2xPAPA P_flank_BPNLS	GCCCCCTGCTCCAGCCCCCTCTGCCCC CGCCCCTAAGCGGACCGCCGACGGC TCTGAATTCGAGAGCCCCAAGAAGA AAAGAAAGGTGCCCCGCCCTGCCCC ACCTGCTCCCGCCCCCTGCC (SEQ ID NO: 134)	APAPAPPAPAPKRTADGSEFESPK KKRKVPAPAPPAPAPA (SEQ ID NO: 135)
Linker_2xEAA AK_flank_BPN LS	GCCGAGGCCGCGCCAAGGAAGCC GCCGCCAAGAAGCGGACCGCCGAC GGCAGCGAGTTCGAGTCTCCTAAGA	AEAAAKEAAAKKRTADGSEFESF KKRKVEAAAKEAAKA (SEQ ID NO: 137)

Component	Nucleotide Sequence	Amino Acid Sequence
	AGAAAAGAAAGGTGGAAGCCGCTG CTAAAGAGGCCGCTGCCAAGGCC (SEQ ID NO: 136)	
Linker_AEAAA KGGGSGGG GSEAAKA	GCTGAAGCCGCCCAAGGGCGGC GGCGGATCTGGCGGAGGCGGCAGCG AGGCCGCTGCCAAGGCC (SEQ ID NO: 138)	AEAAAKGGGSGGGGSEAAKA (SEQ ID NO: 139)
Linker_APAPA PGGGSGGG GSPAPAPA	GCCCCTGCCCTGCCCGGCGGCG GCGGATCTGGCGGAGGCGGCAGCCC CGCCCCTGCTCCAGCC (SEQ ID NO: 140)	APAPAPGGGSGGGGSPAPAPA (SEQ ID NO: 141)
Linker_2xGGG GS	GGCGGCGGAGGCAGCGGCGGCGGC GGATCT (SEQ ID NO: 142)	GGGSGGGGS (SEQ ID NO: 143)
FLAG	GATTACAAAGACGATGACGATAAGG ATTACAAAGACGATGACGATAAGGA TTACAAAGACGATGACGATAAG (SEQ ID NO: 144)	DYKDDDDKDYKDDDDKDYKDD DDK (SEQ ID NO: 145)

Table 16. Configurations of Library Components

Config	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8	Part 9	Part 10
1	NLS	NLS	Nuc /Nick	Linker	RT	FLAG				
2	Nuc /Nick	Linker	RT	FLAG	NLS	NLS				
3	NLS	Nuc /Nick	Linker	RT	FLAG	NLS	Linker	NLS		
4	Nuc /Nick	Linker	RT	NLS	FLAG	NLS	Linker	NLS		
5	NLS	Linker	Nuc /Nick	Linker	RT	FLAG	Linker	NLS		
6	NLS	Linker	Nuc /Nick	Linker	RT	FLAG	Linker	NLS	NLS	NLS
7	NLS	Linker	RT	Linker	Nuc /Nick	FLAG	Linker	NLS	NLS	NLS
8	RT	Linker	Nuc /Nick	NLS	FLAG	NLS	Linker	NLS		
9	NLS	Nuc /Nick	Linker	NLS	Linker	RT	FLAG			
10	NLS	Nuc /Nick	Linker	NLS	Linker	RT	FLAG	Linker	NLS	
11	NLS	RT	Linker	NLS	Linker	Nuc /Nick	FLAG			
12	NLS	RT	Linker	NLS	Linker	Nuc/N ick	FLAG	Linker	NLS	
13	NLS	Nuc /Nick	Linker	RT	NLS	FLAG				
14	NLS	Nuc	Linker	RT	Linker	Linker	NLS	FLAG		

		/Nick								
--	--	-------	--	--	--	--	--	--	--	--

The plasmid library was screened in HEK293T cells using the lipid-based transient transfection method described in Example 5. Each well of the 96-well plate was transfected with a plasmid encoding a unique CRISPR nuclease-reverse transcriptase fusion polypeptide or CRISRP nickase-reverse transcriptase fusion polypeptide. The editing template RNA sequence of SEQ ID NO: 70, designed to introduce a 6-nucleotide substitution into an EMX1_T2 target, was also transfected into each well. Quantification of edits was performed as described in Example 6.

FIG. 4A and FIG. 4B show the CRISPR nuclease-reverse transcriptase fusion polypeptides and CRISPR nickase-reverse transcriptase fusion polypeptides, respectively, that installed the highest percentage of edits encoded by the editing template RNA of all tested constructs. Results are the average of 2 technical replicates. The dotted line depicts the percentage of reads comprising the 6-nucleotide substitution installed by the control CRISPR nuclease-reverse transcriptase fusion polypeptide (FIG. 4A) or CRISPR nickase-reverse transcriptase fusion polypeptide (FIG. 4B). As shown in FIG. 4A and FIG. 4B, several fusion constructs resulted in increased installation of the 6-nucleotide substitution relative to the control CRISPR nuclease-reverse transcriptase and CRISPR nickase-reverse transcriptase fusion polypeptides. The sequences of the top-performing CRISPR nickase-reverse transcriptase fusion polypeptides are shown in Table 17.

Table 17. Configurations and Sequences of Top-Performing Fusion Constructs

Construct Name	Protein Configuration	Complete AA Sequence
FC1	CRISPR nickase-Linker01-Variant MMLV-NLS01-FLAG-NLS10-Linker04-NLS03	MNKPYSIGLDIGTNSVGWSII TDDYKVPAAKMRVLGNTDKEYIKKNLIG ALLFDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSF HRLDSFLVEEDKRGSKYP IFATLQEEKDYHEKFPTIYHLRKEADKQE KADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFD FENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKL IVGNQADFKKHFNLEDKTPIQFAKDSYDEDLNLLGQIGDEFADLFSA KKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKAS LPEKYQEIFADSSKDGAGYIEGKTNQEAIFYKYSKLLTKQEDSENFLE KIKNEFLRKQRTFDNGSIPHQVHLTELKAI IRRQSEYYPFLENQDRI EKILTRIPYYIGPLAREKSDFAWMTRKTDD SIRPWNFEDLVDKEKSAE AFIHRMTNDFYLPPEEKVLPKHS LIYEKFTVYNELTKVRYKNEQGETYF FDSNICKQEIFDGVFKEHRKVSKKLLDFLAKEEYEFRIVDVIGLDKENK AFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMICKR LENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLI DDGRSNRNFQMLINDDGLSFKSII SKAQAGSHSDNLKEVVGELAGSPAI KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQRYKL LDDGVKNLASDLNGNILEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDAI IPQAFIKDDSIDNRVLVSSAKNRGKSDDVP SLEIVK DCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQ

Construct Name	Protein Configuration	Complete AA Sequence
		ITKHVARILDERFNNELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKI REVNDYHHAHDAYLNAVAKAAILTKYPQLEPEFVYGLYRRKKLSKIVRE DKEEKYSEATRKMFYNSLMNMFKRVRRLADGSIVVRPVIETGRYMGKT AWDKKKHFATVRKVLSTYPQVNIKKTEIQTGGFSKESILAHGNSDKLIP RKT KD IYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGIT IMERSKFENPSAFLESKGYLNIRDDKMLIPKYSLFELENGRRRLAS AGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSY FDDILQLINDFSKRVLADANLEKINKLYQDNKENISVDELANNIINLF TFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSI TGLYETRIDL SKLGEDSGSSGGSSGSETPGTSESATPESSGGSSGSSSTLNIEDEYRL HETSKPEVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPV SIKQYPMSEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTND YRPVQDLREVNRKVEDIHPVTPNPYNLLSGLPPSHQWYTVLDDKDAFFC LRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTL FNEALHRD LADFRIQHPDLILQYVDDLLLAATSELDCCQGTALLQLNGLRAS AKKAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKTPRQLREFL GKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQKAYQEI KQALLT APALGLPDLTKPFELFVDEKQGYAKGVLTQKLGWRRPVAYLSKKLDPV AAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVNGPDRW LSNARMTHYQALLD TDRVQFGPVVALNPATLLPLPEEGLQHNCDLILA EAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTETETEWIWA KALPAGTSAQRAELIALTQALKMAEGKKLVNYTDSRYAFATAHIHGEIY RRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSABAR GNRMADQAARKAAITETPTDSTLLIENSSPKRPAATKKAGQAKKKDYK DDDDKDYKDDDDKDYKDDDDKPAAKKKKLDSSGSKRPAATKKAGQAKKK K (SEQ ID NO: 146)
FC2	NLS03-Linker02-CRISPR nickase-Linker01-Variant MMLV-FLAG-Linker02-NLS03	MKRPAAATKKAGQAKKKKSAGGGGGGGGGGGGGGNPKPYSIGLDIGTNS VGWSIITDDYKVP AKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRRL KRTARRRYTRRRNRILYLQETFAEEMSKVDDSFHRLDESF LVEEDKRG SKYPIFATLQEEKDYHEKFPTIYHLRKEADKQEKADRLRIYLALAHII KFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLKSNQVDEVA ILTDKISAKKDRILAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLNLGQIGDEFADLFSAAKKLYDSVLLSGILTV IDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKD GYAGYIEGTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLNRKQRTFD NGSIPHQVHLTELKAIIRQSEYYPFLKENQDRIEKILTRIPYYIGPL AREKSDFAWMTRKTDDSI RPWNFEDLVDEKESAEAFIHRMTNNDFYLPE EKVLPKHSIIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFK EHRKVSKKLLDLFLAKEYEFRIVDVIGLDKENKAFNASLGTYHDLLEKI LDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKK LYRRHYTGWGRLSAKLINGIRDKESQKTIIDYLIIDGRSNRNFQMQLIND DGLSFKSIIISKAQAGSHSDNLKEVVGELAGSPAIIKKGILQSLKIVDELV KVMGYEPEQIVVEMARENQTTNQGRNRSRQRYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDNLSQYDIDAIIP QAFIKDSDIDNRVLVSSAKNRGKSDDVPSLEIVKDCVKVFWKKLLDAKLM SQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNN ELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLN AVVAKAAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATRKMF YNSLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKVL STYPQVNIKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGG FDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFENPSAF LESKGYLNIRDDKMLIPKYSLFELENGRRRLASAGELQKGNELALPTQ FMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKR VLADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNSTLIHQSI TGLYETRIDL SKLGEDSGSSGGSS GSETPGTSESATPESSGGSSGSSSTLNIEDEYRLHETSKPEVSLGSTW LSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSEARLG IKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRKVE

Construct Name	Protein Configuration	Complete AA Sequence
		DIHPTVPNPNLLSGLPPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ YVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAE MAAPLYPLTKPGTLFNWGPDQQKAYQEIQAALLTAPALGLPDLTKPFEL FVDEKQGYAKGVLTKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLD TDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLP DADHTWYTDGSSLLQEGQQRKAGAAVTTETETEWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKN KDEILALLKALFLPKRLSIHCPGHQKGHSAEARGNRMADQAARKAAIT ETPDTSTLLIENSSPDYKDDDDKDYKDDDDKDYKDDDDKSAGGGGSGGG GSGGGSGGKRPAATKKAGQAKKKK (SEQ ID NO: 147)
FC3	NLS03-Linker02-CRISPR nickase-Linker01-Variant MMLV-FLAG-Linker02-NLS04	MKRPAATKKAGQAKKKKSAGGGGSGGGGSGGGSGGNKPYSIGLDIGTNS VGWSIITDDYKVPAAKMRVLGNTDKEYIKKNLIGALLFDGNGTAADRRRL KRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDSFLVEEDKRG SKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRLIYLALAHII KFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDITFENNHLSSQNVDEVA ILTDKISKSAKKDRILAQYPNQKSTGIFAELKLIIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTV IDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKD GYAGYIEGKTNQEAFYKYLKLLTKQEDSENFLEKIKNEDFLRKQRTFD NGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTRIPYYIGPL AREKSDFAWMTRKTDDSIKRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPE EKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFK EHRKVSKKLLDLFLAKEYEFRIVDVIGLDKENKAFNASLGTYHDLKI LDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLINYKDLFTSESQKK LYRRHYTGWGRLSAKLINGIRDKESQKTIIDYLDIDGRSNRNFQMQLIND DGLSFKSIISKAQAGSHSDNLKEVVGELAGSPAIIKKGILQSLKIVDELV KVMGYEPEQIVVEMARENQTTNQGRNRSRQRYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDNLSQYDIDAIIP QAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCCKVFWKKLLDAKLM SQRKYDNLTKAERGGTSDDKARFIQRLVETRQITKHVARILDERFNN ELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLN AVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEYSEATRKMFF YSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKVL SYPQVNIKKTEIQTGGFSKESILAHGNSDKLIPRKTIDYLDPKKYGG FDSPIVAYSVLVADIKKGAKKLKTVTELLGITIMERSFEKPNSAFL ESKGYLNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQ FMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRV ILADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFD KIVDRKRYTSTKEVLNSTLIHQSIITGLYETRIDLKLGEDSGGSSGGSS GSETPGTSESATPESSGGSSGSSSTLNIEDEYRLHETSKEPDVSLGSTW LSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIIKQYPMSEQEARLGI KPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNRKVE DIHPTVPNPNLLSGLPPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQ YVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVYLG YLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAE MAAPLYPLTKPGTLFNWGPDQQKAYQEIQAALLTAPALGLPDLTKPFEL FVDEKQGYAKGVLTKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLD TDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLP DADHTWYTDGSSLLQEGQQRKAGAAVTTETETEWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKN KDEILALLKALFLPKRLSIHCPGHQKGHSAEARGNRMADQAARKAAIT ETPDTSTLLIENSSPDYKDDDDKDYKDDDDKDYKDDDDKSAGGGGSGGG GSGGGSGGKRTADGSEFESPKKKRKV (SEQ ID NO: 148)

Construct Name	Protein Configuration	Complete AA Sequence
FC4	NLS04-Linker02-Variant MMLV-Linker01-CRISPR nickase-FLAG-Linker02-NLS04-NLS03-NLS10	MKRTADGSEFESPKKKRKVSAGGGGSGGGGSGGGGSGTLNIEDEYRLHE TSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS KQYPMSEARLGIKPHIQRLLDQILVPCQSPWNTPLLVPKKPGTNDYR PVQDLREVNRKVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDADFCLR LHPTSQPLFAFEWRDPEMGISGQLTWTRLPGGFKNSPTLFNEALHRDLA DFRIQHPDLILLQYVDDLLAATSELDCQGGTRALLQTLGNLGYRASAK KAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIQALLTAP ALGLPDLTKPFELFVDEKQGYAKGVLTKLGPWRRPVAYLSKKLDPVAA GWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS NARMTHYQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETVIWAKA LPAGTSAQRAELIALTQALKMAEGKKNVYTDSTRYAFATAHIGHEYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGN RMAQQAARKAAITETPDTSTLLIENSSPSGGSSGSSGSGTSESAT PESSGGSSGGSSNKPYSIGLDIGTNSVGWSIIITDDYKVPAKMKMRVLGNT DKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQETFA EEMSKVDDSSFFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIY HLRRELADKQEKADLRILIYALAHIIKFRGHFLIEDDRFDVNDLIQKQ YQAFLEIFDITTFENHLLSQNVDEAILTDKISKSAKKDRILAQYFNQK STGIFAEFLKLIVGNQADFKKHFNLEDKTPQLQFAKDSYDEDLENLLGQI GDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMTQRYDEHRED LKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAIFYKLSKLL TKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRQYSEY YPFLKENQDRIEKILTRIPYYIGPLAREKSDFAWMTRKTDSSIRPWNF EDLVDKEKSAEAFIHRMTNNDIFYLPEEKVLPKHSLEYEFTVYNELTKV RYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKLLDFLAKEYEEFRI VDVIGLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQTLT LFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDK ESQKTILDYLIIDGRSNRNFMLINDDGLSFKSIIISKAQAGSHSDNLKE VVGELAGSPAIAKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQ GRRNSRQRYKLLDDGVKNLASDLNGNIIKEYPTDNQALQNERLFLYLLQ NGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDDSIDNRVLVSSAKNRGK SDDVP SLEIVKDCVFWKKLLDAKLMSQRKYDNLTKAERGGTSSDKAR FIQRQLVETRQITKHVARILDERFNNELDSKGRRIKRVKIVTLKSNLVS NFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVGLY RRKKLSKIVREDKEEKYSEATRKMFFYSNLMNMFKRVRRLADGSIVVRP VIETGRYMGKTAWDKKKHFAVTRKVLSTYPQVNIIVKKEIQTGGFSKESI LAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVADIKKGAQK LKTVTELLGITIMERSKFEKNPSAFLESKGYLNIRDCLKMILPKYSLFE LENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKKGPEETEK KQEFVNQHVSYFDDILQLINDFSKRVLADANLEKINKLYQDNKENISV DELANNIINLFTTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQ ITGLYETRIDLKLGEDDYKDDDDKDYKDDDDKDYKDDDDKSAGGGGSG GGGSGGGSGKRTADGSEFESPKKKRKVKRPAATKKAGQAKKKKPAKK KKLD (SEQ ID NO: 149)
FC5	NLS04-Linker02-Variant MMLV-Linker01-CRISPR nickase-FLAG-Linker02-NLS04-NLS03-NLS04	MKRTADGSEFESPKKKRKVSAGGGGSGGGGSGGGGSGTLNIEDEYRLHE TSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS KQYPMSEARLGIKPHIQRLLDQILVPCQSPWNTPLLVPKKPGTNDYR PVQDLREVNRKVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDADFCLR LHPTSQPLFAFEWRDPEMGISGQLTWTRLPGGFKNSPTLFNEALHRDLA DFRIQHPDLILLQYVDDLLAATSELDCQGGTRALLQTLGNLGYRASAK KAQICQKQVKYLGILLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIQALLTAP ALGLPDLTKPFELFVDEKQGYAKGVLTKLGPWRRPVAYLSKKLDPVAA GWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLS NARMTHYQALLDTRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETETVIWAKA LPAGTSAQRAELIALTQALKMAEGKKNVYTDSTRYAFATAHIGHEYRR

Construct Name	Protein Configuration	Complete AA Sequence
		RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGN RMADQAARKAAITETPDTSTLLIENSSPSGGSSGGSSGSETPGTSESAT PESGGSSGGSSNKPYISIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNT DKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIY HLRKELADKQEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQ YQAFLEIFDITTFENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPNQK STGIFAEFLKLIVGNQADFKKHFNLEDKTPLOFAKDSYDEDLENLLGQI GDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHRED LKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAIFYKLSKLL TKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEY YPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDDSIKPNWF EDLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHSIYIEKFTVYNELTKV RYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKLLDLFLAKEEYEFRI VDVIGLDKENKAFNASLGTYHDLEKILDKDFLDNPDNIESILEDIVQTLT LFEDREMIKKRLNYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDK ESQKTILDYLIDGRSNRNFQMQLINDDGLSFKSIISKAQAGSHSDNLKE VVGELAGSPAIAKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQ GRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYLLQ NGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDDSIDNRLVSSAKNRGK SDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILDERFNNELDKSGRRIRKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHHAHDAYLNAVAKAILTKYPQLEPEFVYGLY RRKKLSKIVREDKEEKYSEATRKMFYFYNLMNMFKRVRRLADGSIIVVRP VIETGRYMGKTAWDKKKHFAVRKVLSPQVNIIVKKEIQTGGSFSES LAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVVADIKKGAQK LKTVTTELLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMILPKYSLFE LENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKGPETIEK KQEFVNQHVSYFDDILQLINDFSKRVIADANLEKINKLYQDNKENISV DELANNIINLFTFTSLGAPAAFKFDFKIVDRKRYTSTKEVLNSTLIHQ ITGLYETRIDLKSLIGEDDYKDDDDKDYKDDDDKDYKDDDDKSAAGGGGSG GGSGGGGGSGKRTADGSEFESPKKKRKVKRPAATKKAGQAKKKKRTAD GSEFESPKKKRKV (SEQ ID NO: 150)
FC6	NLS04-CRISPR nickase-Linker01- NLS03-Linker01- Variant MMLV- FLAG	MKRTADGSEFESPKKKRKVNKPYSIGLDIGTNSVGWSIITDDYKVPAKK MRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRIL YLQEIFAEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYH EKFTIYHLRKELADKQEKADLRLIYLALAHIIKFRGHFLIEDDRFDVR NTDIQKQYQAFLEIFDITTFENNHLLSQNVDVEAILTDKISKSAKKDRIL AQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLEDKTPLOFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQR YDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAIFY KLSKLLTKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHLTELKAI IRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDD SIRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHSIYIEKFTV YNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKLLDLFLAK EYEEFRIVDVIGLDKENKAFNASLGTYHDLEKILDKDFLDNPDNIESILE DIVQTLTLTFEDREMIKKRLNYKDLFTESQLKKLYRRHYTGWGRLSAKL INGIRDKESQKTILDYLIDGRSNRNFQMQLINDDGLSFKSIISKAQAGS HSDNLKEVVVGELAGSPAIAKKGILQSLKIVDELVKVMGYEPEQIVVEMAR ENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNER LFLYLLQNGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDDSIDNRLVVS SAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGL TSDDKARFIQRQLVETRQITKHVARILDERFNNELDKSGRRIRKVKIVT LKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVAKAILTKYPQLEP EFVYGLYRRKKLSKIVREDKEEKYSEATRKMFYFYNLMNMFKRVRRLAD GSIIVVRPVIETGRYMGKTAWDKKKHFAVRKVLSPQVNIIVKKEIQTG GFSKESILAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVVADI KKGAQKLKTVTTELLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMIL PKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKG

Construct Name	Protein Configuration	Complete AA Sequence
		KPEEIEKKQEFVNQHVS YFDD ILQLINDFSKRVLADANLEKINKLYQD NKENISVDELANNIINLFTFTSLGAPAAFKFDDKIVDRKRYTSTKEVLN STLIHQSIITGLYETRIDL SKLGEDSGGSSGGSSGSETPGTSESATPESS GGSSGGSSKRPAATKKAGQAKKKKSGGSSGGSSGSETPGTSESATPESS GGSSGGSSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGMGL AVRQAPLI IPLKATSTPVSIKQYPMSQEARLG I KPHIQRLDQGI L VPC QSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVEDI HPTV PNPYNLLSGL PPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRL PQGFKNSTL FNEALHRDLADFR I QHPDLILLQYVDDLLLAATSELD CQ QGTRALLQTLGNLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDPVAGWPPCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQF GPVVALNPAT LLPLPEEGLQHNC LD ILAEAHGTRPDLTDQPLPDADHTWYD TGSLLSGL GQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKNLV YTD SRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL KALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPD TSTLLIENSSPD YKDDDDKDYKDDDDKDYKDDDDK (SEQ ID NO: 151)
FC7	NLS04-CRISPR nickase-Linker01- NLS03-Linker01- Variant MMLV- FLAG-Linker04- NLS01	MKRTADGSEFESPKKKRKVNKPYSIGLDIGTNSVGSII TDDYKVPACK MRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRIL YLQEIFAEEMSKVDDSFHRL EDSFLVEEDKRGSKYPIFATLQEEKD YH EKFPTIYHLRKE LADKQEKADLR LIYLALAH I IKFRGHFLIEDDRFDVR NTDIQKQYQAFLEIFD TTFENNHL LSNQNVDEAILTDKISKS AKKDRIL AQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLEDKTP LQFAKDSYDEDL ENLLGQIGDEFADLFSAAKLYDSVLLSGILTVIDLSTKAPLSASMIQR YDEHREDLKLQKQFVKASLPEKYQEIFADSSKDGYAGYIEGKTNEAFY KYL SKLLTKQEDSENFL EKIKNEDFLRKQRTFDNGSIPHQVHLTELKAI IRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDD SIRPWNFEDLV DKEKSAAEFIHRMTNND FYLP EEKVLPKHS LIYEKFTV YNELTKVRYKNEQG ETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAK EYEEFRIVDVI GLDKENKAFNASLGT YHDL EKILD KDFLGNPDNESILE DIVQTLTLTFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKL INGIRDKESQKTILDY LIDGRSNRNF MQ LINDDGLSFKS I ISKAQAGS HSDNLKEVVGELAGSPA IKKGILQSLKIVDELVKVMGYEPQI VVEMAR ENQTTNQGRNSRQRYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNER LFLY LQNGRDMYTGEALDIDNLSQYD IDAIIPQAF IKDSDIDNRVLVS SAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGL TSDDKARFTQRQLVETRQITKHVARILDERFNNELDSKGRRI R KVKIVT LKS NLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEP EFVYGLYRRKKLSKIVREDKEEKYSEATRK MFFYSNLMNMFKRVRRLAD GSIVVRPVIETGRYMGKTAWDKKKHFATVRKVL SYPQVNI VKKTEIQTG GFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFDSPIVAYSVLVVADI KKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMIL PKYSLFELENGRRRL LASAGELQKGNELALP TQFMKFLYLASRYNESKG KPEEIEKKQEFVNQHVS YFDD ILQLINDFSKRVLADANLEKINKLYQD NKENISVDELANNIINLFTFTSLGAPAAFKFDDKIVDRKRYTSTKEVLN STLIHQSIITGLYETRIDL SKLGEDSGGSSGGSSGSETPGTSESATPESS GGSSGGSSKRPAATKKAGQAKKKKSGGSSGGSSGSETPGTSESATPESS GGSSGGSSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGMGL AVRQAPLI IPLKATSTPVSIKQYPMSQEARLG I KPHIQRLDQGI L VPC QSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVEDI HPTV PNPYNLLSGL PPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDP EMGISGQLTWTRL PQGFKNSTL FNEALHRDLADFR I QHPDLILLQYVDDLLLAATSELD CQ QGTRALLQTLGNLGYRASAKKAQICQKQVKYLG YLLKEGQRWLTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDQQKAYQEI KQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDPVAGWPPCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQF GPVVALNPAT

Construct Name	Protein Configuration	Complete AA Sequence
		LLPLPEEGLQHNCILAEAHGTRPDLTDQFLPDADHTWYTDGSSLLQE GQRKAGAAVTTETETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLNV YTD SRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPD YKDDDDKDYKDDDDKDYKDDDDKSGGSPKKKKRV (SEQ ID NO: 152)
FC8	NLS04-CRISPR nickase-Linker01- NLS03-Linker01- Variant MMLV- FLAG-Linker04- NLS10	MKRTADGSEFESPKKKKRVNKPYSIGLDIGTNSVGWSIITDDYKVPAAK MRVLGNTDKEYIKKNLIGALLFDGGNTAADRLKRTARRRYTRRRNRIL YLQEIFAEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYH EKFTIYHLRKELADQEKADLRLIYLALAHIIKFRGHFLIEDDRFDVR NTDIQKQYQAFLEIFDTTFENNHLSSQNVDVEAILTDKISKSAKKDRIL AQYPNQKSTGIFAELFKLIVGNQADFKKHFNLEDKTPQLQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQR YDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAFY KYLKLLTKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHLTELKAI IRROSEYYPFLKENQDRIEKILTFRIPIYYIGPLAREKSDFAWMTRKTD SIRPNWFEDLVDEKSAEAFIHRMTNNDLYLPEEKVLPKHSIIYKFTV YNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKLLDFLAK EYEEFRIVDVIGLDKENKAFNASLGTYHDEKILDKDFLNDPNDSILE DIVQTLTLTFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKL INGIRDKESQKTIIDYLDIDGRSNRNFMLINDDGLSFKSIIKSAQAGS HSDNLKEVVGELAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMAR ENQTTNQGRNRNRQRYKLLDDGVKNLASDLNGNILEYPTDNQALQNER LFLYYLQNGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDDSIDNRVLVS SAKNRGKSDDVPSLEIVKDCVFWKKLLDAKLMSQRKYDNLTKAERGG TSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKGRRIKVKIVT LKSNLVSFNRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYQLEP EFVYGLYRRKKLSKIVREDKEEYSEATRMFFYSNLMNMFKRVRVRLAD GSIVVRPVIETGRYMGKTAWDKKKHFAFTRKVLVSYPQVNIYVKKTEIQTG GFSKESILAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVADI KKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGYLNRDDDKMIL PKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKG KPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVLADANLEKINKILYQD NKENISVDELANNIINLFTFTSLGAPAAFKFDDKIYDRKRYTSTKEVLN STLIHQSIITGLYETRIDLKSLGEDSGSSGGSSGSETPGTSESATPES GGSSGGSSKRPAATKAGQAKKKKSGSSGGSSGSETPGTSESATPES GGSSGGSSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIIKQYPMSEARLGIKPHIQRLLDQGIIVPC QSPWNTPLLPVKKPGTNDYRPVQDLREVNRVEDIHPVTPNPYNLLSGL PPSHQWYTVLIDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISSQLTWTRL PQGFKNSTPLFNEALHRDLADFRIQHPDLILLQYVDDLLAATSELDCQ QGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYYLKEGQRLWTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDQOKAYQEIQAALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTOK LGPWRRPVAYLSKKLDPVAAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFQGPVVALNPAT LLPLPEEGLQHNCILAEAHGTRPDLTDQFLPDADHTWYTDGSSLLQE GQRKAGAAVTTETETEVIAKALPAGTSAQRAELIALTQALKMAEGKKLNV YTD SRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPD YKDDDDKDYKDDDDKDYKDDDDKSGGSPAIAKKKKLD (SEQ ID NO: 153)
FC9	NLS04-CRISPR nickase-Linker01- NLS03-Linker01- Variant MMLV- FLAG-Linker04- NLS03	MKRTADGSEFESPKKKKRVNKPYSIGLDIGTNSVGWSIITDDYKVPAAK MRVLGNTDKEYIKKNLIGALLFDGGNTAADRLKRTARRRYTRRRNRIL YLQEIFAEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYH EKFTIYHLRKELADQEKADLRLIYLALAHIIKFRGHFLIEDDRFDVR NTDIQKQYQAFLEIFDTTFENNHLSSQNVDVEAILTDKISKSAKKDRIL AQYPNQKSTGIFAELFKLIVGNQADFKKHFNLEDKTPQLQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQR YDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAFY

Construct Name	Protein Configuration	Complete AA Sequence
		KYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAI IRRQSEYYFPFLKENQDRIEKILTFRIPIYYIGPLAREKSDFAWMTRKTDD SIRPWNFEDLVDKESAEAFIHRMTNNDFYLPPEEKVLPKHSLIYEKFTV YNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAK EYEEFRIVDVIGLDENKAFNASLGTYHDLKILDKDFLDNPDNSESILE DIVQTLTLFEDREMIKKRLKENYKDLFTESQLKKLYRRHYTGWGRLSAKL INGIRDKESQKTIILDYLIIDGRSNRNFQMLINDDGLSFKSIISKAQAGS HSDNLKEVVGELAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIIVEMAR ENQTTNQGRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNER LFLYYLQNGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDSDIDNRVLVS SAKNRGKSDDVPSLEIVKDCVFWKLLDAKLMSQRKYDNLTKAERGLL TSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKGRRIKVKIVT LKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKALLTKYPQLEP EFVYGLYRRKKLSKIVREDKEEKYSEATRMKFFYSNLMNMFKRVRVRLAD GSIIVVRPVIETGRYMGKTAWDKKKHFAFATVRKVLSTYQVNVIVKKTETQG GFSKESILAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVVADI KKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMIL PKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKG KPEIEIKKQEFVNQHVSYFDDILQLINDFSKRVLADANLEKINKLEYQD NKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLN STLIHQSIITGLYETRIDLKSLGEDSGSSGGSSGGSETPGTSESATPES GGSSGGSSSKRPAATKAGQAKKKKSGSSGGSSGGSETPGTSESATPES GGSSGGSSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIIKQYPMSEARLGKPHIQRLDQGIIVPC QSPWNTPLLPVKKPGTNDYRVPVQDLREVNRVEDIHPTVNPYNLLSGL PPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPPEMGSQGLTWTRL PQGFKNSTPLFNEALHRDLADFRIQHPDLILLQYVDDLLAATSELDCC QGTALLQTLGNLGYRASAKKAQICQKQVKYLGILLKEGQRLWTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDQKQAYQEIQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDPVAAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQE GQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNV YTDSTRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPD YKDDDDKDYKDDDDKDYKDDDDKSGGSKRPAATKAGQAKKKK (SEQ ID NO: 154)
FC10	NLS04-Variant MMLV-Linker01- NLS03-Linker01- CRISPR nickase- FLAG-Linker04- NLS03	MKRTADGSEFESPKKRKVTNLNIEDEYRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIIKQYPMSEARLGKPHIQ RLLDQGIIVPCQSPWNTPLLPVKKPGTNDYRVPVQDLREVNRVEDIHPT VNPYNLLSGLPPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPPEM GISQGLTWTRLPPQGFKNSTPLFNEALHRDLADFRIQHPDLILLQYVDDL LLAATSELDCCQGTALLQTLGNLGYRASAKKAQICQKQVKYLGILLKE GQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPL YPLTKPGTLFNWGPDQKQAYQEIQALLTAPALGLPDLTKPFELFVDEK QGYAKGVLTQKLGWRRPVAYLSKKLDPVAAAGWPPCLRMVAAIAVLTKD AGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQA LKMAEGKKLNVYTDSTRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEIL ALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDT STLLIENSSPSGGSSGGSSGGSETPGTSESATPESGGSSGGSSSKRPAAT KKAGQAKKKKSGSSGGSSGGSETPGTSESATPESGGSSGGSSSKRPYSI GLDIGTNSVGWSIITDDYKVPAAKMRVLGNTDKEYIKKNLIGALLFDGG NTAADRRLLKRTARRRYTRRRNRIYLYLQEIFAEEMSKVDDSFHRLSDSF LVEEDKRGSKYPIFATLQEEKDYEKFPITYHLRRELADKQEKADLRILI YLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHL SQNVDEAILTDKISKSAKKDRILAQYPNQKSTGIFAELKLIVGNQAD

Construct Name	Protein Configuration	Complete AA Sequence
		FKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSV LLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQE IFADSSKDGAGYIEGKTNQEAIFYKLSKLLTKQEDSENFLKIKNEDF LRKQRTFDNGSIPHQVHTELKAIIRRQSEYYPFLKENQDRIEKILTFR IPYYIGPLAREKSDFAWMTRKTDDSIKPNWFEDLVDKEKSAAEAFIHRMT NNDFYLPPEKVLPHKSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQ EIFDGVFKEHRKVSKKKLLDFLAKEYEFEFRIVDVIKLDKENKAFNASLG TYHDLKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLKNYKDL FTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTIIDYLLIDGGRSNR NFMQLINDDGLSFKSIISKAQAGSHSDNLKEVVGELAGSPAIKKGILQS LKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQRYKLDDGVKN LASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDNLSQ YDIDAIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCVFWK KLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVAR ILDERFNNELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYH HAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYS EATRKMFYSNLMNMFKRVRRLADGSIVVRPIETGRYMGKTAWDKKKH FATVRKVLSPQVNIKKTEIQTGGFSKESILAHGNSDKLIPRKTIDY LDPKKYGGFDSPIVAYSVLVVADIKKGAQKLTVTLLGITIMERSKF EKNPSAFLESKGYLNIRDDKLMILPKYSLFELENGRRRLASAGELQKG NELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQL INDFSKRVLADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGA PAAFKFDDKIVDRKRYTSTKEVLNSTLIHQSIITGLYETRIDLKSLGEDD YKDDDDKDYKDDDDKDYKDDDDKSGGSKRPAATKKAGQAKKKK (SEQ ID NO: 155)
Control	NLS04-CRISPR nickase-Linker01- Variant MMLV- NLS03-FLAG	MKRTADGSEFESPKKKRKVNKPYSIGLDIGTNSVGWSIITDDYKVPAAK MRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLLKRTARRRYTRRRNRIL YLQEIFAEEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYH EKFPITYHLRRELADKQEKADLRLYLALAHIIKFRGHFLIEDDRFDVR NTDIQKQYQAFLEIFDITTFENNHLSSQNVDEAILTDKISKSAKKDRIL AQYPNQKSTGIFAELKLVGNQADFKKHFNLEDKTPLQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQR YDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAIFY KLSKLLTKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHTELKAI IRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDD SIRPWNFEDLVDKEKSAAEAFIHRMTNNDFYLPPEKVLPHKSLIYEKFTV YNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAK EYEEFRIVDVIKLDKENKAFNASLGTYHDLKILDKDFLDNPDNESI DIVQTLTLFEDREMIKKRLKNYKDLFTESQLKKLYRRHYTGWGRLSAKL INGIRDKESQKTIIDYLLIDGGRSNRNMQLINDDGLSFKSIISKAQAGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMAR ENQTTNQGRNRSRQRYKLDDGVKNLASDLNGNILKEYPTDNQALQNER LFLYYLQNGRDMYTGEALDIDNLSQYDIDAIIPQAFIKDDSIDNRVLV SAKNRGKSDDVPSLEIVKDCVFWKLLDAKLMSQRKYDNLTKAERGGL TSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKGRIRKVKIVT LKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEP EFVYGLYRRKKLSKIVREDKEEKYSEATRKMFYSNLMNMFKRVRRLAD GSIVVRPIETGRYMGKTAWDKKKHFATVRKVLSPQVNIKKTEIQTG GFSKESILAHGNSDKLIPRKTIDYLDPKKYGGFDSPIVAYSVLVVADI KKGAQKLTVTLLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMIL PKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKG KPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVLADANLEKINKLYQD NKENISVDELANNIINLFTFTSLGAPAAFKFDDKIVDRKRYTSTKEVLN STLIHQSIITGLYETRIDLKSLGEDSGGSSGGSSGSETPGTSESATPESS GGSSGSSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIKQYPMSEARLGIKPHIQRLDQGIIVPC QSPWNTPLLPVKKPGTNDYRPVQDLREVNRKVEDIHTVPNPNYLLSLG PPSHQWYTVLDLKDFAFFCLRLHPTSQPLFAFEWRDPEMGSQGLTWTRL PQGFKNSTPLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCC

Construct Name	Protein Configuration	Complete AA Sequence
		QGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYYLLKEGQRWLTEARKE TVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDOOKAYOEIKOALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTOK LGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLV ILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDREVQFGPVVALNPAT LLPLPEEGLQHNCILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQE GQRKAGAAVTTEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLVN YTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPK RPAATKKAGQAKKKDKYKDDDDKYKDDDDKYKDDDDK (SEQ ID NO: 156)

This Example thus shows that incorporation of edits into a target can be improved through optimizing the configurations of CRISPR nuclease-reverse transcriptase and CRISPR nickase-reverse transcriptase fusion polypeptides.

5 **Example 8: Further Screening of CRISPR Nickase-Reverse Transcriptase Fusion Polypeptides**

This Example describes the design and implementation of a reporter-based HEK293T stable cell line that can be used to measure activity of CRISPR nickase-reverse transcriptase fusion polypeptides. This system is an orthogonal readout to the NGS-based assay used in the previous Example.

A modified version of the traffic light reporter (TLR) assay described in Glaser et al., Mol Ther Nucleic Acids 5(7): e334 (2016) was used to create a stable cell line with an integrated Blue Fluorescent Protein (BFP) reporter target. Editing by conversion of BFP to eGFP through a two amino acid change (SH>TY) yields the ability to measure eGFP intensity as a percentage of the total cell population. In this assay, cells comprising the integrated reporter are mCherry-positive.

The sequences of the BFP target and the editing template RNA sequence designed to convert BFP to eGFP are shown in **Table 18**. The editing template RNA was cloned and transfected into the stable cell line as described in **Example 6**. The CRISPR nickase-reverse transcriptase fusion polypeptide library described in the previous Example was also transfected.

Table 18. Sequences for TLR Screen

Component	Sequence
BFP Target	GGCTGAAGCACTGGACGCCG (SEQ ID NO: 157)
Editing Template RNA	GGCUGAAGCACUGGACGCCGGUUUUAGAGCUGUGCUGAAAAAG CACAGCACGUUAAAAUAAGGCAGUGAUUGAAAAAUCCAGUCC

	GUAUUCAGCUUGAAAAAGUGAGCACCGAAUCGGUGCUUAGCU GCCUCUGGCCUAAGGACAGCCUGAAGUCGACCACCCUGACCUA CGGCGUCCAGUGCUUC (SEQ ID NO: 158)
--	--

Analysis for the TLR screen was performed by imaging live cells at 72 hours post transfection on the Operetta CLS (Perkin Elmer) and with its Harmony software. Quantification of eGFP was performed and compared to the total mCherry positive cell population. The mCherry population represents the total number of cells that contain the integrated reporter. Imaging data was collected and quantified as a percentage of eGFP positive cells relative to the mCherry positive cell population.

The top-performing CRISPR nickase-reverse transcriptase fusion polypeptides from the TLR screen are shown in **FIG. 5**, illustrated as the percentage of eGFP positive cells in the mCherry-positive population. The hits are rank ordered by activity compared to the control CRISPR nickase-reverse transcriptase polypeptide and correlate with the NGS data from **FIG. 4A** and **FIG. 4B**. One additional trend that was observed was that increasing flexible linker length (Linker_16xGGGGS > Linker_8xGGGGS > Linker_4xGGGGS > Linker_1xGGGGS) between the CRISPR nickase component and the reverse transcriptase component yielded increasing BFP→eGFP editing, translating to increased eGFP-positive cells in the reporter assay. Therefore, increasing the length of the linker between the CRISPR nickase components and the reverse transcriptase components of the CRISPR nickase-reverse transcriptase fusion polypeptides in **Table 17** may be beneficial in further increasing editing efficiency.

Next, to compare the robustness of the reporter-based eGFP quantitation from the TLR screen, the top-performing hits were graphed and compared to the NGS-quantified edits for the conversion of BFP to eGFP (**FIG. 6**). The percentage of edits versus the percentage of eGFP positive cells was positively correlated; higher edit installation resulted in higher percentages of eGFP positive cells. Finally, edits at the EMX1_T2 target from the previous Example were compared to BFP-to-eGFP edits for the top-performing CRISPR nickase-reverse transcriptase fusion polypeptides. As shown in **FIG. 7**, for each CRISPR nickase-reverse transcriptase fusion polypeptide, there was a positive correlation between edits at both targets. Additionally, edits by each CRISPR nickase-reverse transcriptase fusion polypeptide exceeded those by the control CRISPR nickase-reverse transcriptase fusion polypeptide.

This Example thus shows that optimized CRISPR nickase-reverse transcriptase fusion polypeptides introduce edits at multiple target loci.

Example 9: Design of Additional CRISPR Nucleases

This Example describes design of additional CRISPR nucleases with desired bioactivities.

Additional variants of the CRISPR nickases described herein are engineered. Each individual amino acid residue of the CRISPR nickase with an H845A substitution from Example 3 is replaced with the remaining nineteen available amino acids (except for the H845A residue). Nucleic acids encoding the CRISPR nickase variants are cloned into a pcDNA3.1 vector (Invitrogen) comprising a CMV promoter. The expression vectors are introduced into host cells for expression of the CRISPR nickase variants. The CRISPR nickase variants expressed in host cells are purified and their nickase activities are evaluated following the procedures provided in Example 3 above.

Additional CRISPR nuclease variants are engineered to increase double-strand nuclease activity. The variants of Table 19 are cloned and evaluated as described in Example 2.

Table 19. Variant CRISPR Nuclease Sequences

Variant	Amino Acid Sequence	SEQ ID NO
A	MNKPYSIGLDIGTNSVGWSIITDDYKVPAAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKEELADKQEKADLRLLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVDVEAILTDKISKSAAKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKTNQEAFYKYLKLLTKQEDSENFLKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTRKTDDSIIRPWNFEDLVDKEKSAAFIHRMTNNDKYLPEEKVLPKHSIYKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEEYEEFRIVDVIIGLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLNENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKEAQKTILDYLIIDGRSNRNFQMLINDDGLSFKSIISKAGAGSHSDNLKEVVGEVLAGSPAIAKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQRYKRLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDKCVFVWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIKRQLVETRQITKHVARILDERFNNELDSKGRRIKRVKIVTLKSNLVSFRKEFGFYKIREVNDYHHAHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKATRKMFYFYNLMNMFKRVRVRLADGSIVVRPIETGRYMGKTAWDKKKHFATVRKVLSTYPQVNIKKTEIQTGGFSKESILAKRNSDKLIPRKKKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGYLNIRDDKLMKLPKYSLEFELNGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSTITGLYETRIDLSKLGED	159
B	MNKPYSIGLDIGTNSVGWSIITDDYKVPAAKMRVLGNTDKEYIKKNLIGALLFDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	160

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFL EKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDKYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIKRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESIRAKRNSDKLIPRKKKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMKLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
C	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRR LKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFL EKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDKYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIKRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILPKRNSDKLIPRKKKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMKLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	161
D	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRL EDS	162

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRLIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPF LKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDKYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIKRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKATR KMF FYSNLMMNMFKRVR LADGSIVVRPIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESIRPKRNSDKLIPRKKKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMKLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTRLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
E	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRLIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPF LKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDKYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSR QRYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIKRQLVETRQITKH VARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYH AHADAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKAT RKMFFYSNLMMNMFKRVR LADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNI VKKTEIQTGGFSKESILAKRNSDKLIPRKKKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMKLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA ASRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNI INLFTF TSLGAPAAFKFFDKIVDRKRYRSTKE VLNSTLIHQSI TGLYETRIDLSKLGED	163
F	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	164

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIRPWNFEDLVDKEKSAEAFIHRMTNNDKYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKG I LQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSR QRYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRKQLVETRQITKH VARILDERFNNELDSKGRRIKVKIVTLKSNLVS NFRKEFGFYKIREVNDYH HAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSKAT RKMFFYSNLMMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNI VKKTEIQTGGFSKESIRAKRNSDKLIPRKKKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY YLNIRDDKLMRLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYL ASRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFFDKIVDRKRYRSTKE VLNSTLIHQSIITGLYETRIDL SKLGED	
G	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRR LKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKG I LQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRKQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDL SKLGED	165
H	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSF FHRLEDS	166

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYKYLKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKA I LTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESIRARRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
I	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRR LKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYKYLKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKA I LTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILPRRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	167
J	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSFHRL EDS	168

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKRLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFVWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRIRKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMF FFYSNLMMNMFKRVRRLADGSIVVRPIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESIRPRRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMRLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
K	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNRSR QRYKRLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFVWKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKH VARILDERFNNELDSKGRIRKVKIVTLKSNLVS NFRKEFGFYKIREVNDYH AHADAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR RKMFFYSNLMMNMFKRVRRLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNI VKKTEIQTGGFSKESILARRNSDKLIPRKRKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA ASRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFFDKIVDRKRYRSTKE VLNSTLIHQSIITGLYETRIDLSKLGED	169
L	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	170

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSR QRYKRLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKH VARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVN DYH HAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRAT RKMFFYSNLMMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNIVKKTEIQTGGFSKESIRARRNSDKLIPRKRKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMRLPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYL ASRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNIINLFTFTRLGAPAAFKFDDKIVDRKRYRSTKE VLNSTLIHQSIITGLYETRIDL SKLGED	
M	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSR QRYKRLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKH VARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVN DYH HAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRAT RKMFFYSNLMMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNIVKKTEIQTGGFSKESILPRRNSDKLIPRKRKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKG YLNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYL ASRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNIINLFTF TSLGAPAAFKFDDKIVDRKRYTSTKE VLNSTLIHQSIITGLYETRIDL SKLGED	171
N	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSF FHRLEDS	172

Variant	Amino Acid Sequence	SEQ ID NO
	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSR QRYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGE ALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKD CKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKH VARILDERFNNELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYH HAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRAT RKMFFYSNLMNMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK VLSYPQVNI VKKTEIQTGGFSKESIRPRRNSDKLIPRKRKDIYLDPKKYGGF DSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY YLNIRDDKLMILPKYSLFELENGRRRLASARELQKGNELALPTQFMKFLYLA ASRYNESK GKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKI NKLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYRSTKE VLNSTLIHQSIITGLYETRIDL SKLGED	
D462R_ F501R_ H657R_ L784R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTP LQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGRYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFYYSNLMNMFKRVRVLADGSIVVRPIETGRYMGKTAWDKKKHFATVRK LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESK GKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYRSTKEV LNSTLIHQSIITGLYETRIDL SKLGED	173
D462R_ F501R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	174

Variant	Amino Acid Sequence	SEQ ID NO
H657R_Q923R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
D462R_F501R_L784R_Q923R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	175
D462R_H657R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFA EEMSKVDDSFHRL EDS	176

Variant	Amino Acid Sequence	SEQ ID NO
L784R_ Q923R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQE FVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ H657R_ L784R_ Q923R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQE FVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	177
D462R_ F501R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRLSDS	178

Variant	Amino Acid Sequence	SEQ ID NO
H657R_I 1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
D462R_ F501R_I 1206R_Q 923R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	179
D462R_ H657R_I	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	180

Variant	Amino Acid Sequence	SEQ ID NO
1206R_Q923R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
F501R_H657R_I1206R_Q923R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	181
D462R_F501R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	182

Variant	Amino Acid Sequence	SEQ ID NO
H657R_ G1120R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
D462R_ F501R_ G1120R_ I1206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	183
D462R_ H657R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	184

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRL LASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ H657R_ G1120R_ I1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRL LASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	185
F501R_ H657R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	186

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ T1130R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ H657R_ G1120R_ I1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLSSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	187
F501R_ H657R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	188

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ G1120R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	189
H657R_ G1120R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSF FHRLEDS	190

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
H657R_ L784R_ G1120R_ T1130R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	191
H657R_ L784R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	192

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
H657R_ L784R_T I130R_I 1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	193
L784R_ G1120R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	194

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
H657R_ L784R_ H1119R_ T1130R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	195
H657R_ L784R_	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	196

Variant	Amino Acid Sequence	SEQ ID NO
H1119R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
H657R_ L784R_T I1130R_I I1206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	197
H657R_ H1119R_	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	198

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_11206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
L784R_ H1119R_ T1130R_ 11206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKRLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	199
F501R_ Q923R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	200

Variant	Amino Acid Sequence	SEQ ID NO
H1119R_ T1130R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ Q923R_ H1119R_ I1206R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEA FYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSII SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	201
F501R_ Q923R_	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	202

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPA IKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ H1119R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPA IKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	203
Q923R_ H1119R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	204

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDIFYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
F501R_ Q923R_ G1120R_ T1130R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	205
F501R_ Q923R_	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSF FHRLEDS	206

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESI LEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_ Q923R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESI LEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	207
F501R_ G1120R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	208

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_11206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHSIYEEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDYLDIDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYGGF SPIVAYSVLVVADIKKGKAQKLKTIVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
Q923R_11206R G1120R_11206R T1130R_11206R	MNKPYSIGLDIGTNSVGWSIITDDYKVPKAKMRVLGNNDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHSIYEEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDYLDIDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRRLADGSIVVRPIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYGGF SPIVAYSVLVVADIKKGKAQKLKTIVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	209
F501R_E 1037R_G	MNKPYSIGLDIGTNSVGWSIITDDYKVPKAKMRVLGNNDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	210

Variant	Amino Acid Sequence	SEQ ID NO
1120R_T 1130R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_E 1037R_G 1120R_I 1206R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKTD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	211
F501R_E 1037R_T	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	212

Variant	Amino Acid Sequence	SEQ ID NO
I130R_I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKEITQTGGFSKESILAHGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
F501R_G1120R_T1130R_I1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSIISK AQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKEITQTGGFSKESILAHNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	213
E1037R_G1120R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEMSKVDDSFHRLSDS	214

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
H657R_ E1037R_ G1120R_ T1130R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	215
H657R_ E1037R_	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	216

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKA I LTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
H657R_ E1037R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILIY LQE IFAEEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKA I LTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	217
H657R_ G1120R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILIY LQE IFAEEMSKVDDSF FHRLEDS	218

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRRYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VVKTEIQTGGFSKESILAHNRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
D462R_ E1037R_ G1120R_ T1130R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDL EKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VVKTEIQTGGFSKESILAHNRNSDKLIPRKRKDIYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	219
D462R_ E1037R_	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	220

Variant	Amino Acid Sequence	SEQ ID NO
G1120R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKT KD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
D462R_ E1037R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYL ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	221
D462R_ G1120R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	222

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHNRSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
D462R_ E1037R_ H1119R_ T1130R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	223
D462R_ E1037R_	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	224

Variant	Amino Acid Sequence	SEQ ID NO
H1119R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKT KD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
D462R_ E1037R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKD IYLDPKKYGGFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	225
D462R_ H1119R_	MNKPYSIGLDIGTNSVGWSIIITDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	226

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_11206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDIFYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPA IKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
F501R_E 1037R_H 1119R_T 1130R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSRFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPA IKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	227
F501R_E 1037R_H	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	228

Variant	Amino Acid Sequence	SEQ ID NO
1119R_I 1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDITTFENNHLSSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHSIYKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDLAKE EEFRIVDVIGLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDYLDIDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKEITQTGGFSKESILARGNSDKLIPRKTCDIYLDPKKYGDFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	
F501R_E 1037R_T 1130R_I 1206R	MNKPYSIGLDIGTNSVGWSIITDDYKVPKAKMRVLGNNDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDITTFENNHLSSQNVD VEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSIIRPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHSIYKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDLAKE EEFRIVDVIGLDKENKAFNASLGTYHDLKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRLNENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDYLDIDGRSNRNFMLINDDGLSFKSIISKAQAGSHSDNLKEVVGE LAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFFATVRKV LSYPQVNIIVKKEITQTGGFSKESILAHGNSDKLIPRKRKDIYLDPKKYGDFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSIITGLYETRIDLSKLGED	229
F501R_ H1119R_	MNKPYSIGLDIGTNSVGWSIITDDYKVPKAKMRVLGNNDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDS	230

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_ I1206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDRYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
E1037R_ H1119R_ T1130R_ I1206R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAELKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEDLLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLPEEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKE EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVS NFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNI INLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	231
Q923R_ E1037R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	232

Variant	Amino Acid Sequence	SEQ ID NO
H1119R_ T1130R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	
Q923R_ E1037R_ H1119R_ I1206R	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYI EGKTNQEAIFYKLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP EEKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQEFVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDLSKLGED	233
Q923R_ E1037R_	MNKPYSIGLDIGTNSVGWSII TDDYKVP AKMRVLGN TDKEYIKKNLIGALL FDGGNTAADRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRL EDS	234

Variant	Amino Acid Sequence	SEQ ID NO
T1130R_11206R	FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSRATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILAHGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRL LASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQE FVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDL SKLGED	
Q923R_ H1119R_ T1130R_ 11206R	MNKPYSIGLDIGTNSVGWSIITDDYKVP AKMVRVLGNTDKEYIKKNLIGALL FDGGNTAADRR LKRTARRRYTRRRNRILYLQEIFAEMSKVDDSF FHRLEDS FLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHLRKELADKQEKADLRILIYLA ALAHIIKFRGHFLIEDDRFDVRNTDIIQKQYQAFLEIFDTTFENNHLLSQNVD VEAILTDKISKSAKKDRI LAQYPNQKSTGIFAEFLKLIVGNQADFKKHFNLE DKTPLQFAKDSYDEEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDL STKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYI EGKTNQEA FYK YLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVH LTELKAIIRRQSEYYPFLENQDRIEKILTFRIPYYIGPLAREKSDFAWMTR KTDDSI RPWNFEDLVDKEKSAEAFIHRMTNNDFYLP E EKVLPKHS LIYEKFT VYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVS KKKLLDFLAKEY EEFRIVDVI GLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQT LTLFEDREMIKKRL ENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKE SQKTILDY LIDDGRSNRNF MQ LINDDGLSF KSI I SKAQAGSHSDNLKEVVGE LAGSPA I KKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQ RYKLLDDGVKNLASDLNGN ILKEYPTDNQALQNERLFLYYLQNGRDMYTGEA LDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDC KVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIRRQLVETRQITKHV ARILDERFNNELDSKGRRI RKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHH AHDAYLNAVVAKAILTKYPQLEPEFVYGLYRRKKLSKIVREDKEEKYSEATR KMFFYSNLMNMFKRVRVLADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKV LSYPQVNI VKKTEIQTGGFSKESILARGNSDKLIPRKRKDIYLDPKKYG GFD SPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKF EKNPSAFLESKGY LNIRDDKLMRLPKYSLFELENGRRRL LASAGELQKGNELALPTQFMKFLYLA SRYNESKGKPEEIEKKQE FVNQHVS YFDDILQLINDFSKRVI LADANLEKIN KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEV LNSTLIHQSI TGLYETRIDL SKLGED	235

Additional CRISPR nuclease variants are engineered and evaluated for their ability to recognize less stringent PAM sequences. The variants of **Table 20** are cloned and evaluated as described in **Example 2** using target sequences adjacent to 5'-NGN-3', 5'-NRN-3', or 5'-NYN-3' PAM sequences, in which N represents any nucleotide, R represents G or A, and Y represents C or T.

Table 20. Variant CRISPR Nuclease Sequences.

Variants	Amino Acid Sequences	SEQ ID NO
L1117R_D1144V_G1227R_E1228F_A1332R_R1345V_T1347R	MNKPYSIGLDIGTNSVGWSIIITDDYKVPAAKKMRVLGNTDKEYIKKN LIGALLFDGGNTAADRRLLKRTARRRYTRRRNRILYLQEIFAEMSK VDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHL RKELADKQEKADRLRIYLALAHIIKFRGHFLIEDDRFDVRNTDIQK QYQAFLEIFDTTFENNHLLSQNVDEAILTDKISKSAKKDRILAQY PNQKSTGIFAEFLKLIVGNQADFKKHFNLEDKTPQLQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASM IQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKT NQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQ VHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREK SDFAWMTRKTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLPEE KVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGV FKEHRKVSKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLGTYH DLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDL FTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTI LDYLIDDGR SNRNFMLINDDGLSFKSII SKAQAGSHSDNLKEVVGELAGSPAIAK KGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNSRQRY KLLDDGVKNLASDLNGNIIKEYPTDNQALQNERLFLYYLQNGRDMY TGEALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDD VPSLEIVKDCVKFVWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILDERFNNELDSKGRRIKRVKIVTLKSN LVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPE FVYGLYRRKKLSKIVREDKEEKYSEATRKMFYFYNLMNMFKRVVRL ADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKVLSYPQVNIVKKT EIQTGGFSKESIRAHGNSDKLIPRKTKDIYLDPKKYGGFVSPIVAY SVLVVADIKKGKAQKLKTIVTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRLLASARFLQKGNELALPTQFM KFLYLASRYNESKGPKEIEKKQEFVNQHVSYFDDILQLINDFSKR VILADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPRAF KFFDKIVDRKVYRSTKEVLNSTLIHQSIITGLYETRIDL SKLGED	236
D1144L_S1145W_G1227K_E1228Q_R1345Q_T1347R	MNKPYSIGLDIGTNSVGWSIIITDDYKVPAAKKMRVLGNTDKEYIKKN LIGALLFDGGNTAADRRLLKRTARRRYTRRRNRILYLQEIFAEMSK VDDSFHRLSDSFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHL RKELADKQEKADRLRIYLALAHIIKFRGHFLIEDDRFDVRNTDIQK QYQAFLEIFDTTFENNHLLSQNVDEAILTDKISKSAKKDRILAQY PNQKSTGIFAEFLKLIVGNQADFKKHFNLEDKTPQLQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASM IQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGAGYIEGKT NQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQ VHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREK SDFAWMTRKTDDSI RPWNFEDLVDKEKSAEAFIHRMTNND FYLPEE KVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGV	237

	FKEHRKVSKKKLLDFLAKEEYEEFRIVDVIGLDKENKAFNASLGTYH DLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDL FTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTIILDYLI DDGR SNRNFQMQLINDDGLSFKSIIISKAQAGSHSDNLKEVVGELAGSPAIAK KGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQRY KLLDDGVKNLASDLNGNIIKEYPTDNQALQNERLFLYYLQNGRDMY TGEALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDD VPSLEIVKDCVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILDERFNNELDSKGRRIKVKIVTLKSN LVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPE FVYGLYRRKKLSKIVREDKEEKYSEATRKMFYFYNLMNMFKRVRRL ADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKVLSPQVNIKKKT EIQTGGFSKESILAHGNSDKLIPRKT KD IYLDPKKYGGFLWP IVAY SVLVVADIKKGKAQKLKT VTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRL LASAKQLQKGNELALPTQFM KFLYLASRYNESKKGPEEIEKKQEFVNQHVSYFDDILQLINDFSKR VILADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPAAF KFFDKIVDRKQYRSTKEVLNSTLIHQSI TGLYETRIDLSKLGED	
D61R_L1117R_D1 144L_S1145W_G1 227K_E1228Q_S1 327R_A1332R_R1 343P_R1345Q_T1 347R	MNKPYSIGLDIGTNSVGWSIIITDDYKVPKMKMRVLGNTDKEYIKKN LIGALLFDGGNTAARRRLKRTARRRYTRRRNRILYLQEIFAEMSK VDDSFHRLLED SFLVEEDKRGSKYPIFATLQEEKDYHEKFPTIYHL RKELADKQEKADRLRIYLA LAHI IKFRGHFLIEDDRFDVRNTDIQK QYQAFLEIFDTTFENNHLSSQNVDVEAILTDKISKSAKKDRILAQY PNQKSTGIFAEFLKLI VGNQADF KKHFNLEDKTP LQFAKDSYDEDL ENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASM IQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDG YAGYIEGKT NQEAFYKYL SKLLTKQEDSENFL EKIKNEDFLRKQRTFDNGSIPHQ VHTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREK SDFAWMTRKTDSDSIRPWNFEDLVDKEKSAEAFIHRMTNND FYLPEE KVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGV FKEHRKVSKKKLLDFLAKEEYEEFRIVDVIGLDKENKAFNASLGTYH DLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDL FTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTIILDYLI DDGR SNRNFQMQLINDDGLSFKSIIISKAQAGSHSDNLKEVVGELAGSPAIAK KGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRNRSRQRY KLLDDGVKNLASDLNGNIIKEYPTDNQALQNERLFLYYLQNGRDMY TGEALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDD VPSLEIVKDCVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKAR FIQRQLVETRQITKHVARILDERFNNELDSKGRRIKVKIVTLKSN LVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPE FVYGLYRRKKLSKIVREDKEEKYSEATRKMFYFYNLMNMFKRVRRL ADGSIVVRPVIETGRYMGKTAWDKKKHFATVRKVLSPQVNIKKKT EIQTGGFSKESIRAHGNSDKLIPRKT KD IYLDPKKYGGFLWP IVAY SVLVVADIKKGKAQKLKT VTELLGITIMERSKFEKNPSAFLESKGY LNIRDDKLMILPKYSLFELENGRRRL LASAKQLQKGNELALPTQFM KFLYLASRYNESKKGPEEIEKKQEFVNQHVSYFDDILQLINDFSKR VILADANLEKINKLYQDNKENISVDELANNIINLFTFTSLGAPRAF KFFDKIVDPKQYRSTKEVLNSTLIHQSI TGLYETRIDLSKLGED	238

Example 10: Effectiveness of Variants of a CRISPR Nuclease with Relaxed PAM Stringency for Targeting of Exemplary Mammalian Genes

5 This Example describes indel assessment on exemplary mammalian targets using

variants of a CRISPR nuclease with a relaxed PAM transfected into HEK293T cells.

Arginine scanning mutagenesis was performed to individually substitute selected non-arginine residues of the CRISPR nuclease variant of SEQ ID NO: 236 to arginine. This resulted in 372 single arginine substitution variants. The variants were cloned and evaluated as described in **Example 2** using the target sequences adjacent to 5'-NGN-3' PAM sequences summarized in **Table 21**.

Table 21. Mammalian Targets and Corresponding crRNAs

Target Identifier	PAM Sequence	Target Sequence	RNA Guide Sequence
AAVS1	TGC	AACTCCCTTTGTGAGAATGG (SEQ ID NO: 239)	AACUCCCUUGUGAGAAUGGGUU UUAGAGCUGUGCUGAAAAGCACA GCACGUUAAAAUAAGGCAGUGAU UGAAAAAUCCAGUCCGUAUUCAG CUUGAAAAAGUGAGCACCGAAUC GGUGCUU (SEQ ID NO: 240)
EMX1	GGA	GGATGGCGACTTCAGGCA CA (SEQ ID NO: 241)	GGAUGGCGACUUCAGGCACAGUU UUAGAGCUGUGCUGAAAAGCACA GCACGUUAAAAUAAGGCAGUGAU UGAAAAAUCCAGUCCGUAUUCAG CUUGAAAAAGUGAGCACCGAAUC GGUGCUU (SEQ ID NO: 242)

HEK293T cells were further transfected, followed by NGS analysis, as described in **Example 2**. Indel activity of the CRISPR nuclease variant of SEQ ID NO: 236 is shown in **Table 22**. The data in **Table 22** is the average of ten control samples, each of which had two bioreplicates and two technical replicates.

Table 22. Percentage of NGS Reads Comprising Indels

Variant	AAVS1	EMX1
SEQ ID NO: 236	5.0	2.6

Next, for each target, indel ratios, referring to the percentage of NGS reads comprising indels, were calculated for the variant CRISPR nuclease (SEQ ID NO: 236) and for each variant CRISPR nuclease. To then calculate fold change in indel ratios, the indel ratio for each variant was divided by the indel ratio for the variant CRISPR nuclease of SEQ ID NO: 236. The indel ratios used for fold change calculations were the average of two technical replicates. As shown in **Table 23**, 3 of the 372 variants with single arginine substitutions (left column)

were characterized as yielding at least a 2X increase in indel ratio relative to the indel ratio for the variant CRISPR nuclease of SEQ ID NO: 236, when averaged across the two targets (right column).

5 **Table 23. Fold Change in Indel Ratios***

Substitution	AAVS1	EMX1	Average Fold Change
A68R	1.1	4.3	2.7
D61R	1.4	3.00	2.2
H494R	1.1	3.1	2.1

* Variant indel ratio/indel ratio of variant set forth in SEQ ID NO: 236

11 variants with single arginine substitutions were analyzed as having indel ratios 1.5X-2X of the reference indel ratios: L64R, S410R, T67R, Q849R, G1110R, F501R, T659R,
10 L784R, Y516R, G55R, and E1037R. 92 variants exhibited indel ratios 1-1.4X of the reference indel ratios: N57R, D720R, A919R, A1294R, Q812R, N700R, H657R, T73R, Q899R, T1347R, I857R, K751R, D327R, I581R, D462R, E331R, A589R, D471R, I699R, N1295R, T470R, I1147R, E130R, S473R, A353R, K40R, K334R, A60R, S1348R, K367R, A1118R, K31R, Q349R, K341R, Q83R, K585R, Q840R, G660R, K527R, G727R, Y42R, L1281R,
15 L122R, Q123R, T1108R, E41R, K1131R, K30R, S872R, I1206R, D1132R, K460R, L80R, E459R, K1182R, L1117R, M696R, K918R, K126R, N721R, G1227R, Q809R, K1091R, K736R, A1332R, K783R, N498R, K723R, E1228R, H1119R, F463R, L594R, D472R, K744R, E365R, G595R, K45R, Y348R, K964R, S1181R, N813R, D407R, S839R, Y658R, E586R, G754R, A730R, Y1015R, D903R, A1333R, S461R, and H1359R. The remaining variants with
20 single arginine substitutions (266 variants) resulted in decreased indel ratios relative to the indel ratios for the variant CRISPR nuclease of SEQ ID NO: 236 (fold change in indel ratios of less than 1.0).

This Example thus shows that the CRISPR nuclease variant of SEQ ID NO: 236 is an active nuclease capable of editing target sequences adjacent to a 5'-NGN-3' PAM (N
25 representing A, C, G, or U) and that particular further arginine substitutions (*e.g.*, D61R, A68R, and/or H494R) increase nuclease activity.

30 **Example 11: mRNA-Mediated Editing of Target Sequences in Primary Human Hepatocytes**

This Example describes genomic editing of the EMX1 and VEGFA genes using mRNA

encoding a CRISPR Nuclease-Reverse Transcriptase fusion polypeptide or a CRISPR Nickase-Reverse Transcriptase fusion polypeptide.

- Nucleic acids encoding the CRISPR Nuclease-Reverse Transcriptase fusion polypeptide of SEQ ID NO: 55 and nucleic acids encoding the CRISPR Nickase-Reverse Transcriptase fusion polypeptide of SEQ ID NO: 57 were individually cloned into an *in vitro* transcription (IVT) backbone comprising a T7 promoter. Research grade and sequence verified plasmids were obtained using a maxi prep kit (Qiagen). mRNAs were generated through *in vitro* transcription of the IVT backbones, adding a 5' cap and 3' polyA tail. The full-length mRNA sequences are shown in **Table 24**. Working solutions of each mRNA were prepared in water.

Table 24. mRNA Sequences

Component	mRNA Sequence
CRISPR Nuclease-Reverse Transcriptase mRNA	GGAGAGACCCAAGCUGUCUAGACGCCACCAUGAAACGGACAGCCGACGGAAGCGAGU UCGAGUCACCAAAGAAGAAGCGGAAAGUCAACAAGCCUACAGCAUCGGCCUGGAUA UCGGCACCAACAGCGUGGGAUGGUCUAUCAUCACAGAUGACUACAAGGUGCCUGCCA AGAAAAUGAGAGUGUUAGGCAACACCGACAAGGAGUACAUCAAGAAGAACUGAUCG GCGCCCUCCUGUUCGACGGCGGAAACACAGCCGCCGAUCGGCGGGCUGAAAAGGACCG CCAGACGGAGAUAUACGCGGCGGAGAAAUAGAAUCCUGUACUACAGGAAAUCUUCG CUGAGGAGAUAGCAAAGUGGACGACUCUUUUUCCACAGACUGGAAGAUAGCUUUU UGGUGGAAGAGGACAAGCGAGGCUCUAAGUACCCUAUCUUCGCUACACUGCAAGAGG AAAAGGACUACCACGAGAAGUUUCCACCAUCUACCACCUGAGGAAAGAGCUGGCGG ACAAACAAGAGAAGGCCGACCUGAGACUGAUCUACCUUGCCCUGGCUCAUAUCAUCA AGUUCAGAGGCCACUCCUGAUCGAGGACGACAGAUUCGACGUGAGAAAUACAGAUUA UCCAGAAGCAGUACCAGGCCUUUCUGGAGAUUCUUCGACACCACCUUUGAAAACAACC ACCUGCUGAGCCAAACGUGGAUGUGGAAGCCAUUUUACCGAUAAGAUCCUCCAAGA GCGCCAAGAAAGAUAGAAUCCUCGCCCAGUAUCCUAAACCAGAAAAGCACCAGGUAUCU UCGCCGAAUUCUGAAGCUGAUCGUGGGCAACCAGGCCGACUUCAAAAAACACUUUA ACCUGGAGGACAAAACACCUCUGCAGUUUGCCAAGGACUCCUACGACGAAGAUCUGG AAAACCUGCUGGGCCAGAUCCGGCGACGAGUUCGCCGAUCUGUUCAGCGCCGCUAAGA AACUGUACGACAGCGUGCUGCUGAGCGGCAUCUUGACAGUGAUCGAUCUGUCCACAA AGGCCCCCUGAGCGCCAGCAUGAUCAGAGAUACGACGACAGAGAGGAUCUGA AGCAGCUGAAGCAGUUUGUAAGGCCUCUCUGCCCGAGAAGUACCAGGAGAUUUUUG CGGACUCCAGCAAGGAUGGAUACGCCGGAUACAUUGAGGGCAAGACCAACCAGGAAG CUUUCUACAAGUACCUGUCUAAACUGCUGACCAAACAGGAAGAUUCCGAGAACUCC UGGAGAAAAUCAAGAACGAAGAUUCCUGAGAAAGCAGAGGACCUUCGACAACGGCA GCAUCCCCCACCAGGUCCACCUGACCGAGCUGAAGGCUAUCAUCAGACGACAGAGUG AGUACUACCCCUCCUGAAGGAAAAUCAAGACAGAAUCCGAGAAGAUUCUGACAUUCA GAAUCCCAUACUAUAUCGGACCUCUGGCCCGUGAAAAGAGCGACUUCGCCUGGAUGA CCAGAAAGACCGAUGAUUCUAUCCGGCCUGGAACUUCGAGGACCUGGUGGACAAGG AAAAGUCUGCCGAGGCCUUCAUCCACCGAUGACCAACAACGACUUCUACCUGCCUG AAGAGAAGGUGCUGCCUAAACAUUCCUGAUCUACGAAAAGUUCACAGUGUACAACG AGCUGACCAAGGUUAGAUACAAGAACGAGCAGGGCGAGACGUACUUCUUCGACUCUA AUUAUCAAGCAGGAAAUUCGAUGGAGUGUUAAGGAGCACCGGAAGGUCUCUAAGA AGAAACUGUAGACUUCUGGCUAAGGAUACGAGGAUUCGGAUCUGUGGACGUGA UCGGACUGGAUAAGGAGAACAAAGCCUUUAACGCCAGCCUGGGAACAUAACCACGACC UGGAGAAGAUCCUCGACAAGGACUUCUGGAUAACCCGACAAUGAGUCCAUCCUGG

	AAGAUUUCGUGCAGACGCGUGACACUGUUCGAGGACAGAGAGAUGAUCAAGAAGAGAC UGGAAAACUACAAGGAUCUGUUCACCGAGAGCCAGCUGAAGAAGUUGUAUAGACGGC ACUACACCGGAUGGGGCGCGUGUCCGCUAAACUGAUCAACGGCAUCAGAGAUAAAG AAAGUCAGAAAACCAUCCUGGACUACUUAUCGACGAUGGCAGAAAGCAACCGGAACU UCAUGCAGCUGAUCAACGAUGAUGGCUUGAGCUUCAAGUCUAUCUUCUUAAGGCC AGGCCGGAAGCCACAGCGAUAAACUGAAGGAAGUGGUGGUGAGCUGGCCGCGCUCUC CUGCCAUCAAGAAGGGCAUCCUGCAAAGUCUAAAGAUUGUCGACGAGCUGGUGAAGG UGAUGGGCUAUGAACCUGAGCAGAUCCUGGUGGAAAUGGCCAGAGAAAACAGACAA CCAAUCAAGGCAGAAAGAAACAGCAGACAGCGGUACAAGCUGCUGGACGACGGCGUGA AGAAUCUGGCCAGCGACCUGAACGGCAACAUCUGAAAGAGUACCCAACCGACAACC AGGCCUGCAGAACGAGCGCCUGUUUCUGUACUACCUGCAGAACGGCAGAGACAUGU ACACCGGCGAAGCCUGGAUAUCGAUAACCGAGCCAGUACGAUAUCGACCAUAUA UCCACAGGCCUUUAUCAAGGAUGAUUCUAUUGAUAAUCGGGUGCUUGUCUCCAGCG CGAAGAACAGAGGAAGAGCGACGACGUGCCAGCCUCGAAAUCGUGAAAGACUGCA AGGUGUUCUGGAAGAAGCUCUGGACGCCAAGCUGAUGAGCCAGAGAAAAGUACGACA ACCUGACCAAGGCCGAGAGAGGAGGCCUGACCAGCGACGAUAAGGCUAGAUUCAUCC AAAGCAGCUGGUCGAAACCAGACAGAUCAACCAAGCAGUGGCCAGGAUCCUGCAGC AGCGGUUCAUAACGAGCUGGACAGCAAGGUAGAAGAAUAGAAAGUUAAGAUUCG UCACCCUGAAGUCUAUCUUGUGUCCAACUUCGGGAAGGAGUUCGGCUUCUACAAGA UCAGAGAGGUGAACGACUACCACCACGCCAUGAUGCCUACCUGAACGCCGUGGUGG CCAAAGCCAUCCUGACCAAUACCCUACUGGAGCCCGAGUUCGUGUACGGCCUGU ACAGACGGGAAGCUGAGCAAAUUCGUGCGGGAAGAUAAAGGAAGAGAAGUACAGCG AAGCCACAAGGAAAAUGUUUUUCUAUAGCAACCUGAUGAACAUUUAAGAGAGUGG UGAGACUGGCCGACGGAUCUAUCGUCGUGCGGCCUGUGAUCGAAACCGGCAGAUACA UGGGCAAAACCGCCUGGGACAAAAGAAGCACUUCGCCACCGUGCGGAAGGUACUGA GCUACCCUCAGGUGAACAUAGUGAAAAAGACAGAGAUCCAGACCGCGCGCUUCUCUA AAGAGUCCAUCCUCGUCACGGCAUAGCGACAAGCUGAUCCCGAGAAAGACAAAAAG ACAUCUACCUGGACCCUAAGAAGUAUGGCGGCUUCGACAGCCCCAUCGUGGCUUACA GCGUCCUGGUUGGCGCAUAUUAAGAAGGGCAAGGCUCAGAAGCUGAAGACGGUGA CAGAGCUGCUGGGCAUAACAAUCAUGGAACGACGCAAAUUCGAGAAACCCCAAGCG CUUUUCUGGAAAGCAAGGGCUACCUGAACAUCCGGGACGAUAAGCUGAUGAUCCUGC CUAAGUACAGCCUGUUCGAACUGGAAAAUGGCAGACGGCGGCGUGCUAGCCAGCGCCG GAGAGCUGCAAAAGGGAAUUGAGCUGGCCUGCCUACCCAGUUAUGAAAUUCUGU AUCUGGCCUCUAGAUUAACGAGAGCAAGGGCAAACCGUAAGAGAUCGAAAAGAAAC AGGAGUUUGUAACCAACACGUGAGCUACUUCGACGACAUCCUGCAGCUGAUAUAG ACUUCAGCAAGAGGGUGAUCCUGGCCGACGCCAACCGGAAAAAUCAACAAACUUA AUCAGGACAACAAGGAGAAUAUCUCCUGGACGAACUGGCUAACAACAUUAUAUC UGUUCACAUUACACGACCUUGGCGCUCUGCCGCCUUUAAGUUCUUCGACAAAAUCG UGGACCGCAAGCGGUACACCAGCACCAAGGAGGUGCUUAACAGCACACUGAUCCACC AGAGCAUCACUGGCCUGUAUGAGACAAGAAUCGACCUGUCUAAGCUGGGCGAGGAUU CUGGAGGAUCUAGCGGAGGAUCCUCUGGCAGCGAGACACCAGGAACAAGCGAGUCAG CAACACCAGAGAGCAGUGGCGGCAGCAGCGGCGGCAGCAGACCCUAAAUAUAGAAG AUGAGUAUCGGCUACAUGAGACCUAAAAGAGCCAGAUUUUCUAGGGUCCACAU GGCUGUCUGAUUUUCUAGGCCUGGGCGGAAACCGGGGGCAUGGGACUGGCAGUUC GCCAAGCUCUCUGAUCAUACCUUGAAAGCAACCUUACCCCGUGUCCAUAUAAAC AAUACCCCAUGUCACAAGAAGCCAGACUGGGGAUCAAGCCCCACAUAACAGAGACUGU UGGACAGGGAAUACUGGUACCCUGCCAGUCCCCUGGAACACGCCCCUGUACCCG UUAAGAAACAGGGACUAAUGAUUAUAGGCCUGUCCAGGAUCUGAGAGAAGUCAACA AGCGGGUGGAAGACAUCCACCCACCGUGCCCAACCCUUAACCCUCUUGAGCGGGC UCCACCGUCCACCAGUGGUACACUGUGCUUGAUUUAAAGGAUGCCUUUUUCUGCC UGAGACUCCACCCACCAGUCAGCCUCUCUUCGCCUUUGAGUGGAGAGAUCCAGAGA UGGGAUUCUAGGACAAUUGACCUGGACCAGACUCCACAGGGUUUCAAACAGUC CCACCCUGUUUAUAGAGGCACUGCACAGAGACCUAGCAGACUUCGGGAUCCAGCACC CAGACUUGAUCCUGCUACAGUACGUGGAUGACUACUGCUGGCCCGCCAUUCUGAGC UAGACUGCCAAACAGGUACUCGGGCCCCUGUUAACAAACCCUAGGGAACCUCCGGUAUC GGGCCUCGGCCAAGAAAGCCCAAUUUGCCAGAAACAGGUCAAGUAUCUGGGGUAUC
--	--

	UUCUAAAAGAGGGUCAGAGAUGGCUGACUGAGGCCAGAAAAGAGACUGUGAUGGGGC AGCCUACUCCGAAGACCCCUCGACAACUAAGGGAGUUCUAGGGAAGGCAGGCUUCU GUCGCCUCUUCAUCCCUGGGUUGCAGAAAUGGCAGCCCCCUGUACCCUCACCA AACCGGGGACUCUGUUUAAUUGGGGCCAGACCAACAAAAGGCCUAUCAAGAAAUCA AGCAAGCUCUGCUAACUGCCCCAGCCCUGGGUUGCCAGAUUUGACUAAGCCCCUUG AACUCUUUGUCGACGAGAAGCAGGGCUACGCCAAAGGUGUCCUAAACGCAAAACUGG GACCUUGGCUGCGCCGGUGGCCUACCUGUCCAAAAAGCUAGACCCAGUAGCAGCUG GGUGGCCCCUUGCCUACGGAUGGUAGCAGCCAUUGCCGUACUGACAAAGGAUGCAG GCAAGCUAACC AUGGGACAGCCACUAGUCAUUCUGGCCCCCCAUGCAGUAGAGGCAC UAGUCAAAACAACCCCCGACCGCUGGCUUCCAACGCCCGGAUGACUCACUAUCAGG CCUUGCUUUUGGACACGGACCGGGUCCAGUUCGGACCGGUGGUAGCCUGAACCCGG CUACGUGCUCCCACUGCCUGAGGAAGGGCUGCAACACAACUGCCUUGAUUACCCUGG CCGAAGCCCACGGAACCCGACCCGACCUAACGGACAGCCGCUCCAGACGCCGACC ACACCUGGUACACGGAUGGAAGCAGUCUCUUACAAGAGGGACAGCGUAAGGCGGGAG CUGCGGUGACCACCGAGACCGAGGUAUUCUGGGCUAAAGCCCUGCCAGCCGGGACAU CCGCUACAGCGGGCUGAACUGAUAGCACUACCCAGGCCCUAAAGAUGGCAGAAGGUA AGAAGCUAAAUGUUUAUACUGAUAGCCGUUAUGCUUUUGCUACGCCCAUUAUCCAG GAGAAAUUAUCAGAAGGCGUGGGUGGCUCACAUAGAGGCAAGAGAUCAAAAAUA AAGACGAGAUUCUUGGCCUACUAAAAGCCUUCUUCUGCCAAAAGACUUAAGCAUAA UCCAUUGUCCAGGACAUCAAAGGGACACAGCGCCGAGGCUAGAGGCAACCGGAUGG CUGACCAAGCGGCCGAAAGGCAGCCAUACAGAGAUCCAGACACCUCUACCCUCC UCAUAGAAAUUAUCACCCAAAAGGCCGGCGGCCACGAAAAGGCCGGCCAGGCAA AAAAGAAAAGUAAUAAGCGGCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA (SEQ ID NO: 243)
CRISPR Nickase- Reverse Transcriptase mRNA	GGAGAGACCCAAGCUGUCUAGACGCCACCAUGAAACGGACAGCCGACGGAAGCGAGU UCGAGUACCAAGAAGAAGCGGAAAGUCAACAAGCCUACAGCAUCGGCCUGGAUA UCGGCACCAACAGCGUGGGAUGGUCUAUCAACAGAUACUACAAGGUGCCUGCCA AGAAAAUGAGAGUGUUAGGCAACACCGACAAGGAGUACAUAAGAAGAACUGAUUCG GCGCCUCCUGUUCGACGGCGGAAACACAGCCGCCGAUCGGCGGCUGAAAAGGACCG CCAGACGGAGAUUACGCGGCGGAGAAUAGAAUCCUGUACUACAGGAAAUUCUUCG CUGAGGAGAUAGAGCAAUGGAGCAGACUCUUUUUCCACAGACUGGAAGAUAGCUUU UGGUGGAAGAGGACAAGCGAGGCUCUAAGUACCCUAUCUUCGCUACACUGCAAGAGG AAAAGGACUACCACGAGAAGUUUCCACCAUCUACCACCUGAGGAAAGAGCUGGCGG ACAAACAAGAGAAGGCCGACCUAGAGACUGAUUACCUUGCCUGGCUCAUAUCAUA AGUUCAGAGGCCACUCCUGAUCGAGGACGACAGAUUCGACGUGAGAAUACAGAUUA UCCAGAAGCAGUACCAGGCCUUUCUGGAGAUUCUACACACCACCUUUGAAAACAACC ACCUGCUGAGCCAAAACUGGAUGUGGAAGCCAUCUUAACCGAUAGAUUCCAAGA GCGCCAAGAAAGAUAGAAUCCUCGCCAGUAUCCUAACCGAAGAACCGGUAUCU UCGCCGAUUCUGAAGCUGAUCGUGGGCAACCAGGCCGACUUCAAAAACACUUUA ACCUGGAGGACAAAACACCUCUGCAGUUUGCCAAGGACUCCUACGACGAAGAUUCUG AAAACCUGCUGGGCCAGAUCCGCGACGAGUUCGCCGAUCUGUUCAGCGCCGCUAAGA AACUGUACGACAGCGUGCUGCUGAGCGGCAUCUUGACAGUGAUCGAUCUGUCCACAA AGGCCCCCCUGAGCGCCAGCAUGAUCCAGAGAUACGACGAGCACAGAGAGGAUCUGA AGCAGCUGAAGCAGUUUGAAGGCCUCUCUGCCCCGAGAAGUACCAGGAGAUUUUUG CGGACUCCAGCAAGGAUGGAUACGCCGGAUACAUUGAGGGCAAGACCAACAGGAAG CUUUCUACAAGUACCGUCUAAACUGCUGACCAAACAGGAAGAUUCCGAGAACUCC UGGAGAAAAUCAAGAACGAAGAUUUCUGAGAAAGCAGAGGACCUUCGACAACGGCA GCAUCCCCCACCAGGUCCACCUGACCGAGCUGAAGGCUAUCUACAGACGACAGAGUG AGUACUACCCUUCUGAAGGAAAAUCAGACAGAAUCGAGAAGAUUCUGACAUUCA GAAUCCCAUACUAUUCGGACCUUCUGGCCGUGAAAAGAGCGACUUCGCCUGGAUGA CCAGAAAGACCGAUGAUUCUAUCCGGCCUGGAACUUCGAGGACCUUGGACAAAGG AAAAGUCUGCCGAGGCCUUAUCCACCGAUGACCAACAACGACUUCUACGCCUG AAGAGAAGGUGCUGCCUAAACAUUCCUGAUCUACGAAAAGUUCACAGUGUACAACG AGCUGACCAAGGUAGAUACAAGAACGAGCAGGGCGAGACGUACUUCUUCGACUCUA

	AUAUCAAGCAGGAAAUCUUCGAUGGAGUGUUCAAGGAGCACCGGAAGGUCUCUAAGA AGAAACUGUUAGACUUCUGGCUAAGGAAUACGAGGAAUCCGGAUCUGGACGUGA UCGGACUGGAUAAGGAGAACAAAGCCUUUAACGCCAGCCUGGGAACAUACCACGACC UGGAGAAGAUCCUCGACAAGGACUUCUGGAUAACCCCGACAAUGAGUCCAUCCUGG AAGAUUUCUGCAGACGCGUGACACUGUUCGAGGACAGAGAGAUGAUAAGAAGAGAC UGGAAAACUACAAGGAUCUGUUCACCGAGAGCCAGCUGAAGAAGUUGUAUAGACGGC ACUACACCGGAUGGGGCCGGCUGUCCGCUAAACUGAUAACGGCAUCAGAGAUAAAGG AAAGUCAGAAAACCAUCCUGGACUACUUAUUCGACGAUGGCAGAAGCAACCGGAACU UCAUGCAGCUGAUAACGAUGAUGGCUUGAGCUUCAAGUCUAUCAUCUCUAAGGCC AGGCCGGAAGCCACAGCGAUAAACUGAAGGAAGUGGUGGGUGAGCUGGCCGGCUCUC CUGCCAUAAGAAGGGCAUCCUGCAAAGUCUAAAGAUUGUCGACGAGCUGGUAAGG UGAUGGGCUAUGAACCGAGCAGAUUCGUGGUGGAAAUGGCCAGAGAAAACAGACAA CCAAUCAAGGCAGAAGAAACAGCAGACAGCGGUACAAGCUGCUGGACGACGGCGUGA AGAAUCUGGCCAGCGACCUGAACGGCAACAUCUGAAAGAUACCCAACCGACAACC AGGCCUCGCAAGACGAGCGCCUGUUCUGUACUACCUGCAGAACGGCAGAGACAUGU ACACCGGCGAAGCCCUGGAUAUCGAUAACCGAGCCAGUACGAUAUCGACGCCAUAA UCCCAAGGCCUUUAUCAAGGAUGAUUCUAUUGAUAUUCGGGUGCUUGUCCAGCG CGAAGAACAGAGGAAAGAGCGACGACGUGCCAGCCUCGAAAUCGUGAAAGACUGCA AGGUGUUCUGGAAGAAGCUCUGGACGCCAAGCUGAUGAGCCAGAGAAAGUACGACA ACCUGACCAAGGCCGAGAGAGGAGGCCUGACCAGCGACGAUAAGGCUAGAUUCAUCC AAAGGCAGCUGGUCGAAACCAGACAGAUACCAAGCAGUGGCCAGGAUCCUCGACG AGCGGUUCAUAACGAGCUGGACAGCAAAGGUAGAAGAAUAGAAAGGUUAAGAUCG UCACCCUGAAGUCUAUUCUUGUGUCCAACUUCGGGAAGGAGUUCGGCUUCUACAAGA UCAGAGAGGUGAAGACUACCAACACCGCCCAUGAUGCCUACCGAACCGCGUGGUGG CCAAAGCCAUCCUGACCAAAUACCCUCAACUGGAGCCCAGAUUCGUGUACGGCCUGU ACAGACGGAAGAAGCUGAGCAAAAUCGUGCGGGAAGAUAGGAAGAGAAGUACAGCG AAGCCACAAGGAAAAUGUUUUUUAUAGCAACCUGAUGAACAUUGUUAAGAGAGUGG UGAGACUGGCCGACGGAUCUAUCGUCGUGCGGCCUGUGAUCGAAACCGGCAGAUACA UGGGCAAAACCGCCUGGGACAAAAGAAGCACUUCGCCACCGUGCGGAAGGUACUGA GCUACCCUCAGGUGAACAUAUGUGAAAAAGACAGAGAUCCAGACCGCGGCUUCUCUA AAGAGUCCAUCCUCGUCACGCGCAAUAGCGACAAGCUGAUCCCCAGAAAGACAAAAG ACAUCUACCUGGACCCUAAGAAGUAUGGCGGCUUCGACAGCCCCAUCGUGGCUUACA GCGUCCUGGUUGUGGCCGAUAUUAAGAAGGGCAAGGCUAGAAGCUGAAGACGGUGA CAGAGCUGCUGGGCAUAACAUAUGGAACGCAGCAAAUUCGAGAAGAACCAGCG CUUUUCUGGAAAGCAAGGGCUACCUGAACAUCCGGGACGAUAAGCUGAUGAUCCUGC CUAAGUACAGCCUGUUCGACUGGAAAUGGCAGACGGCGGCGUGCUAGCCAGCGCGG GAGAGCUGCAAAAGGGAAAUGAGCUGGCCUGCCUACCCAGUUAUGAAAUUCUGU AUCUGGCCUCUAGAUUAACGAGAGCAAGGGCAAACCUGAAGAGAUCGAAAAGAAAC AGGAGUUUGUAACCAACACGUGAGCUACUUCGACGACAUCUGCAGCUGAUAUAG ACUUCAGCAAGAGGGUGAUCCUGGCCGACGCCAACCUGGAAAAAUAACAAACUAU AUCAGGACAACAAGGAGAAUUCUCCUGGACGAACUGGCUAACAACUAUUAUUC UGUUCACAUUCACCAAGCCUUGGCGCUCUCCUGCCGCCUUUAAGUUCUUCGACAAAAUCG UGGACCGCAAGCGGUACACCAGCACCAAGGAGGUGCUUAACAGCACACUGAUCCACC AGAGCAUCACUGGCCUGUAUGAGACAAGAAUCGACCUGUCUAAGCUGGGCGAGGAUU CUGGAGGAUCUAGCGGAGGAUCCUCUGGCAGCGAGACACCAGGAACAAGCGAGUCAG CAACACCAGAGAGCAGUGGCGGCAGCAGCGGCGGCAGCAGCACCCUAAAUAUAGAAG AUGAGUAUCGGCUACAUGAGACCUCAAAGAGCCAGAUGUUUCUAGGGUCCACAU GGCUGUCUGAUUUUCCUCAGGCCUGGGCGGAAACCGGGGGCAUGGGACUGGCAGUUC GCCAAGCUCCUCUGAUAUACCUUCUGAAAGCAACCUCUACCCCGUGUCCAUAUAAAC AAUACCCCAUGUCACAAGAAGCCAGACUGGGGAUCAAGCCCCACAUACAGAGACUGU UGGACCAGGGAAUACUGGUACCCUGCCAGUCCCCUGGAACACGCCCUGCUACCCG UUAAGAAACCAGGGACUAAUGAUUAUAGGCCUGUCCAGGAUCUGAGAGAAGUCAACA AGCGGGUGGAAGACAUCCACCCACCGUGCCCAACCCUUAACAACCUUGAGCGGGC UCCACCGUCCCAAGUAGUGUACACUGUGCUUGAUUUAAAGGAUGCCUUUUUCUGCC UGAGACUCCACCCACAGUCAGCCUCUCUUCGCCUUUGAGUGGAGAGAUCCAGAGA UGGGAUUCACAGGACAAUUGACCUGGACCAGACUCCACAGGGUUUCAAACACAGUC
--	--

	CCACCCUGUUUAAUGAGGCACUGCACAGAGACCUAGCAGACUCCGGAUCCAGCACC CAGACUUGAUCCUGCUACAGUACGUGGAUGACUUACUGCUGGCCGCCACUUCUGAGC UAGACUGCCAAACAAGGUACUCGGGGCCUGUUACAAACCCUAGGGAACCUCGGGUUUC GGGCCUCGGCCAAGAAAGCCCAAUUUGCCAGAAACAGGUCAAGUAUCUGGGGUUUC UUCUAAAAGAGGGUCAGAGAUAGGCGUACUGAGGCCAGAAAAGAGACUGUGAUGGGGC AGCCUACUCCGAAGACCCUCGACAACUAAGGGAGUCCUAGGGAAGGCAGGCUUCU GUCGCCUCUUCUACCCUGGGUUUGCAGAAAUGGCAGCCCCCUGUACCCUCUCACCA AACCGGGGACUCUGUUUAAUUGGGGCCAGACCAACAAAAGGCCUAUCAAGAAAUCA AGCAAGCUCUGCUAACUGCCCCAGCCUGGGGUUGCCAGAUUUGACUAAGCCCUUUG AACUCUUUGUCGACGAGAAGCAGGGCUACGCCAAAGGUGUCCUAACGCAAAAACUGG GACCUUGGCGUCGGCCGGUGGCCUACCUGUCCAAAAGCUAGACCCAGUAGCAGCUG GGUGGGCCCCUUGCCUACGGAUGGUAGCAGCCAUUGCCGUACUGACAAAGGAUGCAG GCAAGCUAACCAUGGGACAGCCACUAGUCAUUCUGGCCCCCCAUGCAGUAGAGGCAC UAGUCAACAACCCCCGACCGCUGGCUUCCAACGCCCGGAUGACUCACUAUCAGG CCUUGCUUUUGGACACGGACCGGGUCCAGUUCGGACCGGUGGUAGCCCUGAACCCGG CUACGCUGCUCCACUGCCUGAGGAAGGGCUGCAACACAACUGCCUUGAUUACCCUGG CCGAAGCCACGGAACCCGACCCGACCUAACGGACAGCCGCUCCAGACGCCGACC ACACCUGGUACACGGAUGGAAGCAGUCUCUACAAGAGGGACAGCGUAAGGCGGGAG CUGCGGUGACCACCGAGACCGAGGUAAUCUGGGCUAAAGCCCUGCCAGCCGGGACAU CCGCUCAGCGGGCUGAACUGAUAGCACUCACCCAGGCCCUAAAGAUGGCAGAAGGUA AGAAGCUAAAUUUUAUACUGAUAGCCGUUAGCUUUUGCUACUGCCCAUAUCCAUG GAGAAAUUACAGAAGGCUGGGUGGCUACAUCAGAAGGCAAAGAGAUCAAAAUA AAGACGAGAUUCUUGGCCCUACUAAAAGCCUUCUUCUGCCAAAAGACUUAGCAUAA UCCAUUGUCCAGGACAUCAAAAAGGGACACAGCGCCGAGGCUAGAGGCAACCGGAUGG CUGACCAAGCGGCCCGAAAGGCAGCCAUACAGAGACUCCAGACACCUCUACCCUCC UCAUAGAAAAUUAUCACCCAAAAGGCCGGCGGCCACGAAAAGGCCGGCCAGGCAA AAAAAGAAAAAGUAAUAAGCGGCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA (SEQ ID NO: 244)
--	--

Editing template RNAs 1 and 2 (**Table 25**) were designed to install a 3-nucleotide insertion at the EMX1 and VEGFA target genes. The editing template RNAs were ordered as desalted synthetic guides from GenScript with the following chemical modifications: 2'-O-methyl for the first three and last three bases and phosphorothioate bonds between the first
5 three and last three bases, as shown in bold in the Guide (RNA) column of **Table 25**.

Table 25. Editing Template RNAs Sequences.

Editing Template RNA	Spacer	Scaffold	RTT	PBS	DNA Encoding Editing Template RNA	Editing Template RNA Sequence
1	GAGCC AGUGU UGCUA GUCAA (SEQ ID NO: 245)	GUUUUA GAGCUG UGCUGA AAAGCA CAGCAC GUUAAA AUAAGG	GCAUG CUGCC CUUGA AG (SEQ ID NO: 247)	ACUA GCAA CACU G (SEQ ID NO: 248)	GAGCCAGTGT GCTAGTCAAGT TTTAGAGCTGT GCTGAAAAGCA CAGCACGTAA AATAAGGCAGT GATTGAAAAAT	mG*mA*mG *CCA GUGUUGCUAGU CAAGUUUUAGA GCUGUGCUGAA AAGCACAGCAC GUUAAAAUAAG GCAGUGAUUGA

		CAGUGA UUGAAA AAUCCA GUCCGU AUUCAG CUUGAA AAAGUG AGCACC GAAUCG GUGC (SEQ ID NO: 246)			CCAGTCCGTATT CAGCTTGAAAA AGTGAGCACCG AATCGGTGCGC ATGCTGCCCTTG AAGACTAGCAA CACTG (SEQ ID NO: 249)	AAAAUCCAGUC CGUAUUCAGCU UGAAAAAGUGA GCACCGAAUCG GUGCGCAUGCU GCCCUUGAAGA CUAGCAACA*mC *mU*mG (SEQ ID NO: 250)
2	UUUUA AACUC UCCAU GGACC (SEQ ID NO: 251)	(SEQ ID NO: 246)	UGGAU GAGCC UGGUA AG (SEQ ID NO: 252)	CCAU GGAG AGUU U (SEQ ID NO: 253)	TTTTAAACTCTC CATGGACCGTT TTAGAGCTGTG CTGAAAAGCAC AGCACGTTAAA ATAAGGCAGTG ATTGAAAAATC CAGTCCGTATT CAGCTTGAAAA AGTGAGCACCG AATCGGTGCTG GATGAGCCTGG TAAGCCATGGA GAGTTT (SEQ ID NO: 254)	mU*mU*mU*UAA ACUCUCCAUGG ACCGUUUAGA GCUGUGCUGAA AAGCACAGCAC GUUAAAAUAAG GCAGUGAUUGA AAAAUCCAGUC CGUAUUCAGCU UGAAAAAGUGA GCACCGAAUCG GUGCUGGAUGA GCCUGGUAAGC CAUGGAGAG*m U*mU*mU (SEQ ID NO: 255)

* = phosphorothioate (PS) bond, mN = 2'-O-methyl modified base

PHH cells from human donors (from Thermo Fisher Scientific or Lonza) were thawed from liquid nitrogen quickly in a 37°C water bath. The cells were added to pre-warmed hepatocyte recovery media (Thermo Fisher Scientific, CM7000) and centrifuged. The cell pellet was resuspended in an appropriate volume of William's E Medium (Thermo Fisher Scientific) supplemented with Hepatocyte Plating Supplement Pack (serum-containing) (Thermo Fisher Scientific). The cells were counted using a trypan blue viability count and a Vi-CELL BLU cell counter. The desired number of viable cells were then washed in PBS and resuspended in P3 buffer + supplement (Lonza, V4SP-3096) and transfection enhancer oligo.

Resuspended cells were dispensed into Lonza 96-well electroporation plates. mRNA effector (1 mg/mL in water) was mixed with synthetic RNA guides (1 mM in water) at a 1:1 volume ratio. mRNA/guide RNA mixtures were added to each reaction at a final mRNA concentration of 25 nM. The plate was electroporated using an electroporation device (program DS-150, Lonza 4D-nucleofector). Following electroporation, pre-warmed Hepatocyte plating medium was added to each well and mixed very gently. For each technical replicate plate, 125,000 cells

of diluted nucleofected cells were plated into a pre-warmed collagen-coated 96-well plate (Thermo Fisher Scientific) containing Hepatocyte plating medium. The cells were then incubated at 37°C. After 4 hours, the media was changed to hepatocyte maintenance media (Williams' Medium E, Thermo Fisher Scientific) supplemented with William's E Medium Cell Maintenance Cocktail, Thermo Fisher Scientific).

3 days post electroporation, cells in the wells were harvested using Accutase (Thermo Fisher Scientific) and transferred to 96-well twin.tec® PCR plates (Eppendorf) and centrifuged. Media was flicked off, and cells were resuspended in DNA extraction buffer (QuickExtract™). Samples were cycled in a PCR machine at 65°C for 15 min, 68°C for 15 min, and 98°C for 10 min. Samples were then frozen at -20°C and subsequently analyzed via NGS as described in **Example 1**.

Edit incorporation in PHH using an mRNA (SEQ ID NO: 243) encoding the CRISPR nuclease-reverse transcription fusion polypeptide of SEQ ID NO: 55 or an mRNA (SEQ ID NO: 244) encoding the CRISPR nickase-reverse transcription fusion polypeptide of SEQ ID NO: 57 is shown in **Table 26** or **Table 27**, respectively. For the CRISPR nuclease-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 12.6% to 13.2%, while the average percentage of NGS reads comprising 3-nucleotide insertion installation ranged from 0.57% to 0.95% (**Table 26**). For the CRISPR nickase-reverse transcription fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 0.38% to 2.71%, while the average percentage of NGS reads comprising 3-nucleotide insertion installation ranged from 0.17% to 1.05% (**Table 27**). Overall, the use mRNA encoding the CRISPR nickase-reverse transcriptase fusion polypeptide and mRNA encoding the CRISPR nuclease-reverse transcriptase fusion polypeptide resulted in similar levels of 3-nucleotide insertion installation at the target loci.

Table 26. Editing Efficiencies with mRNA Encoding CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptide in PHH

Editing Template	Average % Indel	Average % 3-nt insertion installation
1	13.2	0.57
2	12.6	0.95

Table 27. Editing Efficiencies with mRNA Encoding CRISPR Nickase-Reverse Transcriptase Fusion Polypeptide in PHH

Editing Template	Average % Indel	Average % 3-nt
------------------	-----------------	----------------

		insertion installation
1	0.38	0.17
2	2.71	1.05

This Example thus shows that the abovementioned mRNAs combined with synthetic guides edit non-dividing human cells.

5 **Example 12: RNA-Templated Editing of Mouse Genes in Mice Using CRISPR Nuclease-Reverse Transcriptase and CRISPR Nickase-Reverse Transcriptase Fusion Polypeptides**

This Example shows genetic modification of the DNMT1 gene in mice utilizing mRNA encoding a CRISPR Nuclease-Reverse Transcriptase fusion polypeptide or a CRISPR Nickase-Reverse Transcriptase fusion polypeptide.

10 The mRNA molecules provided in **Table 24** of **Example 11** were used. Editing template RNAs provided in **Table 28** below were designed to install a G>C substitution or a CCC insertion into the DNMT1 locus. A 12-nucleotide RTT was used to install the G>C substitution (Editing Template RNA 1), and a 15-nucleotide RTT was used to install the CCC insertion (Editing Template RNA 2). The editing template RNAs provided in **Table 28** were

15 tested in the presence of the RNA guide provided in **Table 29**. These editing template guide RNAs and RNA guides were ordered as HPLC-purified synthetic guides from GenScript with the following chemical modifications: 2'-O-methyl for the first three and last three bases, and phosphorothioate bonds between the first three and last three bases as shown in bold in the Guide (RNA) column of **Table 28**.

20 **Table 28. Sequences of Editing Template RNAs and Components Thereof**

Editing Template RNA	Spacer (RNA)	Scaffold	RTT	PBS	Guide (DNA)	Guide (RNA)
1	CGGGCU GGAGCU GUUCGC GC (SEQ ID NO: 256)	SEQ ID NO: 246 See Table 25 above	AAGAUG G CAGCG (SEQ ID NO: 257)	CGAAC AGCUC CAG (SEQ ID NO: 258)	CGGGCTGGAGCTG TTCGCGCGTTTTA GAGCTGTGCTGAA AAGCACAGCACGT TAAAATAAGGCAG TGATTGAAAAATC CAGTCCGTATTCA GCTTGAAAAAGTG AGCACCGAATCGG TGCAAGATGGCAG	mC*mG*mG *GCUGG AGCUGUUCGCGCGU UUUAGAGCUGUGCU GAAAAGCACAGCAC GUUAAAAUAAGGCA GUGAUUGAAAAAUC CAGUCCGUUUUCAG CUUGAAAAAGUGAG CACCGAAUCGGUGC AAGAUGGCAGCGCG

					CGCGAACAGCTCC AG (SEQ ID NO: 259)	AACAGCUC*mC*mA* mG (SEQ ID NO: 260)
2	SEQ ID NO: 256	SEQ ID NO: 246 See Table 25 above	AAGAUG CCAGCG GGG (SEQ ID NO: 261)	SEQ ID NO: 258	CGGGCTGGAGCTG TTCGCGCGTTT GAGCTGTGCTGAA AAGCACAGCACGT TAAAATAAGGCAG TGATTGAAAAATC CAGTCCGTATTCA GCTTGAAAAAGTG AGCACCGAATCGG TGCAAGATGCCAG CGGGGCGAACAGC TCCAG (SEQ ID NO: 262)	mC*mG*mG*GCUGG AGCUGUUCGCGCGU UUUAGAGCUGUGCU GAAAAGCACAGCAC GUUAAAAUAAGGCA GUGAUUGAAAAAUC CAGUCCGUUUUCAG CUUGAAAAAGUGAG CACCAGAACGUGUG AAGAUGCCAGCGGG GCGAACAGCUC*mC* mA*mG (SEQ ID NO: 263)

* = phosphorothioate (PS) bond, mN = 2'-O-methyl modified base

Table 29. RNA guide sequence

DNA sequence	RNA sequence
CCGCGCGCGCGAAAAAGCCGGTTT TAGAGCTGTGCTGAAAAGCACAGC ACGTTAAAATAAGGCAGTGATTGA AAAATCCAGTCCGTATTCAGCTTGA AAAAGTGAGCACCGAATCGGTGC (SEQ ID NO: 264)	mC*mC*mG*CGCGCGCGAAAAAGCCGG UUUUAGAGCUGUGCUGAAAAAGCACAGC ACGUAAAAUAAGGCAGUGAUUGAAA AAUCCAGUCCGUUUUCAGCUUGAAAA GUGAGCACCGAACGCG*mU*mG*mC (SEQ ID NO: 265)

* = phosphorothioate (PS) bond, mN = 2'-O-methyl modified base

- The following components were loaded into lipid nanoparticles (LNPs):
- 5 a) the mRNA (SEQ ID NO: 243) encoding the CRISPR nuclease-reverse transcriptase (SEQ ID NO: 55) or the mRNA (SEQ ID NO: 244) encoding the CRISPR nickase-reverse transcriptase fusion polypeptides (SEQ ID NO: 57),
- b) the editing template RNA provided in **Table 28**, and
- c) the RNA guide provided in **Table 29**.
- 10 The LNPs contained 46.3% cationic lipid 6-((2-hexyldecanoyl)oxy)-N-(6-((2-hexyldecanoyl)oxy)hexyl)-N-(4-hydroxybutyl)hexan-1-aminium, 9.4% phospholipid 1,2-Distearoyl-*sn*-glycerol-3-phosphocholine (DSPC), 42.7% cholesterol, and 1.6% PEG lipid 2-[(polyethylene glycol)-2000]-N,N ditetradecylacetamide and were formulated with a Molar N/P ratio of ~6. The LNPs were prepared according to the general procedures described in
- 15 Schoenmaker, IJPharm, 601:120586, 2021, the relevant disclosures of which are incorporated by reference herein for the subject matter and purpose referenced herein.
- Male, C57BL6 mice (6-weeks of age, Jackson Laboratories, Bar Harbor, ME) were

used for these studies. Animals were acclimated to the housing facility for a minimum of 3 days prior to study start. Animals were weighed prior to dosing. The study 1 groups were treated with mRNA encoding the CRISPR nickase-reverse transcriptase fusion polypeptide. The study 2 groups were treated with mRNA encoding the CRISPR nuclease-reverse transcriptase fusion polypeptide. The ratio column in **Table 30** and **Table 31** refer to the ratio of the mRNA encoding the fusion polypeptide to the editing template RNA to the RNA guide. On study day 0, editing templates formulated in LNPs were dosed intravenously via retro-orbital injection in a volume of 200 μ l.

10 **Table 30: Study 1 Conditions (CRISPR Nickase-Reverse Transcriptase Fusion Polypeptide)**

Group #	Editing Template RNA	# Mice	Dose	Ratio	Admin Route	Freq.	Dose Day
1	1	5	2 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
2	1	5	3.5 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
3	1	5	5 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
4	2	5	2.5 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
5	2	5	5 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
6	Saline control	5	N/A	N/A	Retro-orbital	1x	D0

15 **Table 31: Study 2 Conditions (CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptide)**

Group #	Editing Template RNA	# Mice	Dose	Ratio	Admin Route	Freq.	Dose Day
1	1	5	3.5 mg/kg	1:0.9:0.1	Retro-orbital	1x	D0
2	2	5	5 mg/kg	2:0.75:0.25	Retro-orbital	1x	D0
3	Saline control	5	N/A	N/A	Retro-orbital	1x	D0

7 days after LNP dosing, animals were euthanized and perfused with PBS. The left liver lobe pieces comprising of 30-50 mg of tissue were collected and frozen on dry ice for tissue processing. The tissue was processed in Quick Extract™ buffer (Lucigen®) with 5 mm beads, with one bead per sample using a TissueLyser II, then resuspended in Quick Extract™ buffer (Lucigen®). Cells were incubated at 65°C for 15 minutes and 98°C for 2 minutes.

Samples were prepared for NGS and analyzed as described in **Example 1**. The percentage of NGS reads containing indels or edits encoded by the editing template RNAs and incorporated into the DNMT1 locus were quantified. Results are shown in **Table 32** and **Table 33**.

Table 32: Study 1 Results (CRISPR Nickase-Reverse Transcriptase Fusion Polypeptide)

Group #	Editing Template RNA	Average % Indel	Average % Edit Installation
1	1	0.42	2.21
2	1	0.63	3.90
3	1	0.79	5.41
4	2	0.85	1.13
5	2	0.98	2.46
6	Saline control	0.38	0.08

Table 33: Study 2 Results (CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptide)

Group #	Editing Template RNA	Average % Indel	Average % edit installation
1	1	5.92	0.186
2	2	13.02	1.12
3	Saline control	0.78	0.04

For study 1 with the CRISPR nickase-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 0.42% to 0.98%, while the average percentage of NGS reads comprising edit installation ranged from 1.13% to 5.41% (**Table 32**). For both editing template RNAs, there was a dose-dependent increase in incorporation of encoded edits (**Table 32**). For study 2 with the CRISPR nuclease-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 5.92% to 13.02%, while the average percentage of NGS reads comprising edit installation ranged from 0.186% to 1.12% (**Table 33**). Overall, the use of the CRISPR nickase-reverse transcriptase fusion polypeptide resulted in higher edit installation at the DNMT1 locus compared to the CRISPR nuclease-reverse transcriptase fusion polypeptide.

Therefore, this Example shows that edits were capable of being installed by the CRISPR nuclease-reverse transcriptase and CRISPR nickase-reverse transcriptase fusion polypeptides at the DNMT1 locus in mice.

5 OTHER EMBODIMENTS

All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

10 From the above description, one skilled in the art can easily ascertain the essential characteristics of the present invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

15 EQUIVALENTS

While several inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

All definitions, as defined and used herein, should be understood to control over

dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

5 All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

10 The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when
15 used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

20 As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one
25 element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

30 As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also

allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

WHAT IS CLAIMED IS:

1. A gene editing system comprising:
 - (a) a fusion polypeptide comprising a CRISPR nuclease polypeptide and a reverse transcriptase (RT) polypeptide, or a first nucleic acid encoding the fusion polypeptide; wherein the CRISPR nuclease polypeptide comprises the amino acid sequence of SEQ ID NO: 1, or is a variant of SEQ ID NO: 1, the variant comprising:
 - (i) one or more mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1 that reduce or eliminate the nuclease activity thereof;
 - (ii) one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions;
 - (iii) one or more mutations for reducing PAM recognition stringency; or
 - (iv) a combination of any of (i), (ii), and (iii);
 - (b) an RNA molecule comprising a guide RNA (gRNA) and a reverse transcription donor RNA (RT donor RNA), or a second nucleic acid encoding the RNA molecule;
 - wherein the gRNA comprises a scaffold sequence recognizable by the CRISPR nuclease and a spacer sequence specific to a target sequence within a genomic site of interest, the target sequence being upstream to a protospacer adjacent motif (PAM); and
 - wherein the RT donor RNA comprises a primer binding site (PBS) and a template sequence.
2. The gene editing system of claim 1, wherein the fusion polypeptide further comprises one or more nuclear localization signals (NLSs) upstream or downstream to the CRISPR nuclease polypeptide, the RT polypeptide, or both.
3. The gene editing system of claim 2, wherein the fusion polypeptide, from N-terminus to C-terminus, comprises a first NLS, the CRISPR nuclease polypeptide, the RT polypeptide, and a second NLS; optionally wherein the fusion polypeptide further comprises a peptide linker between the CRISPR nuclease polypeptide and the RT polypeptide.
4. The gene editing system of claim 2, wherein the fusion polypeptide comprises a first peptide linker located between the CRISPR nuclease polypeptide and the RT polypeptide,

and wherein the fusion polypeptide comprises a first NLS, a second NLS, which are located at the N-terminus and/or the C-terminus of the fusion polypeptide.

5 5. The gene editing system of claim 4, wherein the fusion polypeptide further comprises a third NLS and optionally a fourth NLS.

 6. The gene editing system of claim 4, wherein the fusion polypeptide further comprises a second peptide linker connecting the CRISPR nuclease polypeptide or the RT polypeptide, and the first or second NLS(s); optionally wherein the fusion polypeptide further
10 comprises a third peptide linker connecting two NLSs.

 7. The gene editing system of any one of claims 4-6, wherein the fusion polypeptide comprises, from N-terminus to C-terminus,
15 (i) the first NLS, the second NLS, the CRISPR nuclease polypeptide, the first peptide linker, and the RT polypeptide;
 (ii) the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the first NLS, and the second NLS;
 (iii) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third NLS, the second peptide linker, and the second NLS;
20 (iv) the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third NLS, the second NLS, the second peptide linker, and the first NLS;
 (v) the first NLS, the second peptide linker, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the third peptide linker, and the second NLS;
25 (vi) the first NLS, the second peptide linker, the CRISPR nuclease, the first peptide linker, the RT polypeptide, the third peptide linker, the third NLS, the fourth NLS, and the second NLS;
 (vii) the first NLS, the second peptide linker, the RT polypeptide, the first peptide linker, the CRISPR nuclease polypeptide, the third peptide linker, the third
30 NLS, the fourth NLS, and the second NLS;
 (viii) the RT polypeptide, the first peptide linker, the CRISPR nuclease polypeptide, the third NLS, the second NLS, the second peptide linker, and the first NLS;
 (ix) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the second peptide linker, the second NLS, and the RT polypeptide;

- (x) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the third NLS, the second peptide linker, the RT polypeptide, the third peptide linker, and the second NLS;
- (xi) the first NLS, the RT polypeptide, the first peptide linker, the second NLS, the second peptide linker, and the CRISPR nuclease polypeptide,
- (xii) the first NLS, the RT polypeptide, the first peptide linker, the third NLS, the second peptide linker, the CRISPR nuclease polypeptide, the third peptide linker, and the second NLS;
- (xiii) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, and the second NLS; or
- (xiv) the first NLS, the CRISPR nuclease polypeptide, the first peptide linker, the RT polypeptide, the second peptide linker, the third peptide linker, and the second NLS.
8. The gene editing system of claim 7, wherein the fusion polypeptide comprises, from N-terminus to C-terminus, (iv), (v), (vii), (ix), or (x).
9. The gene editing system of any one of claims 4-8, wherein the peptide linker(s) between the CRISPR nuclease polypeptide and the RT polypeptide is about 20-80 amino acids in length.
10. The gene editing system of any one of claims 1-9, wherein the CRISPR nuclease polypeptide is the variant of SEQ ID NO: 1.
11. The gene editing system of claim 10, wherein the CRISPR nuclease polypeptide is the variant of SEQ ID NO: 1, which comprises one or more mutations in the HNH nuclease domain at positions D844, H845, and/or N868 relative to SEQ ID NO: 1; optionally wherein the mutation is at position H845, which optionally is H845A.
12. The gene editing system of claim 11, wherein:
- (a) the mutation at D844 is an amino acid substitution of D844A, D844G, D844L, or D844S;
- (b) the mutation at H845 is an amino acid substitution of H845A, H845G, H845L, or H845S; and

(c) the mutation at N868 is an amino acid substitution of N868A, N868G, N868L, or N868S.

13. The gene editing system of any one of claims 10-12, wherein the CRISPR
5 nuclease polypeptide comprises a bridge helix (BH) domain, a nucleic acid recognition (REC) domain, a phosphate lock loop (PLL), a wedge (WED) domain, and a PAM-interacting (PID) domain, and wherein one or more arginine and/or lysine substitutions, optionally arginine substitutions, are located in the BH domain, in the REC domain, in the PLL domain, in the WED domain, in the PID domain, or a combination thereof.

10

14. The gene editing system of 10, wherein the CRISPR nuclease variant is any one of those listed in **Table 19** or **Table 20**.

15. The gene editing system of any one of claims 10-14, wherein the CRISPR
15 nuclease polypeptide contains up to 20 arginine and/or lysine substitutions relative to SEQ ID NO: 1; optionally wherein the CRISPR nuclease polypeptide contains up to 15 arginine and/or lysine substitutions relative to SEQ ID NO: 1; preferably wherein the one or more arginine and/or lysine substitutions are at positions K736, L784, Q812, N813, I857, and/or A919 of SEQ ID NO: 1, which optionally is at position I857, optionally I857R.

20

16. The gene editing system of claim 15, wherein the CRISPR nuclease polypeptide contains at least two arginine and/or lysine substitutions relative to SEQ ID NO: 1, and wherein the at least two arginine and/or lysine substitutions are at positions K736, L784, Q812, N813, I857, and/or A919 of SEQ ID NO: 1.

25

17. The gene editing system of claim 16, wherein the CRISPR nuclease polypeptide contains arginine and/or lysine substitutions at the following positions relative to SEQ ID NO:

1:

- 30
- (a) I857, L784, and K736;
 - (b) I857, A919, and K736;
 - (c) I857, N813, and L784;
 - (d) I857, L784, and A919;
 - (e) I857, N813, and K736;
 - (f) I857 and N813;

- (g) L784, A919, and K736;
- (h) I857 and L784; and
- (i) I857 and A919.

- 5 18. The gene editing system of claim 17, wherein the CRISPR nuclease polypeptide comprises the following arginine substitutions relative to SEQ ID NO: 1:
- (a) I857R, L784R, and K736R;
 - (b) I857R, A919R, and K736R;
 - (c) I857R, N813R, and L784R;
 - 10 (d) I857R, L784R, and A919R;
 - (e) I857R, N813R, and K736R;
 - (f) I857R and N813R;
 - (g) L784R, A919R, and K736R;
 - (h) I857R and L784R; and
 - 15 (i) I857R and A919R;
- optionally wherein the CRISPR nuclease polypeptide comprises the arginine substitutions of (a).

- 20 19. The gene editing system of claim 1, wherein the CRISPR nuclease polypeptide comprises a nickase mutation at position H845, optionally H845A, and an arginine and/or lysine substitution at position I857, optionally I857R relative to SEQ ID NO: 1.

- 25 20. The gene editing system of any one of claims 1-19, wherein the CRISPR nuclease polypeptide comprises the one or more mutations for reducing PAM recognition stringency, optionally wherein the one or more mutations are at positions D61, A68, H494, L1117, D1144, S1145, G1227, E1228, S1327, A1332, R1343, R1345, and/or T1347 of SEQ ID NO: 1.

- 30 21. The gene editing system of claim 20, wherein the one or more mutations comprise:
- (i) one or more arginine and/or lysine substitutions, optionally arginine substitutions, at position D61, A68, H494, L1117, G1227, S1327, A1332, and/or T1347 of SEQ ID NO: 1;
 - (ii) one or more amino acid substitutions at position D1144, S1145, E1228,

R1343, and/or R1345, of SEQ ID NO: 1, optionally D1144L, S1145W, E1228Q, R1343P, R1345V and/or R1345Q; or
(iii) a combination of (i) and (ii).

5 22. The gene editing system of claim 21, comprising the following combination of mutations relative to SEQ ID NO: 1:

- (i) L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and A68R;
- (ii) L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and D61R;

or

10 (iii) L1117R, D1144V, G1227R, E1228F, A1332R, R1345V, T1347R, and H494R.

23. The gene editing system of any one of claims 10-22, wherein the CRISPR nuclease polypeptide comprises (a) the one or more mutations in the HNH nuclease domain at positions D844, H845, and/or N868 relative to SEQ ID NO: 1; optionally wherein the mutation
15 is at position H845; and (b) one or more arginine and/or lysine substitutions relative to SEQ ID NO: 1, optionally wherein the arginine and/or lysine substitutions are at positions I857, L784, and K736.

24. The gene editing system of any one of claims 10-23, wherein the CRISPR
20 nuclease polypeptide comprises an amino acid sequence at least 90% identical to SEQ ID NO: 1.

25. The gene editing system of claim 24, wherein the CRISPR nuclease polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 1.
25

26. The gene editing system of claim 25, wherein the CRISPR nuclease polypeptide comprises an amino acid sequence at least 98% identical to SEQ ID NO: 1.

27. The gene editing system of any one of claims 1-26, wherein the RT polypeptide
30 is Moloney Murine Leukemia Virus (MMLV)-RT, optionally the MMLV-RT comprises the amino acid sequence of SEQ ID NO: 53.

28. The gene editing system of claim 1, wherein the fusion polypeptide is set forth in **Table 8** or **Table 17**.

29. The gene editing system of any one of claims 1-28, wherein the system comprises the fusion polypeptide.
- 5 30. The gene editing system of any one of claims 1-29, wherein the system comprises the first nucleic acid encoding the fusion polypeptide.
31. The gene editing system of claim 30, wherein the first nucleic acid is located on a vector, which optionally is a viral vector.
- 10 32. The gene editing system of claim 31, wherein the first nucleic acid is a first messenger RNA (mRNA).
33. The gene editing system of any one of claims 1-32, wherein the spacer sequence
15 in the gRNA of (b) is 15-30-nucleotide in length; optionally 15-20-nucleotide in length.
34. The gene editing system of any one of claims 1-33, wherein the PAM is 5'-NDR-3' or 5'-NGN-3', in which N represents any nucleotide, D represents A, G, or T, and R represents G or A; optionally wherein the PAM is 5'-NRG-3' or 5'-NRR'3', in which N and
20 Rare defined herein; preferably wherein the PAM is 5'-NGG-3', in which N represents any nucleotide.
35. The gene editing system of any one of claims 1-34, wherein the scaffold
25 sequence comprises a nucleotide sequence at least 85% identical to SEQ ID NO: 2.
36. The gene editing system of claim 35, wherein the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2.
37. The gene editing system of any one of claims 1-36, wherein the PBS in the RT
30 donor RNA is 5-50-nucleotide in length; optionally 5-20-nucleotide in length.
38. The gene editing system of any one of claims 1-37, wherein the PBS binds a PBS-targeting site that is adjacent to or overlaps with the target sequence.

39. The gene editing system of any one of claims 1-38, wherein the PBS-targeting site is adjacent to or overlaps with the target sequence.

40. The gene editing system of claim 39, wherein the PBS-targeting site is adjacent to the 5' of the PAM, optionally wherein the 3' end nucleotide of the PBS-targeting site is about 2-15 nucleotides upstream to the PAM.

41. The gene editing system of any one of claims 1-40, wherein the template sequence in the RT donor RNA is 5-100-nucleotide in length; optionally 15-25-nucleotide in length.

42. The gene editing system of any one of claims 1-41, wherein the template sequence in the RT donor RNA is homologous to the genomic site of interest and comprises one or more nucleotide variations relative to the genomic site of interest.

43. The gene editing system of claim 42, wherein at least one nucleotide variation is located within the target sequence; and/or wherein at least one nucleotide variation is located in the PAM.

44. The gene editing system of any one of claims 1-43, wherein the RNA molecule of (b) further comprises a 3' end extension.

45. The gene editing system of any one of claims 1-44, wherein the RNA molecule of (b) further comprises a 5' end protection fragment, a 3' protection fragment, or both, each of the 5' end protection fragment and the 3' end protection fragment forming a secondary structure, which optionally is a hairpin, a pseudoknot, a circularization, or a triplex structure.

46. The gene editing system of any one of claims 1-45, wherein the RNA molecule of (b) comprises, from 5' to 3':

(i) the spacer sequence, the scaffold sequence, the template sequence, and the PBS; or

(ii) the spacer sequence, the scaffold sequence, the template sequence, the PBS, and the 3' extension.

47. The gene editing system of any one of claims 1-46, wherein the system comprises the RNA molecule of (b).

48. The gene editing system of any one of claims 1-46, wherein the system
5 comprises the second nucleic acid encoding the RNA molecule.

49. The gene editing system of claim 48, wherein the nucleic acid is located on a vector, which optionally is a viral vector.

10 50. The gene editing system of any one of claims 1-49, wherein the system comprises one or more lipid nanoparticles (LNPs) associated with one or more of elements (a)-(b).

15 51. The gene editing system of any one of claims 1-50, wherein the system comprises one or more viral vectors, optionally one or more adeno-associated viral (AAV) vectors encoding one or more of elements (a)-(b).

20 52. A pharmaceutical composition comprising the gene editing system of any one of claims 1-51.

53. A kit comprising the elements (a)-(b) of the gene editing system set forth in any one of claims 1-51.

25 54. A gene editing method, comprising delivering the gene editing system of any one of claims 1-51 to a host cell to edit a genomic site targeted by the gRNA of the gene editing system.

55. The gene editing method of claim 54, wherein the host cell is cultured *in vitro*.

30 56. The gene editing method of claim 55, wherein the host cell is located in a subject who needs the gene editing.

57. A fusion polypeptide, comprising a CRISPR nuclease polypeptide set forth in any one of claims 1-26 and a reverse transcriptase polypeptide set forth in claim 1 or claim 27.

58. The fusion polypeptide of claim 57, which comprises the amino acid sequence of SEQ ID NO: 55 or 57.

5 59. A nucleic acid encoding the fusion polypeptide of claim 57 or claim 58.

60. The nucleic acid of claim 59, which comprises the nucleotide sequence of SEQ ID NO: 54, 243, 56, or 244.

10 61. The nucleic acid of claim 60, which is a vector, optionally an expression vector.