REVERSE TRANSCRIPTION-MEDIATED GENE EDITING SYSTEMS AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of the filing date of U.S. Provisional Application No. 63/570,055, filed March 26, 2024, the entire contents of which is incorporated by reference herein.

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 on March 24, 2025, is named 063586-539001WO_Seq-Listing_ST26.xml and is 386,301 bytes 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 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 systems has shown great potentials in genetic editing. CRISPR-guided reverse transcription allows for introduction of desired edits (*e.g.*, nucleotide substitutions, insertions, or deletions) 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.

SUMMARY OF THE INVENTION

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, and optionally one or more nuclear localization signals (NLS) and/or peptide linkers. The CRISPR nuclease polypeptide can be genetically engineered to possess advantageous enzymatic activities (*e.g.*, nickase activity, high indel activities and/or DNA cleavage activities and precise gene editing as designed). The gene editing systems provided herein would be expected to show

superior effectiveness in inserting desired base substitutions at a genomic site of interest.

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.

In the fusion polypeptide, the CRISPR nuclease may be the reference nuclease of SEQ ID NO: 1 or a variant thereof. The RT polypeptide can be any of those provided herein, for example, an MMLV-RT polypeptide.

In some embodiments, the CRISPR nuclease polypeptide contained in the gene editing system disclosed herein is a variant of SEQ ID NO: 1. In some examples, the variant may comprise: (i) one or more mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1; (ii) one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions; (iii) an N-terminal truncation, or (iv) a combination of (i), (ii) and/or (iii).

In some instances, the variant may comprise (i) one or more nickase mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1. The one or more nickase mutations are located at positions H397, D396, N420, D24, E337, and/or D524 of SEQ ID NO: 1. In some examples, the one or more nickase mutations are located at positions D396, H397, and/or N420 relative to SEQ ID NO: 1. For example, the mutation at D396 is an amino acid substitution of D396A, D396G, D396L, or D396S; the mutation at H397 is an amino acid substitution of H397A, H397G, H397L, or H397S; and the mutation at N420 is an amino acid substitution of N420A, N420G, N420L, or N420S. In some specific examples, the mutation is at position H397 (*e.g.*, H397A or H397L substitution). In one specific example, the mutation is at position H397, *e.g.*, H397A.

The CRISPR nuclease polypeptide disclosed herein 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. In some instances, the variant may comprise (ii) one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions, 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. In some examples, the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) are located at one or more of positions D56, E59, G60, E63, I67, G71, S564, D568, W593, T605, E655, E694, and E706 in SEQ ID NO: 1.

In some instances, the variant 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. For example, the variant CRISPR nuclease polypeptide comprises arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at the following positions relative to SEQ ID NO: 1:

a) I67, D568, and E706;

b) I67, D56, and D568;

c) I67, T605, and D568;

d) I67, D56, and E706;

e) I67, T605, and E706;

f) I67, D568, and E694;

g) I67, D56, E694, and E706;

h) I67, D56, T605, and E706; or

i) I67, D568, W593, and E706.

In one specific example, the variant CRISPR nuclease polypeptide comprises arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at I67, D568, and E706. In another specific example, the variant CRISPR nuclease polypeptide comprises arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at I67, D568, W593, and E706.

In some examples, the variant CRISPR nuclease polypeptide contains the arginine substitutions of (a) I67R, D568R, and E706R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (b) I67R, D56R, and D568R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (c) I67R, T605R, and D568R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (d) I67R, D56R, and E706R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (e) I67R, T605R, and E706R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (f) I67R, D568R, and E694R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (g) I67R, D56R, E694R, and E706R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (h) I67R, D56R, T605R, and E706R. In some examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of (i) I67R, D568R, W593R, and E706R. In some specific examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of I67R, D568R, and E706R. In other specific examples, the engineered CRISPR nuclease polypeptide contains the arginine substitutions of I67R, D568R, W593R, and

E706R.

Alternatively or in addition, the variant CRISPR nuclease polypeptide may have an N-terminal truncation relative to the reference CRISPR nuclease of SEQ ID NO: 1. In some examples, the N-terminal truncation is a deletion within residues 1-15 of SEQ ID NO: 1. In some specific examples, the N-terminal truncation is a deletion of residues 1-14 or 1-15 of SEQ ID NO: 1.

Alternatively or in addition, the variant CRISPR nuclease polypeptide may have a C-terminal truncation relative to the reference CRISPR nuclease of SEQ ID NO: 1 (*e.g.*, within the final 1-5 residues of SEQ ID NO: 1).

In some instances, the variant CRISPR nuclease polypeptide disclosed herein may comprise: (a) the one or more mutations in the HNH nuclease domain (*e.g.*, at position H397, D396, and/or N420 relative to SEQ ID NO: 1); and (b) the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at positions I67R, D568R, and E706R relative to SEQ ID NO: 1.

In some instances, the variant CRISPR nuclease polypeptide disclosed herein may comprise: (a) the one or more mutations in the HNH nuclease domain (*e.g.*, at position H397, D396, and/or N420 relative to SEQ ID NO: 1); (b) the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at positions I67R, D568R, and E706R relative to SEQ ID NO: 1; and (c) the N-terminal truncation within residues 1-15 of SEQ ID NO: 1. For example, the N-terminal truncation is the deletion of residues 1-14 or 1-15 of SEQ ID NO: 1. By way of non-limiting examples, a variant CRISPR nuclease polypeptide may comprise: (a) the nickase mutation of H397A; (b) the arginine substitutions of I67R, D568R, and E706R; or the arginine substitutions of I67R, D568R, W593R, and E706R; and (c) the N-terminal truncation comprising deletion of residues 1-15 of SEQ ID NO: 1.

Any of the CRISPR nuclease polypeptide in the fusion polypeptide disclosed herein may comprise an amino acid sequence at least 95% identical to SEQ ID NO: 1. In some examples, the CRISPR nuclease polypeptide may comprise an amino acid sequence at least 98% identical to SEQ ID NO: 1. In some instances, the CRISPR nuclease polypeptide is listed in **Table 1**, **Table 11**, or **Table 14**.

In some embodiments, the RT polypeptide in the fusion polypeptide may be a Moloney Murine Leukemia Virus (MMLV)-RT. In one example, the MMLV-RT comprises the amino acid sequence of SEQ ID NO: 66 or the amino acid sequence of SEQ ID NO: 176 (without the N-terminal M residue, *e.g.*, when the RT polypeptide is located at the C-terminal of the CRISPR nuclease polypeptide or when a functional fragment such as NLS is added to the N-terminus of

the RT polypeptide). Other exemplary MMLV-RT polypeptides are provided in **Table 25** below.

In some embodiments, the fusion polypeptide comprises one or more 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. For example, the fusion polypeptide may comprise a first NLS located at the N-terminus and a second NLS located at the C-terminus of the fusion polypeptide. Alternatively or in addition, the fusion polypeptide comprises a peptide linker located between the CRISPR nuclease polypeptide and the RT polypeptide.

In specific examples, the fusion polypeptide in any of the gene editing systems provided herein may be any one of those provided in **Tables 16**, **21**, and **22**.

The RNA molecule in the gene editing system provided herein comprises a guide RNA (gRNA) and a reverse transcription donor RNA (RT donor RNA). 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). In some embodiments, the PAM is 5'-NGG-3', in which N represents any nucleotide.

In some embodiments, the spacer sequence in the gRNA can be 15-30-nucleotide in length. In one example, the spacer sequence may be 15-20-nucleotide in length.

In some embodiments, the scaffold sequence may comprise a nucleotide sequence at least 85% identical to SEQ ID NO: 2. In some instances, the scaffold sequence comprises one or more deletions, one or more nucleotide substitutions, or a combination thereof, as compared with SEQ ID NO: 2.

In some instances, the scaffold sequence is a truncated variant of SEQ ID NO: 2, which may have a 3' truncation relative to SEQ ID NO: 2. In some instances, the scaffold sequence is a truncated variant of SEQ ID NO: 2, which may have an internal truncation relative to SEQ ID NO: 2. In some instances, the scaffold sequence is a truncated variant of SEQ ID NO: 2, which may have a 3' truncation and an internal truncation relative to SEQ ID NO: 2. In some examples, the truncated variant may be about 110-140-nt in length. In some examples, the 3' truncation may comprise deletions within residues 143-202 of SEQ ID NO: 2. Alternatively or in addition, the truncated variant may further comprise one or more deletions within residues 10-40 of SEQ ID NO: 2. In some specific examples, the deletions may comprise residues 14-20 and/or 25-32 of SEQ ID NO: 2. Alternatively or in addition, the truncated variant may further comprise one or more mutations relative to SEQ ID NO: 2 such as a deletion and/or substitution within residues 81-85 of SEQ ID NO: 2. In one specific examples, the scaffold sequence comprises (*e.g.*,

consists of) the nucleotide sequence of SEQ ID NO: 27. In another specific example, the scaffold sequence comprises (*e.g.*, consists of) SEQ ID NO: 28. Such scaffold sequence may be about 115-130-nt in length.

In some examples, the CRISPR nuclease polypeptide comprised in the gene editing system disclosed herein is the reference CRISPR nuclease of SEQ ID NO: 1 and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 27. Alternatively, the CRISPR nuclease polypeptide in the gene editing system can be a variant of SEQ ID NO: 1, comprising mutations at positions I67, D568, and E706 (*e.g.*, arginine substitutions I67R, D568R, and E706R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 28. Still alternatively, the CRISPR nuclease polypeptide in the gene editing system can be a variant of SEQ ID NO: 1 comprising a deletion within residues 1-15 of SEQ ID NO: 1; optionally a deletion of residues 1-14 or 1-15 of SEQ ID NO: 1; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 28.

The RT donor RNA comprises a primer binding site (PBS) and a template sequence. 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 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, and optionally the 3' end nucleotide of the PBS-targeting site is about 2-15 nucleotides upstream to the PAM.

In some embodiments, the template sequence in the RT donor RNA portion of the RNA 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 or in addition, at least one nucleotide variation may be located in the PAM.

In some embodiments, any of the RNA molecules in the gene editing systems 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 hairpin, a pseudoknot, a circularization, or a triplex structure.

In some examples, the RNA molecule 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. In other examples, the RNA molecule may comprise, from 5' to 3', the spacer sequence, the scaffold sequence, the template sequence, the PBS, a linker, and the 3' extension.

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 messenger RNA (mRNA).

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 other examples, the nucleic acid is a mRNA.

In some embodiments, the gene editing system disclosed herein may further comprise one or more lipid excipients associated with the fusion polypeptide or the encoding nucleic acid and/or the RNA molecule and/or the encoding nucleic acid of the gene editing system; optionally wherein the one or more lipid excipients form lipid nanoparticles (LNPs), which are associated with or encapsulate the fusion polypeptide or the encoding nucleic acid and/or the RNA molecule and/or the encoding nucleic acid of the gene editing system.

Alternatively, the gene editing system may comprise one or more viral vectors, for example, one or more adeno-associated viral (AAV) vectors encoding one or more of the fusion polypeptide or the encoding nucleic acid, and the RNA molecule and/or the encoding nucleic acid Also provided herein are a pharmaceutical composition comprising any of the gene editing systems provided herein, and a kit comprising the fusion polypeptide or the encoding nucleic acid and the RNA molecule and/or the encoding nucleic acid of the gene editing system as disclosed herein.

In other aspects, the present disclosure features a gene editing method, comprising delivering any of the gene editing systems disclosed herein to a host cell to edit a genomic site targeted 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

ID NO: 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, or 90.

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: 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, or 89. 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 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 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 percentages of NGS reads comprising indels at six genetic loci as indicated in the presence or absence of the CRISPR nuclease of SEQ ID NO: 1.

FIGS. 2A-2D are gel images showing *in vitro* cleavage of target or non-target strand of a target DNA substrate and a bar graph showing quantification of nuclease activity. **FIG. 2A** is a gel image captured using an 800 nm channel showing *in vitro* cleavage of the target strand (labelled on the 5' end with an IR800 dye) of the target DNA substrate by the reference CRISPR nuclease, putative HNH-knockout nickases, or putative RuvC-knockout nickases. **FIG. 2B** is a gel image captured using a 700 nm channel showing *in vitro* cleavage of the non-target strand (labelled on the 5' end with an IR700 dye) of a target DNA substrate by the reference CRISPR nuclease, putative HNH-knockout nickases, or putative RuvC-knockout nickases. **FIG. 2C** is an overlaid image captured using 800 nm and 700 nm channels of **FIG. 2A** and **FIG. 2B**. **FIG. 2D** is a bar graph showing 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.

FIGS. 3A-3C are schematic illustrations depicting predicted secondary structures of scaffold sequences. **FIG. 3A**: Reference Scaffold (SEQ ID NO: 2). **FIG. 3B**: Truncated Reference Scaffold (SEQ ID NO: 27). **FIG. 3C**: Scaffold 2 (SEQ ID NO: 28).

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 polypeptide comprising the CRISPR nuclease polypeptide and the RT polypeptide, 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 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.

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 some 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.

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

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.

(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 or SEQ ID NO: 160 (the reference CRISPR nuclease, with or without the N-terminus M residue), 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/or 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.

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 15-53, 308-338, and 472-566 of SEQ ID NO: 1) and a HNH domain (located at residues 339-471 of SEQ ID NO: 1). The RuvC nuclease domain and the HNH nuclease domain coordinate cleavage of the DNA strand adjacent to the 5'-NGG-3' PAM motif, in which N represents any nucleotide. Positions D24, E335 and D524 are deemed the active sites in the RuvC domain and positions D396, H397, and N420 are deemed the active sites in the HNH domain. R506 and H521 may also be important for the nuclease activity of the RuvC domain. In addition to the nuclease domains, the reference CRISPR nuclease of SEQ ID NO: 1 also includes a BH domain (residues 54-89 of SEQ ID NO: 1), a REC domain (residues 90-307 of SEQ ID NO: 1), a PLL domain (residues 567-580 of SEQ ID NO: 1), a WED domain (residues 581-672 of SEQ ID NO: 1), and a PID domain (residues 673-776 of SEQ ID NO: 1).

Compared to the CRISPR Cas9 nuclease, the reference CRISPR nuclease of SEQ ID NO: 1 disclosed herein is smaller. Its cognate scaffold comprises a distinctive structure compared to the Cas9 nuclease scaffold. The distinctive RNA scaffold is expected to allow for decreased size of some domains (*e.g.*, the REC domain) and thus contribute to the smaller nuclease size. Additionally, the cognate scaffold can be miniaturized relative to the reference scaffold of SEQ ID NO: 2. These features would be beneficial for delivery. Arginine and/or lysine substitutions (*e.g.*, arginine substitutions) can be introduced into the CRISPR nuclease of SEQ ID NO: 1 to increase indel activity. Additionally, because the reference CRISPR nuclease of SEQ ID NO: 1 comprises a RuvC domain and an HNH domain, nickase variants can be engineered. The cutting pattern of the reference CRISPR nuclease of SEQ ID NO: 1 is different than that of Cas9 and creates a staggered double-strand break, allowing for additional gene targets to be edited.

Variants of CRISPR Nuclease Polypeptide

The variant CRISPR nuclease polypeptides provided herein are derived from 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 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 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.

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 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 some instances, the decreased enzymatic activity is achieved by reducing or 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, 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 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%, 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 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 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 instances, no alterations are introduced into the RuvC and/or the HNH nuclease domains, or at the active sites and/or sites involved in activity in these domains as provided herein. Alternatively, conservative amino acid substitutions may be introduced into SEQ ID NO: 1, including in the RuvC and/or the HNH nuclease domains.

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, T; (f) Q, N; and (g) E, D.

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

In some instances, the arginine and/or lysine 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 instances, the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) may be located at one or more of positions D56, E59, G60, E63, I67, G71, S564, D568, W593, T605, E655, E694, and E706 in SEQ ID NO: 1. In some specific

examples, the variant CRISPR nuclease polypeptide may comprise arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at the following positions relative to SEQ ID NO: 1: a) I67, D568, and E706; b) I67, D56, and D568; c) I67, T605, and D568; d) I67, D56, and E706; e) I67, T605, and E706; f) I67, D568, and E694; g) I67, D56, E694, and E706; h) I67, D56, T605, and E706; or i) I67, D568, W593, and E706. In one specific example, the engineered variant of the reference CRISPR nuclease comprises arginine and/or lysine substitutions at I67, D568, and E706. In another specific example, the engineered variant of the reference CRISPR nuclease comprises arginine and/or lysine substitutions at I67, D568, W593, and E706.

Any of the disclosed engineered variants of the reference CRISPR nuclease disclosed herein may contain up to 20 arginine and/or lysine substitutions (*e.g.*, arginine substitutions), for example, up to 15 arginine and/or lysine substitutions (*e.g.*, arginine substitutions).

In some examples, the variant CRISPR nuclease polypeptide may contain one or more of the following arginine substitutions: D56R, E59R, G60R, E63R, I67R, G71R, S564R, D568R, T605R, E655R, E694R, E706R. By way of non-limiting examples, the one or more arginine substitutions may include: a) I67R, D568R, and E706R; b) I67R, D56R, and D568R; c) I67R, T605R, and D568R; d) I67R, D56R, and E706R; e) I67R, T605R, and E706R; f) I67R, D568R, and E694R; g) I67R, D56R, E694R, and E706R; h) I67R, D56R, T605R, and E706R; or i) I67R, D568R, W593R, and E706R.

In some instances, the arginine and/or lysine substitution may be located within the RuvC and/or the HNH nuclease domains. In some examples, the arginine and/or lysine substitutions may be within the RuvC nuclease domain to reduce or inactivate the RuvC domain (*e.g.*, at positions D24, E335, R506, H521, and/or D524 in the RuvC domain). In other examples, the arginine and/or lysine substitutions may be within the HNH domain, *e.g.*, at positions D396, H397, and/or N420 in the HNH domain. In yet other examples, the arginine and/or lysine substitutions may be within both the RuvC domain and the HNH domain to reduce or diminish the nuclease enzymatic activity.

Alternatively, the arginine and/or lysine substitution may not be at the active sites and/or sites involved in activity in the RuvC and/or the HNH nuclease domains (e.g., not at positions D24, E335, R506, H521, and/or D524 in the RuvC domain and/or at positions D396, H397, and/or N420 in the HNH domain). In some examples, the arginine and/or lysine substitution may not be in the RuvC and/or HNH domains.

It is reported herein that arginine substitutions at the following positions in SEQ ID NO: 1 led to diminished or no nuclease activity, indicating that these positions are not tolerable to mutations with respect to nuclease activity: Y383R, K423R, M469R, F511R, Y108R, V321R,

I270R, V382R, S309R, L651R, T377R, N674R, F640R, L658R, G31R, V107R, A455R, K672R, W509R, I23R, P743R, V42R, T596R, L452R, M169R, H289R, M336R, L630R, T30R, V34R, V35R, G22R, G659R, L85R, C111R, G115R, V317R, V725R, Y683R, V595R, M50R, L615R, I736R, V526R, A529R, I606R, Q662R, A414R, C384R, I406R, V48R, L700R, Y611R, I466R, H397R, A635R, L523R, T347R, F399R, I527R, G657R, C415R, A525R, V664R, L310R, V182R, D524R, F186R, E337R, V675R, L765R, L483R, L32R, K450R, S473R, G499R, P581R, L21R, H521R, D396R, C416R, I398R, Y641R, T673R, L739R, G27R, V413R, M663R, L745R, P400R, C286R, T389R, N339R, A393R, I474R, T502R, N411R, A33R, F341R, S338R, G26R, L528R, G475R, N420R, D24R, Y358R, V334R, A472R, I343R, and S470R.

In some embodiments, a variant CRISPR nuclease provided herein exhibits nuclease activity and may not have mutations at the above positions. Alternatively, a variant CRISPR nuclease may be a dead nuclease (*e.g.*, for use in base editing) having mutations at one or more of those positions.

In some examples, a variant CRISPR nuclease comprising arginine and/or lysine substitutions at a combination of 2 or more (*e.g.*, 3 or 4) positions of D56, I67, D568, T605, E694, and E706. Examples include, but are not limited to, (a) I67R, D568R, and E706R; (b) D56R, I67R, and D568R; (c) I67R, D568R, and T605R; (d) D56R; I67R, and E706R; (e) I67R, T605R, and E706R; (f) I67R, D568R, and E694R; (g) D56R, I67R, E694R, and E706R; (h) D56R, I67R, T605R, and E706R; and (i) I67R, D568R, W593R, and E706R. In some specific examples, the engineered CRISPR nuclease polypeptide comprises the arginine substitutions of I67R, D568R, and E706R. In other specific examples, the engineered CRISPR nuclease polypeptide comprises the arginine substitutions of I67R, D568R, W593R, and E706R.

Alternatively or in addition, the engineered CRISPR nuclease polypeptide disclosed herein may have an N-terminal truncation relative to the reference CRISPR nuclease of SEQ ID NO: 1. Such an engineered CRISPR nuclease polypeptide has a fragment at the N-terminus of the reference CRISPR nuclease deleted. In some instances, the deleted N-terminal fragment has up to 80 amino acid residues, for example, up to 70 amino acid residues, up to 60 amino acid residues, up to 50 amino acid residues, up to 40 amino acid residues, up to 30 amino acid residues, or up to 20 amino acid residues. In some examples, the N-terminal truncation is a deletion of residues 1-14 or 1-15 of SEQ ID NO: 1. In some instances, the truncated variant may further comprise one or more of the mutations disclosed herein, for example, one or more arginine/lysine substitutions, one or more mutations resulting in nickase activity, or a combination thereof. In some specific examples, the truncated variant further comprises arginine substitutions of I67R, D568R, E706R relative to SEQ ID NO: 1. In other specific examples, the

truncated variant further comprises arginine substitutions of I67R, D568R, W593R, and E706R relative to SEQ ID NO: 1.

Alternatively or in addition, the engineered CRISPR nuclease polypeptide disclosed herein may have a C-terminal truncation relative to the reference CRISPR nuclease of SEQ ID NO: 1. Such an engineered CRISPR nuclease polypeptide has a fragment at the C-terminus of the reference CRISPR nuclease deleted. In some instances, the deleted C-terminal fragment has up to 10 amino acid residues, for example, up to 5 amino acid residues, up to 4 amino acid residues, up to 3 amino acid residues, up to 2 amino acid residues, or up to 1 amino acid residues. In some examples, the C-terminal truncation is a deletion of the final residue, the final two residues, the final three residues, the final four residues, or the final five residues of SEQ ID NO: 1. In some instances, the C-terminal truncated variant may further comprise one or more of the mutations disclosed herein, for example, one or more arginine/lysine substitutions, one or more mutations resulting in nickase activity, or a combination thereof. In some specific examples, the C-terminal truncated variant further comprises arginine substitutions of I67R, D568R, E706R relative to SEQ ID NO: 1. In other specific examples, the C-terminal truncated variant further comprises arginine substitutions of I67R, D568R, W593R, and E706R relative to SEQ ID NO: 1. In some examples, the C-terminal truncated variant further comprises an N-terminal truncation. In some examples, the C-terminal truncated variant further comprises a deletion of residues 1-14 or 1-15 of SEQ ID NO: 1.

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 domain to reduce or eliminate the nuclease activity of the target domain, thereby producing a variant with nickase activity. 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. The variant CRISPR nuclease polypeptides may share a high sequence homology relative to the reference CRISPR nuclease (*e.g.*, at least 85% sequence identity).

The one or more 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 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.

Positions D24, E337, and D524 are identified as putative catalytic residues in the RuvC domain, and positions D396, H397, and N420 are identified as putative catalytic residues in the HNH domain. In some examples, the one or more mutations may be within the RuvC nuclease domain to reduce or inactivate the RuvC domain (*e.g.*, at positions D24, E337, and/or D524 in the RuvC domain). In some specific examples, the variant CRISPR nuclease polypeptide may contain substitution(s) of D24A, E337A, and/or D524A. In other examples, any of D24, E337, and D524 may be substituted by an amino acid residue similar to A, for example, G, S, or L.

In some examples, the one or more mutations may be within the HNH domain to reduce or inactivate the HNH domain (*e.g.*, at positions D396, H397, and/or N420 in the HNH domain). In some specific examples, the variant CRISPR nuclease polypeptide may contain substitutions of H397A, D396A, and/or N420A. Alternatively, any of H397, D396, and N420 may be replaced with an amino acid residue similar to A, for example, G, L, or S. In one specific example, the variant CRISPR nuclease polypeptide may contain substitution of H397A or H397L.

In yet other examples, the one or more mutations may be within both the RuvC domain and the HNH domain to reduce or diminish the nuclease enzymatic activity.

In some embodiments, the variant CRISPR nuclease polypeptide provided herein may comprise both arginine/lysine substitutions, e.g., at one or more positions provided herein, and nickase mutations, which are mutations that lead to nickase activity. For example, the variant CRISPR nuclease polypeptide may comprise arginine/lysine substitutions (*e.g.*, arginine substitutions) at positions I67, D568, and/or E706, and further comprise an amino acid substitution at H397 (*e.g.*, H1397A or H397L) in SEQ ID NO: 1.

In some examples, the variant CRISPR nuclease polypeptide that contains one or more of the mutations noted herein (e.g., arginine and/or lysine substitutions, nickase mutation, and/or N-terminal truncation) may share a sequence identity at least 90% (*e.g.*, 95%, 97%, 98%, 99%, 99.5%, or greater) with SEQ ID NO: 1. In specific examples, the variant CRISPR nuclease polypeptide may comprise (*e.g.*, consist of) an amino acid sequence of any one of SEQ ID NOs: 6-13, 34, 36, 38, 40, 42, 44, 46, 58, 62, and 64 (*e.g.*, SEQ ID NO: 6, SEQ ID NO: 36).

Exemplary CRISPR nuclease polypeptides for use in the gene editing systems provided herein are disclosed in **Table 1**, **Table 11**, or **Table 14** below, each of which is within the scope

of the present disclosure.

In some embodiments, the CRISPR nuclease polypeptide in the gene editing systems 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 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.

In some instances, the CRISPR nuclease portion in the fusion polypeptide may be the reference CRISPR nuclease of SEQ ID NO: 1. Alternatively, the CRISPR nuclease portion in the fusion polypeptide may be a variant CRISPR nuclease derived from SEQ ID NO: 1 as those disclosed herein. Exemplary additional 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 control factor, a chemically inducible factor, a chromatin visualization factor, or a combination thereof.

In some embodiments, the additional functional moiety may comprise an 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-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 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 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.orjp/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 **Table 1** 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. 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 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.

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 25**, a variant thereof, or an ortholog thereof.

Table 25. Exemplary Reverse Transcriptase Sequences

Reverse Transcriptase	Sequence	SEQ ID NO
MMLV-RT	*See Table 15 below*	66 *With no N-terminus M residue (SEQ ID NO: 176)*
MMLV-RT_2	MTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEA RLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREV NKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQP LFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQH PDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQIC QKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFC RLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEILQALLTAPALG LPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGW PPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSN ARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEA HGTRPDL	178 *With no N-terminus M residue (SEQ ID NO: 179)*
MMTV-RT	MFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQ PVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKS GKWRLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFF NIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFV DKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHG LVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLL	180 *With no N-terminus M residue (SEQ ID NO: 181)*

	GNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMN ERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEK EDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIV IFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQP FNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYI GHIRGHTGLPGPLAQGNAYADSLTRIL	
MarathonRT	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNG ETIKGQLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQ VLTPIYEEQFHDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEK FFDTVNHDKLMTLIGRTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGT PQGGNLSPLLANIMLNELDKEMEKRGLNFVRYADDCIIMVGSEMSANR VMRNISRFIEEKLGLKVNMTKSKVDRPSGLKYLGFGFYFDPRAHQFKA KPHAKSVAKFKKRMKELTCRSWGVSNSYKVEKLNQLIRGWINYFKIGS MKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLVKLGIDRNTARRVA YTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKCVTC	182 *With no N-terminus M residue (SEQ ID NO: 183)*
RTX reverse transcriptase	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHGTVVTVKRVEKVQKKFLGRPVEVWKLYFTHPQDVPAI MDKIREHPAVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIET LYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVVSTEREMIK RFLRVVKEKDPDVLITYNGDNFDFAYLKKRCEKLGINFALGRDGSEPK IQRMGDRFAVEVKGRIHFDLYPVIRRTINLPTYTLEAVYEAVFGQPKE KVYAEEITTAWETGENLERVARYSMEDAKVTYELGKEFLPMEAQLSRL IGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARRHQSH EGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLNREGCKEYD VAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERKLLD YRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYLTMTIKEI EEKYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALE LEYEGFYKRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQA RVLEALLKDGDVEKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKD YKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEF DPTKHKYDAEYYIEKQVLPAVERILRAFGYRKEDLRYQKTRQVGLSAR LKPKGT	184 *With no N-terminus M residue (SEQ ID NO: 185)*
RTX-exoMinus	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHGTVVTVKRVEKVQKKFLGRPVEVWKLYFTHPQDVPAI MDKIREHPAVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIET LYHEGEEFAEGPILMISYADEEGARVITWKNVDLPYVDVVSTEREMIK RFLRVVKEKDPDVLITYDGDNFDFAYLKKRCEKLGINFALGRDGSEPK IQRMGDRFAVEVKGRIHFDLYPVIRRTINLPTYTLEAVYEAVFGQPKE KVYAEEITTAWETGENLERVARYSMEDAKVTYELGKEFLPMEAQLSRL IGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARRHQSH EGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLNREGCKEYD VAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERKLLD YRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYLTMTIKEI EEKYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALE LEYEGFYKRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQA RVLEALLKDGDVEKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKD YKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEF DPTKHKYDAEYYIEKQVLPAVERILRAFGYRKEDLRYQKTRQVGLSAR LKPKGT	186 *With no N-terminus M residue (SEQ ID NO: 187)*
B11 RT	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAI EEVKKITAERHSTVVTVKRVEKVQKKFLGRSVEVWKLYFTHPQDVPAI MDKIREHPAVIDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLALDIGT PCHEGEVFAEGPILMISYADEEGTRVITWRNVDLPYVDVLSTEREMIQ RFLRVVKEKDPDVLITYNGDNFDFAYLKKRCEKLGINFTLGREGSEPK IQRMGDRFAVEVKGRIHFDLYPVIRRTVNLPIYTLEAVYEAVFGQPKE KVYAEEITTAWETGENLERVARYSMEDAKVTYELGKEFMPMEAQLSRL IGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNKPDEKELARRHQSH EGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLNREGCKEYD VAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERKLLD YRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYITMTIKEI EEKYGFKLIYSDTDGFFATIPGAEAETVKKKAMEFLKYINAKLPGALE	188 *With no N-terminus M residue (SEQ ID NO: 189)*

	LEYEGFYKRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQA RVLEALLKDGDVEKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKD YKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIVDRAIPFDEF DPTKHKYDAEYYIENQVLPAVERILRAYGYRKEDLWYQKTRQVGLSAR LKPKGT	
RT CaMV	MDHLLLKTQTQTEQVMNVTNPNSIYIKGRLYFKGYKKIELHCFVDTGA SLCIASKFVIPEEHWVNAERPIMVKIADGSSITISKVCKDIDLIIAGE IFRIPTVYQQESGIDFIIGNNFCQLYEPFIQFTDRVIFTKNKSYPVHI AKLTRAVRVGTEGFLESMKKRSKTQQPEPVNISTNKIENPLEEIAILS EGRRLSEEKLFITQQRMQKIEELLEKVCSENPLDPNKTKQWMKASIKL SDPSKAIKVKPMKYSPMDREEFDKQIKELLDLKVIKPSKSPHMAPAFL VNNEAEKRRGKKRMVVNYKAMNKATVGDAYNLPNKDELLTLIRGKKIF SSFDCKSGFWQVLLDQESRPLTAFTCPQGHYEWNVVPFGLKQAPSIFQ RHMDEAFRVFRKFCCVYVDDILVFSNNEEDHLLHVAMILQKCNQHGII LSKKKAQLFKKKINFLGLEIDEGTHKPQGHILEHINKFPDTLEDKKQL QRFLGILTYASDYIPKLAQIRKPLQAKLKENVPWRWTKEDTLYMQKVK KNLQGFPPLHHPLPEEKLIIETDASDDYWGGMLKAIKINEGTNTELIC RYASGSFKAAEKNYHSNDKETLAVINTIKKFSIYLTPVHFLIRTDNTH FKSFVNLNYKGDSKLGRNIRWQAWLSHYSFDVEHIKGTDNHFADFLSR EFNKVNSSGGS	190 *With no N-terminus M residue (SEQ ID NO: 191)*
RT retron	MGIHGVPAAMKSAEYLNTFRLRNLGLPVMNNLHDMSKATRISVETLRL LIYTADFRYRIYTVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKL SSSPFSIGFEKHQSILNNATPHIGANFILNIDLEDFFPSLTANKVFGV FHSLGYNRLISSVLTKICCYKNLLPQGAPSSPKLANLICSKLDYRIQG YAGSRGLIYTRYADDLTLSAQSMKKVVKARDFLFSIIPSEGLVINSKK TCISGPRSQRKVTGLVISQEKVGIGREKYKEIRAKIHHIFCGKSSEIE HVRGWLSFILSVDSKSHRRLITYISKLEKKYGKNPLNKAKT	192 *With no N-terminus M residue (SEQ ID NO: 193)*
HIV2	MEKEGQLEEAPPTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVTQDFT EIQLGIPHPAGLAKKRRITVLDVGDAYFSIPLHEDFRPYTAFTLPSVN NAEPGKRYIYKVLPQGWKGSPAIFQHTMRQVLEPFRKANKDVIIIQYM DDILIASDRTDLEHDRVVLQLKELLNGLGFSTPDEKFQKDPPIHWMGY ELWPTKWKLQKIQLPQKEIWTVNDIQKLVGVLNWAAQLYPGIK	194 *With no N-terminus M residue (SEQ ID NO: 195)*
HIV1	MEEEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFW EVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSIN NETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRIKNPEIVIYQYM DDLYVGSDLEIGQHRTKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGY ELHPDRWTVQPIDLPEKDSWTVNDIQKLVGKLNWASQIYAGIK	196 *With no N-terminus M residue (SEQ ID NO: 197)*
HIV-1 RT p51	MPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKI SKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIP HPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIR YQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGS DLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKW TVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTK ALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQG QWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIW GKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEK EPIVGAETF	198 *With no N-terminus M residue (SEQ ID NO: 199)*
MMLV RT lacking the RNase H domain	STLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAP LIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWN TPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQG FKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFN WGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALN PATLLP	200 *With N-terminus M residue (SEQ ID NO: 201)*

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.

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, 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 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.

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% 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 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, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or 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 15** 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

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

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 or SEQ ID NO: 160 (with no N-terminal M residue). 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 variant, *e.g.*, having any of the mutations noted above in the HNH domain or in the RuvC domain relative to the reference CRISPR nuclease of SEQ ID NO: 1, for example, the nickase of SEQ ID NO: 36. 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. In some instances, the CRISPR nuclease polypeptide in the fusion polypeptide may not include its native N-terminus M residue, for example, when the CRISPR nuclease polypeptide has a functional fragment (e.g., NLS) added to its N-terminus or when the CRISPR nuclease is located at the C-terminal portion of the fusion polypeptide. Exemplary variants of SEQ ID NO: 1 for use in constructing the fusion polypeptides disclosed herein are provided in **Table 1**, **Table 11**, or **Table 14**.

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. 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** and **17** below. See also **Examples 1**, **6**, and **7**.

In some examples, the CRISPR nuclease-RT fusion polypeptide provided herein comprises a peptide linker located between the CRISPR nuclease polypeptide and the RT polypeptide. In some instances, the CRISPR nuclease polypeptide is N-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 place between the CRISPR nuclease 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 two NLSs. 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 some examples, the CRISPR nuclease-RT fusion polypeptide disclosed herein may comprise, from N-terminus to C-terminus, a first NLS, the CRISPR nuclease, the RT polypeptide, and a second NLS. 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. Examples of the CRISPR nuclease-RT polypeptides are provided in **Tables 16**, **21**, and **22** below, all of which are within

the scope of the present disclosure.

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.

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.

(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 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.

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 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 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, 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 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 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 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 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 metallothionine 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.

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 spacer can be adjacent to a protospacer adjacent motif (PAM) of 5'-NGG-3', in which N represents any nucleotide. 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). In some embodiments, a first sequence is adjacent to a second sequence if the two sequences are separated by about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides. In some embodiments, a first sequence is adjacent to a second sequence if the two sequences are separated by up to 2 nucleotides, up to 5 nucleotides, up to 8 nucleotides, up to 10 nucleotides, up to 12 nucleotides, or up to 15 nucleotides. In some embodiments, a first sequence is adjacent to a second sequence if the two sequences are separated by 2-5 nucleotides, 4-6 nucleotides, 4-8 nucleotides, 4-10 nucleotides, 6-8 nucleotides, 6-10 nucleotides, 6-12 nucleotides, 8-10 nucleotides, 8-12 nucleotides, 10-12 nucleotides, 10-15 nucleotides, or 12-15 nucleotides.

In specific examples, the spacer targets a nucleotide sequence (target sequence) that is immediately adjacent to the PAM motif. In other specific examples, the target sequence and the PAM motif have a small gap of less than 5 (*e.g.*, 1, 2, 3, 4, or 5) nucleotides.

A spacer sequence as disclosed herein may have a length of from about 15 nucleotides to about 30 nucleotides. For example, the spacer 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.

GUUACAGUUAAGGCUCUUUGGAAACAAAGAAGCCUUAAUUGUAAAACGCC
UAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUC
UACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGACGAUAUUUAAAAAAC
UUUCUGCGGCUUGCUUACUUUAGUAAGCUUUGUGGCUGAGGCAGAAUUCC
UU (SEQ ID NO: 2)

The predicted secondary structure of the reference scaffold sequence of SEQ ID NO: 2 is provided in **FIG. 3A**.

In other instances, the scaffold sequence may be a variant derived from SEQ ID NO: 2. 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 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 some embodiments, the scaffold sequence can be a truncated variant of SEQ ID NO: 2. Any of the truncation variants provided herein may be about 110-140-nucleotide in length.

In some examples, the truncated variant scaffold sequence may have a 3' truncation relative to SEQ ID NO: 2. Such a 3' truncation may remove one or more of the stem structures P6a, P6b, and P6c depicted in **FIG. 3A**. In specific examples, the 3' truncation may comprise a deletion within residues 143-202 of SEQ ID NO: 2, *e.g.*, deletion of the 143-202 fragment of SEQ ID NO: 2. S*ee*, **FIG. 3B**.

Alternatively or in addition, the variant scaffold sequence can be a truncated variant of SEQ ID NO: 2 having an internal truncation relative to SEQ ID NO: 2. For example, such a truncation variant may comprise a deletion of whole or part of the stem structure P1 depicted in **FIG. 3A**. In some examples, the truncation variant may comprise a deletion within residues 14-20 of SEQ ID NO: 2, within residues 25-32 of SEQ ID NO: 2, or a combination thereof. In specific examples, the truncation variant may comprise deletions of the 14-20 fragment and 25-32 fragment of SEQ ID NO: 2. In some instances, the truncation variant may comprise a deletion

in the loop consisting of residues 77-86 of SEQ ID NO: 2 (*see*, **FIG. 3C**), which may result in shortening of the stem structure P5 depicted in **FIG. 3A**. In some examples, the truncation variant may comprise a deletion within residues 81-85 (*e.g.*, deletion of the 81-85 fragment) of SEQ ID NO:2.

In some examples, the variant scaffold sequence may comprise a combination of the 3' truncation and one or more of the internal truncations as disclosed herein. Alternatively or in addition, the variant scaffold sequence may comprise one or more nucleotide variations relative to the corresponding residues in SEQ ID NO: 2.

In one specific examples, the scaffold sequence comprises (e.g., consists of) the nucleotide sequence of SEQ ID NO: 27. In another specific example, the scaffold sequence comprises (e.g., consists of) SEQ ID NO: 28. Such a scaffold sequence may be about 115-130 - nt in length.

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 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 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 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.

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-paring. 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 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, 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. 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 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 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. 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 specific 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 7** below.

(iv) Modification 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 in a gene editing system as disclosed herein, *e.g.*, the editing template RNA or any of the nucleic acid sequences encoding the components, may include one or more modifications, 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 the 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 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-

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, 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-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-methyladenosine, N6-methyladenosine, N6-isopentenyladenosine, N6-(cis-hydroxyisopentenyl)adenosine, 2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine, N6-glycinylcarbamoyladenosine, 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 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 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, 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 11** below, such as the nickase variant containing the H397 mutation, *e.g.*, H397A). 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.*, D56, E59, G60, E63, I67, G71, S564, D568, T605, E655, E694, and/or E706 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.*, I67, D568, and E706. In other specific examples, the CRISPR nuclease polypeptide in the fusion polypeptide may comprise both the arginine and/or lysine substitutions as disclosed herein and the nickase mutations as also disclosed herein. Examples are provided in **Table 14** below. In some instances, the CRISPR nuclease polypeptide in the fusion polypeptide may have the native N-terminus M residue removed, for example, when it is located at the C-terminal portion of the fusion polypeptide or when a functional fragment such as an NLS is fused to its N-terminus.

In some embodiments, the RT polypeptide in the fusion polypeptide may be an MMLV variant, for example, SEQ ID NO: 66 provided in **Example 6** below or SEQ ID NO: 176, which does not include the N-terminus M residue. In some instances, the RT polypeptide in the fusion polypeptide may have the native N-terminus M residue removed, for example, when it is located at the C-terminal portion of the fusion polypeptide or when a functional fragment such as an NLS is fused to its N-terminus.

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.

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 **Tables 16**, **21** and **22** 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).

In other examples, an exemplary gene editing system provided herein may comprise any of the CRISPR-RT fusion polypeptides provided in **Tables 16**, **21** and **22** below, 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).

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 **Tables 16**, **21** and

22 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). The nucleic acid may comprise an encoding nucleotide sequence that is codon optimized.

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 **Tables 16**, **21** and **22** below, 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. 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.

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 Systems to 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 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 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.*, 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 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 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 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.

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 conjugation, delivery of plasmids or transposons; particle-based methods, such as using a gene gun, magnectofection 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 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 (*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 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 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 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 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.

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.

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 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 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 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, 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 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 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.

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 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

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 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 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 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

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 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, 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* (lRL 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 invention to its fullest extent. The following specific

embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

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

This Example describes the genomic editing of 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.

The CRISPR nuclease was tagged with N-terminal SV40 nuclear localization signal (NLS) and a C-terminal 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. The reference and NLS-tagged sequences used are in **Table 1**. Plasmids were received from Genscript.

RNA guides were designed and cloned and 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 received from GenScript.

Table 1. Amino Acid Sequences of Reference and Exemplary Variant CRISPR Polypeptides

CRISPR Nuclease	Amino Acid Sequence	SEQ ID
Reference	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVADILEGRREQRGIRRTGQTKKKRLREL KTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEMP KEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVTK DVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKVE DCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDKE GREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHIL EHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRDKSNKGTPLYDKNPLSIK	1 *160 (with no N-terminal M residue)*

CRISPR Nuclease	Amino Acid Sequence	SEQ ID
	NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVEKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	
SV40 NLS	KRTADGSEFESPKKKRKV	3
Nucleoplasmin NLS	KRPAATKKAGQAKKKK	4
NLS-tagged Reference	*KRTADGSEFESPKKKRKV*RNCKKDLYVEVTPMQKELVLGIDY GGKYTGLAVVDRRNNQVLYARTVKMRDDVADILEGRREQR GIRRTGQTKKKRLRELKTYLKNIGGKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKEKTREEIEAMSADEKKQWEKEKKE WEETERNSRHRDEVTKDVRKIMSERDATEQQISHVIKIFNKQ YRPKRFNNRILTKCKVEDCGVNTPLRKNVRDLLIKDIVRFLPL TQSEKAELEAHVLDKEGREKVKTFFRRHRINEFLRKQIYDIAF NQKLSGRTVFCKKHILEHEQHTKEERKVFRLAPSLKTKINNV LAVVKDEILPQFGINKVIMESNNFDIAARTKGKKRLAKDEYG KGHKEAKESLIEVLLRETKNQCVYCGKVITVTTANTDHIFPR KAGGINVFANLVACCRECNEKKKGRTPLESGVLPQPEITAHL ENDLKKKILEDARNMNELNLNSIKYMSYASIGWRHMRDRLR GLIGNEELPIERQSGILTAYFRRWWGFRKERGNDIHHALDAVI LASRKDYTDDGVDMTLKPKSSNGREFDPEKHLPDTISFKRDK SNKGTPLYDKNPLSIKNSLITRRWMVTEIEHGKEDTIISDEYR KKLKEAFKHWGTSKGKCLTDEEARQTGFYQRKNGNFMSLR CVEKGLGWGQMVKINNNAFKTNVHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPEIKGKKLYTLRRGDYVTHESEDSV YRIKKLGTSPVLEAVVKGSIGRIRNSASATKLTKVSKPQKVFS GGSSGGSSGSETPGTSESATPESSGGSSGGSS*KRPAATKKAGQA KKKK*	5
I67R/D568R/ E706R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVADILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRRKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	6 *161 (with no N-terminal M residue)*

CRISPR Nuclease	Amino Acid Sequence	SEQ ID
I67R/D56R/ D568R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVARILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRRKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVEKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	7 *162 (with no N-terminal M residue)*
I67R/T605R/ D568R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVADILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRRKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDRIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVEKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	8 *163 (with no N-terminal M residue)*
I67R/D56R/ E706R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVARILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRDKSNKGTPLYDKNPLSIK	9 *164 (with no N-terminal M residue)*

CRISPR Nuclease	Amino Acid Sequence	SEQ ID
	NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	
I67R/T605R/ E706R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVADILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDRIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	10 *165 (with no N-terminal M residue)*
I67R/D568R/ E694R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVADILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRRKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYRIKNPRLLKHFVEKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	11 *166 (with no N-terminal M residue)*
I67R/D56R/ E694R/E706R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVARILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR	12 *167 (with no N-terminal M residue)*

CRISPR Nuclease	Amino Acid Sequence	SEQ ID
	ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDTIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYRIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	
I67R/D56R/ T605R/E706R variant	MRNCKKDLYVEVTPMQKELVLGIDYGGKYTGLAVVDRRNN QVLYARTVKMRDDVARILEGRREQRGRRRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNEPFMTVYKLCHKRGYDYVEM PKEKTREEIEAMSADEKKQWEKEKKEWEETERNSRHRDEVT KDVRKIMSERDATEQQISHVIKIFNKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIKDIVRFLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLRKQIYDIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAPSLKTKINNVLAVVKDEILPQFGINK VIMESNNFDIAARTKGKKRLAKDEYGKGHKEAKESLIEVLLR ETKNQCVYCGKVITVTTANTDHIFPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQPEITAHLENDLKKKILEDARNMN ELNLNSIKYMSYASIGWRHMRDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHALDAVILASRKDYTDDGVDMT LKPKSSNGREFDPEKHLPDTISFKRDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKEDRIISDEYRKKLKEAFKHWGTSKG KCLTDEEARQTGFYQRKNGNFMSLRCVEKGLGWGQMVKIN NNAFKTNVHNIGVAVYLDEKGKKKAYEIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTHESEDSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKPQKVF	13 *168 (with no N-terminal M residue)*

Table 2. Target and RNA Guide Sequences

Guide	PAM	Target Sequence	RNA Guide Sequence*
AAVS1-T3	AGG	TTGTGAGAATGGTGCGTCCT (SEQ ID NO: 14)	UUGUGAGAAUGGUGCGUCCUguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuugggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 15)
AAVS1-T6	TGG	CTTTGTAGCCTCTCCCGCTC (SEQ ID NO: 16)	CUUUGUAGCCUCUCCCGCUCguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuugggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 17)
EMX1-T2	AGG	TGGATGGCGACTTCAGGCAC (SEQ ID NO: 18)	UGGAUGGCGACUUCAGGCACguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuugggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 19)

EMX1-T7	GGG	CTTTAGCAAGGGACTATTCA (SEQ ID NO: 20)	CUUUAGCAAGGGACUAUUCAguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuuggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 21)
VEGFA-T3	TGG	TTGTTGACATTGTCCACACC (SEQ ID NO: 22)	UUGUUGACAUUGUCCACACCguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuuggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 23)
VEGFA-T6	AGG	TTTGCCCATACTGGGGACCA (SEQ ID NO: 24)	UUUGCCCAUACUGGGGACCAguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuuggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 25)

* 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 50-90% 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™ (ThermoFisher Scientific) to the center of each well and incubating at 37ºC 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 65ºC for 15 minutes, 68ºC for 15 minutes, and 98ºC for 10 minutes.

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

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 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 % indels was calculated as the number of indel-containing reads divided by the number of reads analyzed (reads passing filters up to 50,000). The QC standard for the minimum number of reads passing filters was 10,000.

For each target, indel ratios, referring to the percentage of NGS reads comprising 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**, four of the six targets tested demonstrated a greater level of indels observed when the CRISPR nuclease plasmid was present.

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

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

Arginine scanning mutagenesis was performed to individually substitute each non-arginine residue 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 710 single arginine substitution variants. Nucleic acids encoding the reference CRISPR nuclease and each CRISPR nuclease variant were then individually cloned into a pcda3.1 backbone (Invitrogen), and the plasmids were maxi-prepped and diluted. The plasmids comprised a CMV promoter, a first NLS (KRTADGSEFESPKKKRKV; SEQ ID NO: 3) upstream of the coding sequence, and an XTEN linker (SGGSSGGSSGSETPGTSESATPESSGGSSGGSS; SEQ ID NO: 9) followed by a second NLS (KRPAATKKAGQAKKKK; SEQ ID NO: 4) downstream of the coding sequence. See also **Example 1** above.

The RNA guide and target sequences are shown in **Table 3**. RNA guides were cloned into a pUC19 backbone (New England Biolabs®). The plasmids were then maxi-prepped and diluted.

Table 3. Mammalian Targets and Corresponding crRNAs*

Target Identifier	Guide	PAM Sequence	Target Sequence	RNA Guide Sequence
AAVS1	AAVS1-T3	AGG	SEQ ID NO: 14	SEQ ID NO: 15
VEGFA	VEGFA-T6	AGG	SEQ ID NO: 24	SEQ ID NO: 25

* See Table 2 above for the corresponding sequences.

HEK293T cells were transfected with plasmids expressing the CRISPR nuclease polypeptides and the gRNAs provided herein following the methods provided in **Example 1** above. Gene editing efficiencies were determined by NGS following the methods also provided in **Example 1**. The NGS results were analyzed also following the methods provided in **Example 1** above.

For each target, indel ratios, referring to the percentage of NGS reads comprising 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 at a particular target, the indel ratio for each variant was divided by the indel ratio for the reference. **Table 4** shows fold change in indel ratios for each target and as an average of both targets. Numbering is relative to the reference nuclease of SEQ ID NO: 1 (*i.e.*, without an NLS). For example, at the AAVS1 target, the indel ratio for the I67R variant was 7.07 times that of the reference indel ratio, and at the VEGFA target, the indel ratio for the I67R variant was 4.43 times that of the reference indel ratio. The average fold change in indel ratio for the I67R variant at both targets was 5.75 times that of the indel ratio for the reference at both targets.

As shown in **Table 4**, 21 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 both targets (right column).

Table 4. Fold Change in Indel Ratios*

Substitution	AAVS1	VEGFA	AAVS1 and VEGFA Average Fold Change
I67R	7.07	4.43	5.75
E63R	7.22	3.67	5.45
D56R	3.58	3.60	3.59
T605R	2.87	3.49	3.18
D568R	2.91	2.72	2.82
E694R	2.57	2.98	2.77
G72R	3.06	1.77	2.41
E706R	2.09	2.71	2.40

E655R	1.68	2.76	2.22
E59R	2.67	1.19	1.93
G60R	2.22	1.55	1.88
E599R	1.60	2.16	1.88
S564R	1.76	1.96	1.86
E444R	2.04	1.46	1.75
Q72R	1.84	1.49	1.67
G661R	1.15	1.99	1.57
L588R	1.73	1.40	1.56
E264R	1.64	1.48	1.56
K656R	1.62	1.47	1.55
I598R	1.23	1.85	1.54
W593R	1.60	1.47	1.53

* Variant indel ratio/Reference indel ratio

155 variants were analyzed as having indel ratios 1X-1.5X of the reference indel ratios, when averaged across both targets: E206R, K74R, S280R, A319R, I96R, A354R, D323R, F265R, E120R, H442R, V284R, E364R, C653R, A365R, N586R, E294R, E490R, N459R, E460R, E144R, S95R, N571R, S587R, P708R, K163R, G485R, V387R, I711R, W148R, K100R, G722R, A132R, Q328R, T201R, K322R, Q277R, N87R, K378R, N758R, I563R, G330R, D207R, K86R, G359R, K360R, K49R, T741R, I519R, E248R, E299R, K333R, Q41R, V12R, E152R, V210R, D53R, Q380R, M105R, K145R, S548R, S761R, K28R, V390R, K17R, D537R, K585R, Q772R, K351R, T255R, K287R, K146R, G66R, A441R, E491R, K302R, N340R, T98R, T83R, S752R, E149R, A55R, S650R, K355R, N489R, N457R, I753R, E94R, V774R, K572R, E371R, K566R, N462R, D561R, K366R, L369R, K546R, Y693R, S368R, K773R, Q16R,P771R, G385R, S180R, E11R, K614R, V10R, K422R, E129R, H243R, G209R, Q436R, K770R, K737R, K767R, K579R, K764R, G97R, K224R, P709R, L8R, K298R, K750R, Q177R, D533R, D356R, S170R, T395R, K449R, K252R, Y271R, E251R, W141R, K714R, T726R, K167R, I370R, K184R, K91R, E418R,D173R, Q295R, E131R, K350R, N332R, Y9R, A242R, K288R, E142R, N154R, F775R, H600R, L266R, E438R, V749R, and S134R. The remaining variants (534 variants) resulted in decreased indel ratios relative to the reference indel ratios (fold change in indel ratios of less than 1.0): G488R, E300R, K348R, K424R, P437R, K557R, S729R, K702R, G713R, E128R, V654R, S547R, T233R, A135R, T542R, T766R, K386R, T13R, Q64R, E453R, H293R, L543R, D536R, K75R, K363R, P560R, Q140R, K143R, K467R, P192R, M15R, K123R, L245R, K5R, T297R, K278R, K76R, S235R, H621R, E367R, K689R, D246R, Y534R, I756R, V768R, K268R, L434R, T391R, K139R, S465R, E236R, K6R, K82R, K665R, D7R, K138R, Y29R, V119R, E137R, N39R, E556R, N379R, C4R, K247R, T151R, A238R, S742R, T126R, D446R, V372R, Y471R, E357R, N549R, T392R, Q234R,

I130R, L701R, P698R, G538R, D272R, I678R, E710R, N517R, M133R, K613R, D159R, E292R, K626R, E552R, E430R, T562R, G516R, K311R, M594R, D164R, A274R, T535R, P122R, T70R, F275R, N677R, K687R, N263R, A747R, L559R, S570R, S155R, G550R, K237R, D52R, K620R, G751R, K254R, T763R, T440R, P575R, N196R, K109R, E495R, K188R, K216R, F304R, S769R, V205R, Q269R, L291R, G249R, E633R, K628R, E239R, K691R, K644R, K712R, K715R, A760R, N667R, P493R, I226R, E176R, G754R, L684R, H260R, D554R, H727R, N276R, S625R, Q637R, N315R, E160R, A403R, N3R, V433R, K696R, K202R, K402R, D609R, L200R, F229R, V744R, L429R, E18R, T718R, I607R, T73R, H296R, L374R, L78R, D481R, V748R, Q189R, E147R, S583R, H181R, E728R, E150R, E171R, A174R, D220R, K690R, K532R, Y724R, E124R, S759R, Y717R, H703R, P14R, K707R, A681R, L43R, L716R, C629R, L19R, K204R, G740R, T106R, K125R, T175R, E421R, M121R, K113R, K602R, D136R, G361R, I487R, K193R, E617R, G349R, L326R, G623R, K616R, E610R, I695R, E686R, Q497R, E746R, E597R, H558R, N445R, I168R, K313R, E241R, G688R, Q642R, D518R, V733R, K544R, D93R, H362R, L279R, N187R, S732R, K569R, L214R, D225R, V244R, P308R, N464R, G646R, P231R, L412R, I57R, F565R, W508R, Y504R, T162R, Y190R, E730R, D731R, G573R, H478R, G404R, N668R, Q178R, W622R, D540R, S431R, L221R, I199R, T590R, L501R, M541R, G627R, D685R, G601R, I439R, H112R, N407R, V705R, D454R, A344R, N647R, W476R, T631R, N217R, D632R, E80R, N101R, V539R, L486R, E324R, I335R, N211R, N645R, L110R, L492R, V253R, S498R, F409R, G639R, L447R, F195R, V20R, P555R, K738R, V161R, I223R, T638R, T283R, S608R, Y92R, E514R, P428R, T624R, L58R, V218R, K448R, I179R, M458R, I88R, I183R, A692R, Y577R, I325R, L222R, A45R, E102R, A670R, G89R, F329R, T427R, T574R, E634R, D117R, G432R, L232R, V303R, C381R, W660R, D36R, G281R, I584R, A307R, D604R, G679R, G510R, P327R, N316R, L318R, I262R, V227R, F256R, Y84R, N580R, L443R, V320R, F257R, D723R, V54R, F553R, D342R, L353R, T212R, M649R, Y116R, G425R, F104R, H157R, Y25R, A503R, I666R, I314R, I589R, G405R, L230R, T312R, I500R, I290R, N40R, Y44R, P435R, P103R, L582R, Y468R, V165R, P213R, C419R, G90R, F99R, F704R, F505R, F648R, L306R, A618R, L463R, I388R, N197R, A522R, P545R, H676R, F285R, Y118R, N669R, Y734R, N697R, V680R, H520R, L373R, I273R, K513R, L240R, F671R, C203R, E376R, L461R, N394R, L719R, I185R, M479R, V682R, A345R, E603R, T47R, F619R, A762R, L576R, D578R, A410R, I451R, I494R, L81R, I331R, V408R, C208R, S530R, Y383R, K423R, M469R, F511R, Y108R, V321R, I270R, V382R, S309R, L651R, T377R, N674R, F640R, L658R, G31R, V107R, A455R, K672R, W509R, I23R, P743R, V42R, T596R, L452R, M169R, H289R, M336R, L630R, T30R, V34R, V35R, G22R, G659R, L85R, C111R, G115R, V317R,

V725R, Y683R, V595R, M50R, L615R, I736R, V526R, A529R, I606R, Q662R, A414R, C384R, I406R, V48R, L700R, Y611R, I466R, H397R, A635R, L523R, T347R, F399R, I527R, G657R, C415R, A525R, V664R, L310R, V182R, D524R, F186R, E337R, V675R, L765R, L483R, L32R, K450R, S473R, G499R, P581R, L21R, H521R, D396R, C416R, I398R, Y641R, T673R, L739R, G27R, V413R, M663R, L745R, P400R, C286R, T389R, N339R, A393R, I474R, T502R, N411R, A33R, F341R, S338R, G26R, L528R, G475R, N420R, D24R, Y358R, V334R, A472R, I343R, and S470R.

Based on this experiment, the following substitutions were selected for further engineering: I67R, E63R, D56R, T605R, D568R, E694R, G71R, E706R, and E655R.

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. 106 combination CRISPR variants were tested.

Each CRISPR nuclease variant and RNA guide (*see* **Table 5**) was cloned as described in **Example 2**. 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 **Table 6** were calculated as the average of two bioreplicates, each of which contained two technical replicates.

Table 5. Mammalian Targets and Corresponding crRNAs*

Target Identifier	Guide	PAM Sequence	Target Sequence	RNA Guide Sequence
AAVS1	AAVS1-T3	AGG	SEQ ID NO: 14	SEQ ID NO: 15
VEGFA	VEGFA-T6	AGG	SEQ ID NO: 24	SEQ ID NO: 25
EMX1	EMX1-T7	GGG	SEQ ID NO: 20	SEQ ID NO: 21

* See Table 2 for corresponding sequences

Table 6. Indel Ratios for Mammalian Targets

Substitutions	AAVS1 average indel ratio	EMX1 average indel ratio	VEGFA average indel ratio	Average indel ratio targets
I67R, D568R, E706R	0.370	0.337	0.325	0.344
I67R, D56R, D568R	0.369	0.344	0.296	0.336
I67R, T605R, D568R	0.357	0.330	0.319	0.335

Substitutions	AAVS1 average indel ratio	EMX1 average indel ratio	VEGFA average indel ratio	Average indel ratio targets
I67R, D56R, E706R	0.356	0.332	0.287	0.325
I67R, T605R, E706R	0.339	0.319	0.303	0.320
I67R, D568R, E694R	0.326	0.336	0.296	0.319
I67R, D56R, E694R, E706R	0.301	0.336	0.281	0.306
I67R, D56R, T605R, E706R	0.285	0.340	0.277	0.300
I67R, T605R	0.317	0.297	0.273	0.296
I67R, D56R, D568R, E706R	0.293	0.312	0.281	0.295
I67R, T605R, E694R	0.283	0.320	0.273	0.292
I67R, D56R, E694R	0.300	0.318	0.256	0.291
T605R, D568R, E706R	0.268	0.289	0.315	0.291
I67R, D568R	0.339	0.267	0.264	0.290
I67R, D56R, T605R	0.302	0.308	0.260	0.290
I67R, D56R, D568R, E694R	0.273	0.320	0.268	0.287
E63R, T605R	0.329	0.272	0.260	0.287
I67R, T605R, D568R, E706R	0.271	0.302	0.281	0.285
I67R, D56R	0.331	0.284	0.238	0.284
D56R, D568R, E706R	0.267	0.293	0.289	0.283
I67R, T605R, D568R, E694R	0.238	0.324	0.284	0.282
E63R, E706R	0.361	0.238	0.245	0.282
D56R, T605R, D568R	0.255	0.306	0.277	0.279
T605R, D568R, E694R	0.233	0.293	0.288	0.271
E63R, E655R	0.296	0.265	0.251	0.271
E63R, D568R	0.335	0.234	0.239	0.270
I67R, E694R	0.289	0.279	0.240	0.270
T605R, D568R	0.258	0.270	0.278	0.269
I67R, D568R, E694R, E706R	0.265	0.280	0.253	0.266
D56R, T605R, E706R	0.227	0.299	0.268	0.264
D56R, T605R	0.229	0.298	0.250	0.259
I67R, E706R, E655R	0.226	0.286	0.259	0.257
I67R, T605R, E694R, E706R	0.239	0.284	0.247	0.257
I67R, D56R, T605R, D568R	0.241	0.279	0.248	0.256
I67R, E706R	0.299	0.226	0.230	0.252
D56R, D568R, E694R	0.232	0.263	0.259	0.251
I67R, E694R, E706R	0.278	0.247	0.216	0.247
I67R, D568R, E655R	0.217	0.269	0.250	0.245
E63R, E694R	0.276	0.239	0.217	0.244
D56R, D568R	0.245	0.257	0.222	0.241
D56R, T605R, E694R	0.191	0.291	0.241	0.241
D568R, E706R, E655R	0.179	0.261	0.279	0.240
I67R, E655R	0.233	0.245	0.238	0.239
D56R, E694R	0.205	0.280	0.228	0.238
I67R, D56R, D568R, E655R	0.180	0.292	0.237	0.236
T605R, E694R	0.190	0.272	0.242	0.235

Substitutions	AAVS1 average indel ratio	EMX1 average indel ratio	VEGFA average indel ratio	Average indel ratio targets
I67R, D56R, T605R, E694R	0.195	0.278	0.219	0.231
G71R, E655R	0.258	0.221	0.191	0.224
D56R, D568R, E655R	0.167	0.255	0.241	0.221
D56R, D568R, E694R, E706R	0.183	0.247	0.233	0.221
D568R, E694R	0.197	0.241	0.219	0.219
I67R, D56R, E706R, E655R	0.166	0.267	0.223	0.219
D56R, T605R, D568R, E706R	0.176	0.242	0.235	0.218
D568R, E706R	0.219	0.204	0.227	0.217
D56R, E706R	0.201	0.243	0.205	0.216
I67R, D56R, E655R	0.210	0.242	0.196	0.216
E694R, G71R	0.250	0.213	0.181	0.214
I67R, D568R, E706R, E655R	0.147	0.258	0.238	0.214
I67R, G71R	0.314	0.171	0.158	0.214
D56R, T605R, E694R, E706R	0.160	0.253	0.223	0.212
D568R, G71R	0.290	0.160	0.186	0.212
D56R, G71R	0.255	0.219	0.158	0.211
D568R, E655R	0.164	0.228	0.238	0.210
T605R, G71R	0.224	0.198	0.205	0.209
T605R, E694R, E706R	0.155	0.254	0.216	0.208
T605R, E706R	0.171	0.225	0.225	0.207
D56R, E694R, E706R	0.169	0.252	0.196	0.206
D568R, E694R, E706R, E655R	0.129	0.239	0.246	0.205
D568R, E694R, E706R	0.178	0.219	0.211	0.203
I67R, E63R	0.336	0.179	0.088	0.201
I67R, E694R, E655R	0.152	0.244	0.207	0.201
D568R, E694R, E655R	0.126	0.245	0.231	0.201
D56R, E655R	0.143	0.248	0.205	0.199
E63R, G71R	0.334	0.138	0.123	0.198
I67R, T605R, D568R, E655R	0.129	0.240	0.221	0.197
T605R, D568R, E694R, E706R	0.149	0.204	0.215	0.189
D56R, E706R, E655R	0.126	0.227	0.214	0.189
I67R, T605R, E655R	0.143	0.220	0.192	0.185
I67R, D56R, T605R, E655R	0.123	0.237	0.178	0.179
I67R, E694R, E706R, E655R	0.120	0.205	0.188	0.171
D56R, D568R, E706R, E655R	0.110	0.199	0.204	0.171
E694R, E655R	0.100	0.218	0.195	0.171
D56R, T605R, D568R, E694R	0.123	0.200	0.187	0.170
G71R, E706R	0.227	0.131	0.148	0.169
I67R, T605R, E706R, E655R	0.102	0.206	0.196	0.168
T605R, D568R, E655R	0.093	0.201	0.209	0.168
I67R, D56R, E694R, E655R	0.103	0.222	0.163	0.163
E706R, E655R	0.105	0.181	0.197	0.161
T605R, D568R, E706R, E655R	0.080	0.185	0.209	0.158

Substitutions	AAVS1 average indel ratio	EMX1 average indel ratio	VEGFA average indel ratio	Average indel ratio targets
I67R, D568R, E694R, E655R	0.091	0.199	0.179	0.156
D56R, E694R, E655R	0.080	0.198	0.172	0.150
E694R, E706R, E655R	0.077	0.186	0.173	0.145
D56R, E694R, E706R, E655R	0.080	0.187	0.165	0.144
D56R, T605R, D568R, E655R	0.076	0.166	0.174	0.139
D56R, T605R, E655R	0.062	0.180	0.150	0.131
D56R, D568R, E694R, E655R	0.069	0.153	0.153	0.125
T605R, E706R, E655R	0.052	0.140	0.152	0.115
D56R, T605R, E706R, E655R	0.052	0.146	0.141	0.113
T605R, D568R, E694R, E655R	0.048	0.129	0.146	0.107
I67R, T605R, E694R, E655R	0.056	0.140	0.126	0.107
T605R, E694R, E655R	0.044	0.137	0.124	0.102
T605R, E655R	0.045	0.118	0.120	0.094
T605R, E694R, E706R, E655R	0.041	0.113	0.121	0.091
E694R, E706R	0.065	0.109	0.093	0.089
D56R, T605R, E694R, E655R	0.037	0.113	0.112	0.087
Reference	0.044	0.033	0.049	0.042
E63R, D56R	0.001	0.004	0.001	0.002

As shown in **Table 6**, all but one of the CRISPR nuclease variants with combinations of amino acid substitutions exhibited higher indel activity than the reference CRISPR nuclease (SEQ ID NO: 1). 8 CRISPR nuclease variants resulted in indel ratios of over 0.3 when averaged across the three targets, indicating that over 30% of NGS reads comprised indels. These 8 CRISPR nuclease variants comprised the following substitution combinations: a) I67R, D568R, and E706R; b) I67R, D56R, and D568R; c) I67R, T605R, and D568R; d) I67R, D56R, and E706R; e) I67R, T605R, and E706R; f) I67R, D568R, and E694R; g) I67R, D56R, E694R, and E706R; and h) I67R, D56R, T605R, and E706R. 64 CRISPR nuclease variants resulted in indel ratios of over 0.2 when averaged across the three targets, indicating that over 20% of NGS reads comprised indels. 29 CRISPR nuclease variants resulted in indel ratios of over 0.1 when averaged across the three targets, indicating that over 10% of NGS reads comprised indels.

Based on this experiment, the CRISPR nuclease variant comprising the following substitutions was selected for further testing: I67R, D568R, and E706R. This CRISPR nuclease variant exhibited an over 8-fold increase in indel activity compared to the reference CRISPR nuclease.

Example 4: Editing of Human Target Genes by RNA-Guided Nuclease in HEK293T Cells and Additional RNA Scaffold Sequences

This Example describes the genomic editing of the exemplary target genes by the reference CRISPR nuclease of SEQ ID NO: 1 and the I67R, D568R, E706R variant CRISPR nuclease of SEQ ID NO: 6 when combined with RNA guides comprising alternative scaffold sequences.

RNA guides were designed using the RNA scaffold sequences of **Table 7** and cloned into a pUC19 plasmid following the U6 PolIII promoter and terminated with a 6x polyT sequence, unless noted otherwise. RNA guides were designed to be specific to the AAVS1-T3, EMX1-T7, and/or AAVS1-T3b target sequences (*see* **Example 1 and Table 7**). 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 8**. Plasmids were received from GenScript.

Table 7. RNA Guide Component Sequences

Component	Sequence
AAVS1-T3b Spacer	GAGAAAGAGAAAGGGAGUAG (SEQ ID NO: 26)
Truncated Reference Scaffold	guuacaguuaagggaaaccuuaauuguaaaacgccuauaugguaaagugauguacguuugg guauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacg auauu (SEQ ID NO: 27)
Scaffold 2	guuacaguuaagggaaaccuuaauuguaaaacgccuauacggugaagggauauacgcuugg guuuguccagccugagccucuaugccagaaauggcgccuucaucgugggguuaggaca (SEQ ID NO: 28)

Table 8. RNA Guide Sequences

Description	Sequence
Truncated Reference Scaffold (AAVS1-T3)	UUGUGAGAAUGGUGCGUCCUguuacaguuaagggaaaccuuaauuguaaaa cgccuauaugguaaagugauguacguuugggguauauaucgccagccugaaccucuacgcca gaaauggcagcuuuaucaugggguuaggacgauauu (SEQ ID NO: 29)
Truncated Reference Scaffold (EMX1-T7)	CUUUAGCAAGGGACUAUUCAguuacaguuaagggaaaccuuaauuguaaaa cgccuauaugguaaagugauguacguuugggguauauaucgccagccugaaccucuacgcca gaaauggcagcuuuaucaugggguuaggacgauauu (SEQ ID NO: 30)
Scaffold 2 (AAVS1-T3b)	GAGAAAGAGAAAGGGAGUAGguuacaguuaagggaaaccuuaauuguaaaa cgccuauacggugaagggauauacgcuuggguuuguccagccugagccucuaugccagaaa uggcgccuucaucgugggguuaggaca (SEQ ID NO: 31)

* *Spacer sequences in upper case and scaffold sequences in lower case*

The RNA guides of **Table 8** were combined with either the reference CRISPR nuclease or the I67R/D568R/E706R variant CRISPR nuclease and were introduced into HEK293T cells by lipid-based transient transfection as described in **Example 1**. The AAVS1-T3 and EMX1-T7

RNA guides from **Example 1** were further used as controls. Genomic DNA was recovered approximately 72 hours post-transfection, and samples were prepared for NGS and analyzed as described in **Example 1**. The percentage of NGS reads comprising indels shown in **Table 9** were calculated as the average of two biological replicates, which each comprised two technical replicates. The percentage of NGS reads comprising indels shown in **Table 10** were calculated as the average of three technical replicates, unless indicated otherwise.

Table 9. Indel activity of the Reference CRISPR Nuclease

RNA Guide	Average Indel Percentage
Reference Scaffold (AAVS1-T3)	4.85
Reference Scaffold (EMX1-T7)	4.10
Truncated Reference Scaffold (AAVS1-T3)	12.89*
Truncated Reference Scaffold (EMX1-T7)	9.93
No guide (AAVS1-T3)	0.10
No guide (EMX1-T7)	0.24

**One technical replicate was excluded because fewer than 10,000 NGS reads were generated*

Table 10. I67R/D568R/E706R Variant CRISPR Nuclease Indel Activity

RNA Guide	Average Indel Percentage
Reference Scaffold (AAVS1-T3)	35.90
Reference Scaffold (EMX1-T7)	28.97
Scaffold 2 (AAVS1-T3b)	36.79*
No guide (AAVS1-T3)	0.34
No guide (EMX1-T7)	0.66
No guide (AAVS1-T3b)	0.15

**One technical replicate was excluded because fewer than 10,000 NGS reads were generated*

As shown in **Table 9**, RNA guides comprising the truncated reference scaffold exhibited higher indel activity than the RNA guides comprising the reference scaffold when paired with the reference CRISPR nuclease. As shown in **Table 10**, RNA guides comprising the reference scaffold or scaffold 2 each exhibited high indel activity with the I67R, D568R, E706R variant

CRISPR nuclease. In all cases, the average percent indel observed was significantly higher than background controls comprising either the reference CRISPR nuclease or the I67R, D568R, E706R variant CRISPR nuclease in the absence of an RNA guide.

This Example thus shows that the truncated reference scaffold, and scaffold 2 are capable of being recognized by the CRISPR nuclease and a CRISPR nuclease variant. The reference scaffold and scaffold 2 were used to install edits into human gene targets in **Example 7** as examples. Scaffold 2 can be used as an exemplary scaffold to install edits into human gene targets, as described in **Examples 9 and 10**.

Example 5: Engineering and Effectiveness of CRISPR Nickase Variants for Targeting Mammalian Genes

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. D396, H397, and N420 were identified as putative catalytic residues of the HNH domain. D24, E337, and D524 positions were identified as putative catalytic residues of the RuvC domain. These positions were identified by analyzing models generated with AlphaFold2 (Jumper *et al.*, 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 PDB IDs: 5h0m, 7eu9, 6ltu, 7odf, 7lys, 8dc2, 4cmp, 4oo8, 7z4j, 5axw, 5b2o, 6kc8, 7utn, 8csz, 8ctl, 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 active site residues of the HNH and RuvC domains. A leucine mutant was also cloned for position H397. Research grade plasmids were received from GenScript. The engineered nickase sequences are shown in **Table 11**. 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 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 11. CRISPR Nuclease and Nickase Sequences

CRISPR Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
Reference	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtatacc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacaccgaccacatctttccgc gtaaggcgggtggtatcaacgtctttgcaaatctggttgcctgctg tcgtgagtgcaatgaaaaaaaaaaggccgtaccccgctggagt ccggtgtgttgccacaaccggaaattactgcgcacctggagaac gatctgaagaagaaaattctggaggacgctcggaacatgaatga gttgaacctgaattcgatcaaatacatgagctatgccagcatcgg ctggcgtcacatgcgtgaccgactccgtggtttaatcggcaacga agaactgccgattgaacgccaaagcggcatcttgaccgcgtact ttcgtcgctggtggggtttccgcaaagaaaggggcaatgacattc atcacgcactggacgcggttattttagcgtctcgcaaggattacac cgacgacggcgttgatatgacgctgaagccgaaatccagcaac ggtcgcgaattcgatccggaaaaacacctgccggataccatcag ttttaaacgtgacaaaagcaacaaaggcacgcctctgtacgaca agaatccgctttcgattaagaacagcttgattacccgtcgctggat ggttactgaaatcgaacatggtaaggaggataccattatttcggat gagtacagaaaaaagctaaaagaagcatttaagcattggggcac gagcaaaggcaaatgcctgacggatgaagaggcgcgtcagac cggcttttatcagcgcaaaaacgggaacttcatgtcactccgctg cgttgaaaaaggtctgggctggggacaaatggtgaaaatcaata acaacgcgttcaaaaccaatgtgcataatattggtgtagcagtgta	SEQ ID NO: 1 or 160 (*see* ***Table 1***)

	cctggatgagaagggtaagaaaaaggcgtatgagatcaagaac ccgcgtctgctcaaacacttcgtggaaaaaccgccggagataaa aggcaagaagctctacacactgcgtcgtggcgattatgttacgca tgaaagtgaggacagcgtttaccgtatcaaaaagctgggtacgtc cccggttctcgaggccgtggttaaaggctctattggtagaatccgt aatagcgcatcggctaccaagttgactaaggtctccaaaccgca gaaagtgttc (SEQ ID NO: 32)	
D396A (putative HNH-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtatacc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacacc**GCG**cacatctttcc gcgtaaggcgggtggtatcaacgtctttgcaaatctggttgcctg ctgtcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctgg agtccggtgtgttgccacaaccggaaattactgcgcacctggag aacgatctgaagaagaaaattctggaggacgctcggaacatgaa tgagttgaacctgaattcgatcaaatacatgagctatgccagcatc ggctggcgtcacatgcgtgaccgactccgtggtttaatcggcaa cgaagaactgccgattgaacgccaaagcggcatcttgaccgcgt actttcgtcgctggtggggtttccgcaaagaaaggggcaatgac attcatcacgcactggacgcggttattttagcgtctcgcaaggatt acaccgacgacggcgttgatatgacgctgaagccgaaatccag caacggtcgcgaattcgatccggaaaaacacctgccggatacc atcagttttaaacgtgacaaaagcaacaaaggcacgcctctgtac gacaagaatccgctttcgattaagaacagcttgattacccgtcgct ggatggttactgaaatcgaacatggtaaggaggataccattatttc ggatgagtacagaaaaaagctaaaagaagcatttaagcattggg gcacgagcaaaggcaaatgcctgacggatgaagaggcgcgtc agaccggcttttatcagcgcaaaaacgggaacttcatgtcactcc	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANT**A**HIFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 34)

	gctgcgttgaaaaaggtctgggctggggacaaatggtgaaaatc aataacaacgcgttcaaaaccaatgtgcataatattggtgtagca gtgtacctggatgagaagggtaagaaaaaggcgtatgagatcaa gaacccgcgtctgctcaaacacttcgtggaaaaaccgccggag ataaaaggcaagaagctctacacactgcgtcgtggcgattatgtt acgcatgaaagtgaggacagcgtttaccgtatcaaaaagctggg tacgtccccggttctcgaggccgtggttaaaggctctattggtaga atccgtaatagcgcatcggctaccaagttgactaaggtctccaaa ccgcagaaagtgttc (SEQ ID NO: 33)	*With no N-terminus M residue (SEQ ID NO: 169)*
H397A (putative HNH-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtatacc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacaccgac**GCG**atctttcc gcgtaaggcgggtggtatcaacgtctttgcaaatctggttgcctg ctgtcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctgg agtccggtgtgttgccacaaccggaaattactgcgcacctggag aacgatctgaagaagaaaattctggaggacgctcggaacatgaa tgagttgaacctgaattcgatcaaatacatgagctatgccagcatc ggctggcgtcacatgcgtgaccgactccgtggtttaatcggcaa cgaagaactgccgattgaacgccaaagcggcatcttgaccgcgt actttcgtcgctggtggggtttccgcaaagaaaggggcaatgac attcatcacgcactggacgcggttattttagcgtctcgcaaggatt acaccgacgacggcgttgatatgacgctgaagccgaaatccag caacggtcgcgaattcgatccggaaaaacacctgccggatacc atcagttttaaacgtgacaaaagcaacaaaggcacgcctctgtac gacaagaatccgctttcgattaagaacagcttgattacccgtcgct ggatggttactgaaatcgaacatggtaaggaggataccattatttc ggatgagtacagaaaaaagctaaaagaagcatttaagcattggg	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTD**A**IFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT

	gcacgagcaaaggcaaatgcctgacggatgaagaggcgcgtc agaccggcttttatcagcgcaaaaacgggaacttcatgtcactcc gctgcgttgaaaaaggtctgggctggggacaaatggtgaaaatc aataacaacgcgttcaaaaccaatgtgcataatattggtgtagca gtgtacctggatgagaagggtaagaaaaaggcgtatgagatcaa gaacccgcgtctgctcaaacacttcgtggaaaaaccgccggag ataaaaggcaagaagctctacacactgcgtcgtggcgattatgtt acgcatgaaagtgaggacagcgtttaccgtatcaaaaagctggg tacgtccccggttctcgaggccgtggttaaaggctctattggtaga atccgtaatagcgcatcggctaccaagttgactaaggtctccaaa ccgcagaaagtgttc (SEQ ID NO: 35)	KVSKPQKVF (SEQ ID NO: 36) *With no N-terminus M residue (SEQ ID NO: 170)*
H397L (putative HNH-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtataccc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacaccgac**CTG**atctttcc gcgtaaggcgggtggtatcaacgtctttgcaaatctggttgcctg ctgtcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctgg agtccggtgtgttgccacaaccggaaattactgcgcacctggag aacgatctgaagaagaaaattctggaggacgctcggaacatgaa tgagttgaacctgaattcgatcaaatacatgagctatgccagcatc ggctggcgtcacatgcgtgaccgactccgtggtttaatcggcaa cgaagaactgccgattgaacgccaaagcggcatcttgaccgcgt actttcgtcgctggtggggtttccgcaaagaaaggggcaatgac attcatcacgcactggacgcggttattttagcgtctcgcaaggatt acaccgacgacggcgttgatatgacgctgaagccgaaatccag caacggtcgcgaattcgatccggaaaaacacctgccggatacc atcagttttaaacgtgacaaaagcaacaaaggcacgcctctgtac gacaagaatccgctttcgattaagaacagcttgattacccgtcgct	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTD**L**IFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE

	ggatggttactgaaatcgaacatggtaaggaggataccattatttc ggatgagtacagaaaaaagctaaaagaagcatttaagcattggg gcacgagcaaaggcaaatgcctgacggatgaagaggcgcgtc agaccggcttttatcagcgcaaaaacgggaacttcatgtcactcc gctgcgttgaaaaaggtctgggctggggacaaatggtgaaaatc aataacaacgcgttcaaaaccaatgtgcataatattggtgtagca gtgtacctggatgagaagggtaagaaaaaggcgtatgagatcaa gaacccgcgtctgctcaaacacttcgtggaaaaaccgccggag ataaaaggcaagaagctctacacactgcgtcgtggcgattatgtt acgcatgaaagtgaggacagcgtttaccgtatcaaaaagctggg tacgtccccggttctcgaggccgtggttaaaggctctattggtaga atccgtaatagcgcatcggctaccaagttgactaaggtctccaaa ccgcagaaagtgttc (SEQ ID NO: 37)	SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 38) *With no N-terminus M residue (SEQ ID NO: 171)*
N420A (putative HNH-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtatacc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacaccgaccacatctttccgc gtaaggcgggtggtatcaacgtctttgcaaatctggttgcctgctg tcgtgagtgc**<u>GCG</u>**gaaaaaaaaaaaggccgtaccccgctgg agtccggtgtgttgccacaaccggaaattactgcgcacctggag aacgatctgaagaagaaaattctggaggacgctcggaacatgaa tgagttgaacctgaattcgatcaaatacatgagctatgccagcatc ggctggcgtcacatgcgtgaccgactccgtggtttaatcggcaa cgaagaactgccgattgaacgccaaagcggcatcttgaccgcgt actttcgtcgctggtggggtttccgcaaagaaaggggcaatgac attcatcacgcactggacgcggttattttagcgtctcgcaaggatt acaccgacgacggcgttgatatgacgctgaagccgaaatccag caacggtcgcgaattcgatccggaaaaacacctgccggatacc	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPR KAGGINVFANLVACCREC **<u>A</u>**EKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK

	atcagttttaaacgtgacaaaagcaacaaaggcacgcctctgtac gacaagaatccgctttcgattaagaacagcttgattacccgtcgct ggatggttactgaaatcgaacatggtaaggaggataccattatttc ggatgagtacagaaaaaagctaaaagaagcatttaagcattggg gcacgagcaaaggcaaatgcctgacggatgaagaggcgcgtc agaccggcttttatcagcgcaaaaacgggaacttcatgtcactcc gctgcgttgaaaaaggtctgggctggggacaaatggtgaaaatc aataacaacgcgttcaaaaccaatgtgcataatattggtgtagca gtgtacctggatgagaagggtaagaaaaaggcgtatgagatcaa gaacccgcgtctgctcaaacacttcgtggaaaaaccgccggag ataaaaggcaagaagctctacacactgcgtcgtggcgattatgtt acgcatgaaagtgaggacagcgtttaccgtatcaaaaagctggg tacgtccccggttctcgaggccgtggttaaaggctctattggtaga atccgtaatagcgcatcggctaccaagttgactaaggtctccaaa ccgcagaaagtgttc (SEQ ID NO: 39)	AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 40) *With no N-terminus M residue (SEQ ID NO: 172)*
D24A (putative RuvC-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatc**GCG**tacggcggtaagtata ccgggctggcggtcgtggatcgtcgcaataaccaagttctgtac gcgcgtaccgtcaagatgcgtgacgatgtagctgatattttagag ggtcgtcgggagcagcgtggtatccgtcgtaccggccagacca agaagaaacgtctacgtgaattgaagacctatttgaaaaacattg gcggtaagtacgacgaaagcatcggtacattcaagaacgaacc gtttatgaccgtgtacaaactgtgtcataagcgtggctatgactac gttgagatgccgaaagagaagacccgtgaagagattgaggctat gagcgcggatgagaagaagcaatgggaaaaagagaaaaagg agtgggaagagaccgaacgcaactctcgtcaccgtgacgaggt aacgaaggacgttcgtaaaatcatgagcgaacgcgacgcgacc gaacagcagatcagccatgttattaaaatcttcaacaaacaatatc gtccaaagcgttttaacaaccgtatcctgaccaagtgcaaagttg aagattgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacc tgctgattaaggacatcgtgcgtttttctgccgctgacgcagagcg aaaaggctgaattggaggcccacgtgctggacaaagaaggcc gcgagaaggttaaaaccttcttccgacgccaccgcattaacgagt tcctgcgcaagcagatttatgacatcgctttcaatcagaaactttct ggtcgtaccgtcttctgcaaaaagcacattctggagcacgaacaa cacacgaaggaggaacgcaaggtgttccgcttggcgccaagct tgaaaaccaagatcaataacgtgttggctgtggttaaagatgaga tccttccgcaatttggtattaataaagtgattatggagtccaacaatt ttgatatcgccgcacgtaccaagggcaaaaagagattagcaaaa gacgagtatggtaagggtcataaagaagcgaaggagtccctcat cgaggtgctgctgagagaaaccaaaaaccaatgtgtttactgcg gtaaggtgatcaccgttacgaccgcgaacaccgaccacatctttc cgcgtaaggcgggtggtatcaacgtctttgcaaatctggttgcct gctgtcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctg gagtccggtgtgttgccacaaccggaaattactgcgcacctgga gaacgatctgaagaagaaaattctggaggacgctcggaacatg aatgagttgaacctgaattcgatcaaatacatgagctatgccagc atcggctggcgtcacatgcgtgaccgactccgtggtttaatcggc aacgaagaactgccgattgaacgccaaagcggcatcttgaccg cgtactttcgtcgctggtggggtttccgcaaagaaaggggcaatg acattcatcacgcactggacgcggttattttagcgtctcgcaagga	MRNCKKDLYVEVTPMQK ELVLGI**A**YGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG

	ttacaccgacgacggcgttgatatgacgctgaagccgaaatcca gcaacggtcgcgaattcgatccggaaaaacacctgccggatac catcagttttaaacgtgacaaaagcaacaaaggcacgcctctgta cgacaagaatccgctttcgattaagaacagcttgattacccgtcg ctggatggttactgaaatcgaacatggtaaggaggataccattatt tcggatgagtacagaaaaaagctaaaagaagcatttaagcattg gggcacgagcaaaggcaaatgcctgacggatgaagaggcgc gtcagaccggcttttatcagcgcaaaaacgggaacttcatgtcac tccgctgcgttgaaaaaggtctgggctggggacaaatggtgaaa atcaataacaacgcgttcaaaaccaatgtgcataatattggtgtag cagtgtacctggatgagaagggtaagaaaaaggcgtatgagatc aagaacccgcgtctgctcaaacacttcgtggaaaaaccgccgg agataaaaggcaagaagctctacacactgcgtcgtggcgattat gttacgcatgaaagtgaggacagcgtttaccgtatcaaaaagctg ggtacgtccccggttctcgaggccgtggttaaaggctctattggt agaatccgtaatagcgcatcggctaccaagttgactaaggtctcc aaaccgcagaaagtgttc (SEQ ID NO: 41)	LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 42) *With no N-terminus M residue (SEQ ID NO: 173)*
E337A (putative RuvC-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtataccc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatg**GCG**tccaacaatttt gatatcgccgcacgtaccaagggcaaaaagagattagcaaaag acgagtatggtaagggtcataaagaagcgaaggagtccctcatc gaggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggt aaggtgatcaccgttacgaccgcgaacaccgaccacatctttcc gcgtaaggcgggtggtatcaacgtctttgcaaatctggttgcctg ctgtcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctgg agtccggtgtgttgccacaaccggaaattactgcgcacctggag aacgatctgaagaagaaaattctggaggacgctcggaacatgaa tgagttgaacctgaattcgatcaaatacatgagctatgccagcatc ggctggcgtcacatgcgtgaccgactccgtggtttaatcggcaa cgaagaactgccgattgaacgccaaagcggcatcttgaccgcgt	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIM**A**SNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHALDAVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH

	actttcgtcgctggtggggtttccgcaaagaaaggggcaatgac attcatcacgcactggacgcggttattttagcgtctcgcaaggatt acaccgacgacggcgttgatatgacgctgaagccgaaatccag caacggtcgcgaattcgatccggaaaaacacctgccggatacc atcagttttaaacgtgacaaaagcaacaaaggcacgcctctgtac gacaagaatccgctttcgattaagaacagcttgattacccgtcgct ggatggttactgaaatcgaacatggtaaggaggataccattatttc ggatgagtacagaaaaaagctaaaagaagcatttaagcattggg gcacgagcaaaggcaaatgcctgacggatgaagaggcgcgtc agaccggcttttatcagcgcaaaaacgggaacttcatgtcactcc gctgcgttgaaaaaggtctgggctggggacaaatggtgaaaatc aataacaacgcgttcaaaaccaatgtgcataatattggtgtagca gtgtacctggatgagaagggtaagaaaaaggcgtatgagatcaa gaacccgcgtctgctcaaacacttcgtggaaaaaccgccggag ataaaaggcaagaagctctacacactgcgtcgtggcgattatgtt acgcatgaaagtgaggacagcgtttaccgtatcaaaaagctggg tacgtccccggttctcgaggccgtggttaaaggctctattggtaga atccgtaatagcgcatcggctaccaagttgactaaggtctccaaa ccgcagaaagtgttc (SEQ ID NO: 43)	WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 44) *With no N-terminus M residue (SEQ ID NO: 174)*
D524A (putative RuvC-knockout)	atgaggaattgtaagaaagatctatatgtagaggttaccccgatg cagaaagagttagtcctgggtatcgattacggcggtaagtatacc gggctggcggtcgtggatcgtcgcaataaccaagttctgtacgc gcgtaccgtcaagatgcgtgacgatgtagctgatattttagaggg tcgtcgggagcagcgtggtatccgtcgtaccggccagaccaag aagaaacgtctacgtgaattgaagacctatttgaaaaacattggc ggtaagtacgacgaaagcatcggtacattcaagaacgaaccgtt tatgaccgtgtacaaactgtgtcataagcgtggctatgactacgtt gagatgccgaaagagaagacccgtgaagagattgaggctatga gcgcggatgagaagaagcaatgggaaaaagagaaaaaggagt gggaagagaccgaacgcaactctcgtcaccgtgacgaggtaac gaaggacgttcgtaaaatcatgagcgaacgcgacgcgaccgaa cagcagatcagccatgttattaaaatcttcaacaaacaatatcgtc caaagcgttttaacaaccgtatcctgaccaagtgcaaagttgaag attgtggtgtcaataccccgctgcgtaaaaacgtgcgtgacctgc tgattaaggacatcgtgcgttttctgccgctgacgcagagcgaaa aggctgaattggaggcccacgtgctggacaaagaaggccgcg agaaggttaaaaccttcttccgacgccaccgcattaacgagttcct gcgcaagcagatttatgacatcgctttcaatcagaaactttctggtc gtaccgtcttctgcaaaaagcacattctggagcacgaacaacac acgaaggaggaacgcaaggtgttccgcttggcgccaagcttga aaaccaagatcaataacgtgttggctgtggttaaagatgagatcct tccgcaatttggtattaataaagtgattatggagtccaacaattttga tatcgccgcacgtaccaagggcaaaaagagattagcaaaagac gagtatggtaagggtcataaagaagcgaaggagtccctcatcga ggtgctgctgagagaaaccaaaaaccaatgtgtttactgcggtaa ggtgatcaccgttacgaccgcgaacaccgaccacatctttccgc gtaaggcgggtggtatcaacgtctttgcaaatctggttgcctgctg tcgtgagtgcaatgaaaaaaaaaaaggccgtaccccgctggagt ccggtgtgttgccacaaccggaaattactgcgcacctggagaac gatctgaagaagaaaattctggaggacgctcggaacatgaatga gttgaacctgaattcgatcaaatacatgagctatgccagcatcgg	MRNCKKDLYVEVTPMQK ELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIG GKYDESIGTFKNEPFMTV YKLCHKRGYDYVEMPKE KTREEIEAMSADEKKQWE KEKKEWEETERNSRHRDE VTKDVRKIMSERDATEQQ ISHVIKIFNKQYRPKRFNN RILTKCKVEDCGVNTPLR KNVRDLLIKDIVRFLPLTQ SEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHI LEHEQHTKEERKVFRLAP SLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPR KAGGINVFANLVACCREC NEKKKGRTPLESGVLPQP EITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGF RKERGNDIHHAL**A**AVILA SRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFK RDKSNKGTPLYDKNPLSIK

	ctggcgtcacatgcgtgaccgactccgtggtttaatcggcaacga agaactgccgattgaacgccaaagcggcatcttgaccgcgtact ttcgtcgctggtggggtttccgcaaagaaaggggcaatgacattc atcacgcactg**<u>GCG</u>**gcggttattttagcgtctcgcaaggattac accgacgacggcgttgatatgacgctgaagccgaaatccagca acggtcgcgaattcgatccggaaaaacacctgccggataccatc agttttaaacgtgacaaaagcaacaaaggcacgcctctgtacga caagaatccgctttcgattaagaacagcttgattacccgtcgctgg atggttactgaaatcgaacatggtaaggaggataccattatttcgg atgagtacagaaaaaagctaaaagaagcatttaagcattggggc acgagcaaaggcaaatgcctgacggatgaagaggcgcgtcag accggcttttatcagcgcaaaaacgggaacttcatgtcactccgc tgcgttgaaaaaggtctgggctggggacaaatggtgaaaatcaa taacaacgcgttcaaaaccaatgtgcataatattggtgtagcagtg tacctggatgagaagggtaagaaaaaggcgtatgagatcaaga acccgcgtctgctcaaacacttcgtggaaaaaccgccggagata aaaggcaagaagctctacacactgcgtcgtggcgattatgttacg catgaaagtgaggacagcgtttaccgtatcaaaaagctgggtac gtccccggttctcgaggccgtggttaaaggctctattggtagaatc cgtaatagcgcatcggctaccaagttgactaaggtctccaaaccg cagaaagtgttc (SEQ ID NO: 45)	NSLITRRWMVTEIEHGKE DTIISDEYRKKLKEAFKH WGTSKGKCLTDEEARQTG FYQRKNGNFMSLRCVEKG LGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKK AYEIKNPRLLKHFVEKPPE IKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLT KVSKPQKVF (SEQ ID NO: 46) *With no N-terminus M residue (SEQ ID NO: 175)*

A linear DNA template sequence encoding an RNA guide was designed and ordered (IDT) with a T7 promoter upstream and a T7Te terminator sequence downstream of the guide. The RNA guide was designed to be specific to a previously tested target sequence, described in **Example 1,** within the coding exon of AAVS1 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: 49. The sequence of the encoded RNA guide and its individual components are shown in **Table 12**.

Table 12. RNA Sequences

CRISPR Nuclease	RNA sequence	SEQ ID NO
Reference sgRNA scaffold	*See Table 2*	2
AAVS1-T3 spacer	UUGUGAGAAUGGUGCGUCCU	47
T7Te terminator	*ggcucaccuucgggugggccuuucugcg*	48
T7 transcribed AAVS1-T3 sgRNA	gUUGUGAGAAUGGUGCGUCCUguuacaguuaaggcucuuuggaa acaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuugggu auauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuag gacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcu gaggcagaauuccuu*ggcucaccuucgggugggccuuucugcg*	49

A DNA target was designed and ordered as a synthesized linear DNA fragment. The target sequence from AAVS1 and 10 bases upstream and downstream within the exon was flanked by 100 bases of unrelated sequence upstream and 200 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' IR800 and 5' IR700 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 13**.

Table 13. Target gBlock and Primer Sequences

CRISPR Nuclease	DNA sequence	SEQ ID NO
AAVS1-T3 target	*See Table 2*	14
AAVS1-T3 upstream 10 nt	*AAAACTCCCT*	50
AAVS1-T3 downstream 10 nt	*AGGTGTTCAC*	51
100 nt upstream sequence	gctcaaagctttagactggaggtaccgtagtactatgagcttaatcgcgctgagggtgatttgggtgttcgtctcatatcttcgattatcggacagtagt	52
200 nt downstream sequence	agtagtatcggaccggttggcgaaccacagcgtataacgagacatggactcgcattccctaccttaagcgggcggttttggagctcggaattcatgtaaactcgcccgtttaacataatctagatggatcttgcgcgcaccgcccgaacaccactcgccacaaaaaaccgccggaacgtccaaaagtacgggttttgctg	53
Full target gblock	gctcaaagctttagactggaggtaccgtagtactatgagcttaatcgcgctgagggtgatttgggtgttcgtctcatatcttcgattatcggacagtagt*AAAACTCCCT*TTGTGAGAATGGTGCGTCCT*AGGTGTTCAC*agtagtatcggaccggttggcgaaccacagcgtataacgagacatggactcgcattccctaccttaagcgggcggttttggagctcggaattcatgtaaactcgcccgtttaacataatctagatggatcttgcgcgcaccgcccgaacaccactcgccacaaaaaaccgccggaacgtccaaaagtacgggttttgctg	54
Forward primer	/5IRD700/GCTCAAAGCTTTAGACTGGAGGTACCG	55
Reverse primer	/5IRD800/CAGCAAAACCCGTACTTTTGGACG	56

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 11** and linear DNA template for the T7 transcribed AAVS1-T3 sgRNA from **Table 12** 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 Odysssey M imaging system using the 800 nm and 700 nm channels to visualize the 5' IR800 and 5' IR700 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 800 nm channel showing cleavage of the target strand. **FIG. 2B** is a gel image captured using the 700 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 **FIGS. 2A-2D**, the reference CRISPR nuclease (SEQ ID NO: 1) cleaved both the target strand and the non-target strand, as expected. Each of the four HNH-knockout nickase constructs (D396A, H397A, H397L, and N420A) showed significantly decreased activity on the target strand while retaining activity on the non-target strand. Two of the three RuvC-knockout nickase constructs (D24A and E337A) 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 H397A variant was chosen to install edits into human gene targets, as described in **Example 6** below.

Additional variant CRISPR nuclease polypeptides are provided in **Table 14** below, each of which is also within the scope of the present disclosure.

Table 14. Variant CRISPR Nuclease Polypeptides

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
I67R, D568R, W593R, E706R	cggaactgcaagaaggacctgtacgtggaagtgacccctatgca gaaagagctggtgctgggcattgactacggcggaaagtacacag gcctggccgtggtggaccggagaaacaaccaggtgctgtacgcc agaaccgtcaagatgagagacgacgtggccgatatcctcgaagg ccggagggaacagcgaggc**AGA**agaagaacaggccagaca aagaagaagagactgagagaactgaaaacctacctgaagaacat cggaggaaaatatgacgagtctatcggcacattcaaaaacgagcc tttcatgacagtgtacaagctgtgtcacaagagaggctacgactac gtggaaatgcctaaggaaaagaccagagaagaaatcgaggccat	RNCKKDLYVEVTPMQKEL VLGIDYGGKYTGLAVVDRR NNQVLYARTVKMRDDVAD ILEGRREQRG**R**RRTGQTKK KRLRELKTYLKNIGGKYDE SIGTFKNEPFMTVYKLCHK RGYDYVEMPKEKTREEIEA MSADEKKQWEKEKKEWEE TERNSRHRDEVTKDVRKIM

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gtctgccgacgagaaaaagcagtgggaaaaagagaaaaaggag tgggaggaaaccgagcggaactccagacaccgggatgaggtga ccaaggacgtgcggaagatcatgagcgagagagatgccaccga gcagcagatcagccacgtgatcaagatcttcaacaagcagtaccg gcccaagcggttcaataatagaatcctgacaaaatgtaaggttgag gattgcggcgtgaacacacctctgagaaagaacgtcagagacct gctgatcaaggacatcgtgagattcctgcccctgacccagagcga gaaagccgagctggaagctcatgtgctggataaggaaggaaga gagaaggtgaagaccttcttcagacggcacagaatcaacgagttc ctgcgcaagcagatctacgacatcgcctttaaccaaaagctgagc ggcagaaccgtgttctgcaagaagcacatcctggagcacgagca acacaccaaagaggaacgcaaggtctttagactggcccctagcct gaaaactaagattaacaacgtgctggctgtggttaaggacgagat cctgccccagttcggcatcaacaaggtcatcatggaatccaacaa cttcgatattgccgctagaaccaagggcaagaagcggctggcca aggatgagtacggcaagggccacaaggaggccaaagagagcc tgatcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtact gcggaaaggtgatcaccgtgaccaccgccaataccgaccacatc ttccccagaaaggcgggcggcatcaacgttttcgccaacctggtg gcctgttgtagagagtgcaacgagaagaagaagggcagaacac ctctggaatccggcgtgcttccacagcctgagatcaccgcccacc tggaaaacgacctcaaaaagaagatccttgaggacgccagaaat atgaacgagctgaatctgaacagcatcaagtacatgagctacgcc agtatcggatggcggcacatgagagatagactgaggggccttatc ggtaatgaagaactgcctatcgagcgccaaagcggcatcctgac agcttacttcagacggtggtggggatttcggaaggaacggggaa atgacattcaccacgccctggacgctgtgatcctggccagcagaa aagactatacagatgacggcgtggatatgaccctgaagcccaaaa gcagcaacggccgagaattcgaccccgagaaacatctgcctgat acaatcagcttcaaaaga**AGA**aagtccaacaaaggaaccctc tgtatgataagaacccattgtctatcaagaattctctgatcacaagaa gg**CGG**atggtgacagagatcgaacacggcaaggaggacacc atcatcagcgacgagtacagaaagaagctgaaggaggctttcaa acactggggcaccagcaagggcaagtgcctgaccgacgaggaa gccagacagaccggcttctaccagagaaagaatggcaacttcatg agcctgcggtgcgtggagaagggcctgggctggggccagatgg tgaagatcaataacaacgcctttaagacgaacgtgcacaacatcg gggtggctgtttacctggacgagaagggcaagaagaaggcctac gaaattaagaaccccagactgctgaaacatttcgtg**AGA**aagc cacctgagatcaagggaaagaagctgtacaccctgcggagagg cgactacgtgacccacgagagcgaggactccgtgtacagaatca aaaaactgggcacaagccctgtgctggaagctgtggtcaagggc tctatcggccgcatccggaacagcgccagcgccacaaaactgac caaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 202)	SERDATEQQISHVIKIFNKQ YRPKRFNNRILTKCKVEDC GVNTPLRKNVRDLLIKDIVR FLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLR KQIYDIAFNQKLSGRTVFCK KHILEHEQHTKEERKVFRL APSLKTKINNVLAVVKDEIL PQFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHKE AKESLIEVLLRETKNQCVY CGKVITVTTANTDHIFPRKA GGINVFANLVACCRECNEK KKGRTPLESGVLPQPEITAH LENDLKKKILEDARNMNEL NLNSIKYMSYASIGWRHMR DRLRGLIGNEELPIERQSGIL TAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRR**R**MVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVF (SEQ ID NO: 203) *With N-terminal M residue (SEQ ID NO: 204)*
I67R, D568R, W593R, E706R with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccggaactgcaagaaggacctgtacgtgga agtgacccctatgcagaaagagctggtgctgggcattgactacgg cggaaagtacacaggcctggccgtggtggaccggagaaacaac caggtgctgtacgccagaaccgtcaagatgagagacgacgtggc cgatatcctcgaaggccggagggaacagcgaggc**AGA**agaa gaacaggccagacaaagaagaagagactgagagaactgaaaac ctacctgaagaacatcggaggaaaatatgacgagtctatcggcac	M*KRTADGSEFESPKKKRKV*R NCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADI LEGRREQRG**R**RRTGQTKKK RLRELKTYLKNIGGKYDESI GTFKNEPFMTVYKLCHKRG YDYVEMPKEKTREEIEAMS

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	attcaaaaacgagcctttcatgacagtgtacaagctgtgtcacaag agaggctacgactacgtggaaatgcctaaggaaaagaccagaga agaaatcgaggccatgtctgccgacgagaaaaagcagtgggaa aaagagaaaaaggagtgggaggaaaccgagcggaactccaga caccgggatgaggtgaccaaggacgtgcggaagatcatgagcg agagagatgccaccgagcagcagatcagccacgtgatcaagatc ttcaacaagcagtaccggcccaagcggttcaataatagaatcctga caaaatgtaaggttgaggattgcggcgtgaacacacctctgagaa agaacgtcagagacctgctgatcaaggacatcgtgagattcctgc ccctgacccagagcgagaaagccgagctggaagctcatgtgctg gataaggaaggaagagagaaggtgaagaccttcttcagacggca cagaatcaacgagttcctgcgcaagcagatctacgacatcgccttt aaccaaaagctgagcggcagaaccgtgttctgcaagaagcacat cctggagcacgagcaacacaccaaagaggaacgcaaggtcttta gactggcccctagcctgaaaactaagattaacaacgtgctggctgt ggttaaggacgagatcctgccccagttcggcatcaacaaggtcat catggaatccaacaacttcgatattgccgctagaaccaagggcaa gaagcggctggccaaggatgagtacggcaagggccacaagga ggccaaagagagcctgatcgaggtgctgctgcgggaaaccaag aaccagtgcgtgtactgcggaaaggtgatcaccgtgaccaccgc caataccgaccacatcttccccagaaaggcgggcggcatcaacg ttttcgccaacctggtggcctgttgtagagagtgcaacgagaagaa gaagggcagaacacctctggaatccggcgtgcttccacagcctg agatcaccgcccacctggaaaacgacctcaaaaagaagatccttg aggacgccagaaatatgaacgagctgaatctgaacagcatcaagt acatgagctacgccagtatcggatggcggcacatgagagataga ctgaggggccttatcggtaatgaagaactgcctatcgagcgccaa agcggcatcctgacagcttacttcagacggtggtggggatttcgg aaggaacggggaaatgacattcaccacgccctggacgctgtgat cctggccagcagaaaagactatacagatgacggcgtggatatga ccctgaagcccaaaagcagcaacggccgagaattcgaccccga gaaacatctgcctgatacaatcagcttcaaaaga**AGA**aagtcca acaaaggaacccctctgtatgataagaacccattgtctatcaagaat tctctgatcacaagaagg**CGG**atggtgacagagatcgaacac ggcaaggaggacaccatcatcagcgacgagtacagaaagaagc tgaaggaggctttcaaacactggggcaccagcaagggcaagtgc ctgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcctg ggctggggccagatggtgaagatcaataacaacgcctttaagacg aacgtgcacaacatcggggtggctgtttacctggacgagaaggg caagaagaaggcctacgaaattaagaaccccagactgctgaaac atttcgtg**AGA**aagccacctgagatcaagggaaagaagctgta caccctgcggagaggcgactacgtgacccacgagagcgaggac tccgtgtacagaatcaaaaaactgggcacaagccctgtgctggaa gctgtggtcaagggctctatcggccgcatccggaacagcgccag cgccacaaaactgaccaaggtgtccaagcctcagaaggtgttttct ggaggatctagcggaggatcctctggcagcgagacaccaggaa caagcgagtcagcaacaccagagagcagtggcggcagcagcg gcggcagcagcaaaaggccggcggccacgaaaaaggccggc caggcaaaaaagaaaaag (SEQ ID NO: 57)	ADEKKQWEKEKKEWEETE RNSRHRDEVTKDVRKIMSE RDATEQQISHVIKIFNKQYR PKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFL PLTQSEKAELEAHVLDKEG REKVKTFFRRHRINEFLRKQ IYDIAFNQKLSGRTVFCKKH ILEHEQHTKEERKVFRLAPS LKTKINNVLAVVKDEILPQF GINKVIMESNNFDIAARTKG KKRLAKDEYGKGHKEAKE SLIEVLLRETKNQCVYCGK VITVTTANTDHIFPRKAGGI NVFANLVACCRECNEKKK GRTPLESGVLPQPEITAHLE NDLKKKILEDARNMNELNL NSIKYMSYASIGWRHMRDR LRGLIGNEELPIERQSGILTA YFRRWWGFRKERGNDIHH ALDAVILASRKDYTDDGVD MTLKPKSSNGREFDPEKHL PDTISFKR**R**KSNKGTPLYDK NPLSIKNSLITRR**R**MVTEIEH GKEDTIISDEYRKKLKEAFK HWGTSKGKCLTDEEARQT GFYQRKNGNFMSLRCVEK GLGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKKA YEIKNPRLLKHFV**R**KPPEIK GKKLYTLRRGDYVTHESED SVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSK PQKVFSGGSSGGSSGSETPG TSESATPESSGGSSGGSS*KR PAATKKAGQAKKKK* (SEQ ID NO: 58) *With no C-terminal NLS (SEQ ID NO: 205)*

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
Truncated (16-775) Reference	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggcattagaagaacaggccagac aaagaagaagagactgagagaactgaaaacctacctgaagaaca tcggaggaaaatatgacgagtctatcggcacattcaaaaacgagc ctttcatgacagtgtacaagctgtgtcacaagagaggctacgacta cgtggaaatgcctaaggaaaagaccagagaagaaatcgaggcc atgtctgccgacgagaaaaagcagtgggaaaaagagaaaaagg agtgggaggaaaccgagcggaactccagacaccgggatgaggt gaccaaggacgtgcggaagatcatgagcgagagagatgccacc gagcagcagatcagccacgtgatcaagatcttcaacaagcagtac cggcccaagcggttcaataatagaatcctgacaaaatgtaaggttg aggattgcggcgtgaacacacctctgagaaagaacgtcagagac ctgctgatcaaggacatcgtgagattcctgcccctgacccagagc gagaaagccgagctggaagctcatgtgctggataaggaaggaa gagagaaggtgaagaccttcttcagacggcacagaatcaacgag ttcctgcgcaagcagatctacgacatcgcctttaaccaaaagctga gcggcagaaccgtgttctgcaagaagcacatcctggagcacgag caacacaccaaagaggaacgcaaggtctttagactggccctag cctgaaaactaagattaacaacgtgctggctgtggttaaggacga gatcctgccccagttcggcatcaacaaggtcatcatggaatccaac aacttcgatattgccgctagaaccaagggcaagaagcggctggc caaggatgagtacggcaagggccacaaggaggccaaagagag cctgatcgaggtgctgctgcgggaaaccaagaaccagtgcgtgt actgcggaaaggtgatcaccgtgaccaccgccaataccgaccac atcttccccagaaaggcgggcggcatcaacgttttcgccaacctg gtggcctgttgtagagagtgcaacgagaagaagaagggcagaa cacctctggaatccggcgtgcttccacagcctgagatcaccgccc acctggaaaacgacctcaaaaagaagatccttgaggacgccaga aatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcct tatcggtaatgaagaactgcctatcgagcgccaaagcggcatcct gacagcttacttcagacggtggtggggatttcggaaggaacggg gaaatgacattcaccacgccctggacgctgtgatcctggccagca gaaaagactatacagatgacggcgtggatatgaccctgaagccc aaaagcagcaacggccgagaattcgaccccgagaaacatctgcc tgatacaatcagcttcaaaagagataagtccaacaaaggaacccc tctgtatgataagaacccattgtctatcaagaattctctgatcacaag aaggtggatggtgacagagatcgaacacggcaaggaggacacc atcatcagcgacgagtacagaaagaagctgaaggaggctttcaa acactggggcaccagcaagggcaagtgcctgaccgacgaggaa gccagacagaccggcttctaccagagaaagaatggcaacttcatg agcctgcggtgcgtggagaagggcctgggctggggccagatgg tgaagatcaataacaacgcctttaagacgaacgtgcacaacatcg gggtggctgtttacctggacgagaagggcaagaagaaggcctac gaaattaagaaccccagactgctgaaacatttcgtggaaaagcca cctgagatcaagggaaagaagctgtacaccctgcggagaggcg actacgtgacccacgagagcgaggactccgtgtacagaatcaaa aaactgggcacaagccctgtgctggaagctgtggtcaagggctct atcggccgcatccggaacagcgccagcgccacaaaactgacca aggtgtccaagcctcagaaggtgttt (SEQ ID NO: 206)	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTDHI FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKRDKS NKGTPLYDKNPLSIKNSLIT RRWMVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FVEKPPEIKGKKLYTLRRGD YVTHESEDSVYRIKKLGTSP VLEAVVKGSIGRIRNSASAT KLTKVSKPQKVF (SEQ ID NO: 207) *With N-terminal M residue (SEQ ID NO: 208)*

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
Truncated (16-775) Reference with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggcattagaa gaacaggccagacaaagaagaagagactgagagaactgaaaac ctacctgaagaacatcggaggaaaatatgacgagtctatcggcac attcaaaaacgagcctttcatgacagtgtacaagctgtgtcacaag agaggctacgactacgtggaaatgcctaaggaaaagaccagaga agaaatcgaggccatgtctgccgacgagaaaaagcagtgggaa aaagagaaaaaggagtgggaggaaaccgagcggaactccaga caccgggatgaggtgaccaaggacgtgcggaagatcatgagcg agagagatgccaccgagcagcagatcagccacgtgatcaagatc ttcaacaagcagtaccggcccaagcggttcaataatagaatcctga caaaatgtaaggttgaggattgcggcgtgaacacacctctgagaa agaacgtcagagacctgctgatcaaggacatcgtgagattcctgc ccctgacccagagcgagaaagccgagctggaagctcatgtgctg gataaggaaggaagagagaaggtgaagaccttcttcagacggca cagaatcaacgagttcctgcgcaagcagatctacgacatcgccttt aaccaaaagctgagcggcagaaccgtgttctgcaagaagcacat cctggagcacgagcaacacaccaaagaggaacgcaaggtcttta gactggcccctagcctgaaaactaagattaacaacgtgctggctgt ggttaaggacgagatcctgccccagttcggcatcaacaaggtcat catggaatccaacaacttcgatattgccgctagaaccaagggcaa gaagcggctggccaaggatgagtacggcaagggccacaagga ggccaaagagagcctgatcgaggtgctgctgcgggaaaccaag aaccagtgcgtgtactgcggaaaggtgatcaccgtgaccaccgc caataccgaccacatcttccccagaaaggcgggcggcatcaacg ttttcgccaacctggtggcctgttgtagagagtgcaacgagaagaa gaagggcagaacacctctggaatccggcgtgcttccacagcctg agatcaccgcccacctggaaaacgacctcaaaaagaagatccttg aggacgccagaaatatgaacgagctgaatctgaacagcatcaagt acatgagctacgccagtatcggatggcggcacatgagagataga ctgaggggccttatcggtaatgaagaactgcctatcgagcgccaa agcggcatcctgacagcttacttcagacggtggtggggatttcgg aaggaacggggaaatgacattcaccacgccctggacgctgtgat cctggccagcagaaaagactatacagatgacggcgtggatatga ccctgaagcccaaaagcagcaacggccgagaattcgaccccga gaaacatctgcctgatacaatcagcttcaaaagagataagtccaac aaaggaaccctctgtatgataagaacccattgtctatcaagaattc tctgatcacaagaaggtggatggtgacagagatcgaacacggca aggaggacaccatcatcagcgacgagtacagaaagaagctgaa ggaggctttcaaacactggggcaccagcaagggcaagtgcctga ccgacgaggaagccagacagaccggcttctaccagagaaagaa tggcaacttcatgagcctgcggtgcgtggagaagggcctgggct ggggccagatggtgaagatcaataacaacgcctttaagacgaac gtgcacaacatcggggtggctgtttacctggacgagaagggcaa gaagaaggcctacgaaattaagaaccccagactgctgaaacattt cgtggaaaagccacctgagatcaagggaaagaagctgtacaccc tgcggagaggcgactacgtgacccacgagagcgaggactccgt gtacagaatcaaaaaactgggcacaagccctgtgctggaagctgt ggtcaagggctctatcggccgcatccggaacagcgccagcgcc acaaaactgaccaaggtgtccaagcctcagaaggtgttttctggag	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRGIRRTGQT KKKRLRELKTYLKNIGGKY DESIGTFKNEPFMTVYKLC HKRGYDYVEMPKEKTREEI EAMSADEKKQWEKEKKEW EETERNSRHRDEVTKDVRK IMSERDATEQQISHVIKIFNK QYRPKRFNNRILTKCKVED CGVNTPLRKNVRDLLIKDIV RFLPLTQSEKAELEAHVLD KEGREKVKTFFRRHRINEFL RKQIYDIAFNQKLSGRTVFC KKHILEHEQHTKEERKVFR LAPSLKTKINNVLAVVKDEI LPQFGINKVIMESNNFDIAA RTKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCVY CGKVITVTTANTDHIFPRKA GGINVFANLVACCRECNEK KKGRTPLESGVLPQPEITAH LENDLKKKILEDARNMNEL NLNSIKYMSYASIGWRHMR DRLRGLIGNEELPIERQSGIL TAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKRDKSNKGTPLY DKNPLSIKNSLITRRWMVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFVEKPP EIKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEAV VKGSIGRIRNSASATKLTKV SKPQKVF<u>SGGSSGGSSGSET PGTSESATPESSGGSSGGSS</u> *KRPAATKKAGQAKKKK* (SEQ ID NO: 60) *With no C-terminal NLS (SEQ ID NO: 209)*

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gatctagcggaggatcctctggcagcgagacaccaggaacaag cgagtcagcaacaccagagagcagtggcggcagcagcggcgg cagcagcaaaaggccggcggccacgaaaaaggccggccagg caaaaaagaaaaag (SEQ ID NO: 59)	
Truncated (16-775) I67R, D568R, E706R	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggc**AGA**agaagaacaggcca gacaaagaagaagagactgagagaactgaaaacctacctgaag aacatcggaggaaaatatgacgagtctatcggcacattcaaaaac gagcctttcatgacagtgtacaagctgtgtcacaagagaggctac gactacgtggaaatgcctaaggaaaagaccagagaagaaatcga ggccatgtctgccgacgagaaaaagcagtgggaaaaagagaaa aaggagtgggaggaaaccgagcggaactccagacaccgggat gaggtgaccaaggacgtgcggaagatcatgagcgagagagatg ccaccgagcagcagatcagccacgtgatcaagatcttcaacaag cagtaccggcccaagcggttcaataatagaatcctgacaaaatgta aggttgaggattgcggcgtgaacacacctctgagaaagaacgtca gagacctgctgatcaaggacatcgtgagattcctgcccctgaccc agagcgagaaagccgagctggaagctcatgtgctggataaggaa ggaagagagaaggtgaagaccttcttcagacggcacagaatcaa cgagttcctgcgcaagcagatctacgacatcgcctttaaccaaaag ctgagcggcagaaccgtgttctgcaagaagcacatcctggagca cgagcaacacaccaaagaggaacgcaaggtctttagactggccc ctagcctgaaaactaagattaacaacgtgctggctgtggttaagga cgagatcctgccccagttcggcatcaacaaggtcatcatggaatc caacaacttcgatattgccgctagaaccaagggcaagaagcggct ggccaaggatgagtacggcaagggccacaaggaggccaaaga gagcctgatcgaggtgctgctgcgggaaaccaagaaccagtgcg tgtactgcggaaaggtgatcaccgtgaccaccgccaataccgacc acatcttccccagaaaggcgggcggcatcaacgttttcgccaacc tggtggcctgttgtagagagtgcaacgagaagaagaagggcaga acacctctggaatccggcgtgcttccacagcctgagatcaccgcc cacctggaaaacgacctcaaaaagaagatccttgaggacgccag aaatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcct tatcggtaatgaagaactgcctatcgagcgccaaagcggcatcct gacagcttacttcagacggtggtggggatttcggaaggaacggg gaaatgacattcaccacgccctggacgctgtgatcctggccagca gaaaagactatacagatgacggcgtggatatgaccctgaagccc aaaagcagcaacggccgagaattcgaccccgagaaacatctgcc tgatacaatcagcttcaaaaga**AGA**aagtccaacaaaggaacc cctctgtatgataagaacccattgtctatcaagaattctctgatcaca agaaggtggatggtgacagagatcgaacacggcaaggaggaca ccatcatcagcgacgagtacagaaagaagctgaaggaggctttc aaacactggggcaccagcaagggcaagtgcctgaccgacgagg aagccagacagaccggcttctaccagagaaagaatggcaacttc atgagcctgcggtgcgtggagaagggcctgggctggggccaga tggtgaagatcaataacaacgcctttaagacgaacgtgcacaacat cggggtggctgtttacctggacgagaagggcaagaagaaggcct	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTDHI FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKR**R**KS NKGTPLYDKNPLSIKNSLIT RRWMVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FV**R**KPPEIKGKKLYTLRRG DYVTHESEDSVYRIKKLGT SPVLEAVVKGSIGRIRNSAS ATKLTKVSKPQKVF (SEQ ID NO: 211) *With N-terminal M residue (SEQ ID NO: 212)*

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	acgaaattaagaaccccagactgctgaaacatttcgtg**AGA**aa gccacctgagatcaagggaaagaagctgtacaccctgcggaga ggcgactacgtgacccacgagagcgaggactccgtgtacagaat caaaaaactgggcacaagccctgtgctggaagctgtggtcaagg gctctatcggccgcatccggaacagcgccagcgccacaaaactg accaaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 210)	
Truncated (16-775) I67R, D568R, E706R with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggc**AGA**ag aagaacaggccagacaaagaagaagagactgagagaactgaaa acctacctgaagaacatcggaggaaaatatgacgagtctatcggc acattcaaaaacgagcctttcatgacagtgtacaagctgtgtcaca agagaggctacgactacgtggaaatgcctaaggaaaagaccaga gaagaaatcgaggccatgtctgccgacgagaaaaagcagtggg aaaaagagaaaaaggagtgggaggaaaccgagcggaactcca gacaccgggatgaggtgaccaaggacgtgcggaagatcatgag cgagagagatgccaccgagcagcagatcagccacgtgatcaag atcttcaacaagcagtaccggcccaagcggttcaataatagaatcc tgacaaaatgtaaggttgaggattgcggcgtgaacacacctctga gaaagaacgtcagagacctgctgatcaaggacatcgtgagattcc tgcccctgacccagagcgagaaagccgagctggaagctcatgtg ctggataaggaaggaagagagaaggtgaagaccttcttcagacg gcacagaatcaacgagttcctgcgcaagcagatctacgacatcgc ctttaaccaaaagctgagcggcagaaccgtgttctgcaagaagca catcctggagcacgagcaacacaccaaagaggaacgcaaggtc tttagactggcccctagcctgaaaactaagattaacaacgtgctgg ctgtggttaaggacgagatcctgccccagttcggcatcaacaagg tcatcatggaatccaacaacttcgatattgccgctagaaccaaggg caagaagcggctggccaaggatgagtacggcaagggccacaa ggaggccaaagagagcctgatcgaggtgctgctgcgggaaacc aagaaccagtgcgtgtactgcggaaaggtgatcaccgtgaccac cgccaataccgaccacatcttccccagaaaggcgggcggcatca acgttttcgccaacctggtggcctgttgtagagagtgcaacgagaa gaagaagggcagaacacctctggaatccggcgtgcttccacagc ctgagatcaccgcccacctggaaaacgacctcaaaaagaagatc cttgaggacgccagaaatatgaacgagctgaatctgaacagcatc aagtacatgagctacgccagtatcggatggcggcacatgagagat agactgaggggccttatcggtaatgaagaactgcctatcgagcgc caaagcggcatcctgacagcttacttcagacggtggtggggatttc ggaaggaacggggaaatgacattcaccacgccctggacgctgtg atcctggccagcagaaaagactatacagatgacggcgtggatatg accctgaagcccaaaagcagcaacggccgagaattcgaccccg agaaacatctgcctgatacaatcagcttcaaaaga**AGA**aagtcc aacaaaggaacccctctgtatgataagaacccattgtctatcaaga attctctgatcacaagaaggtggatggtgacagagatcgaacacg gcaaggaggacaccatcatcagcgacgagtacagaaagaagct gaaggaggctttcaaacactggggcaccagcaagggcaagtgc ctgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcctg ggctggggccagatggtgaagatcaataacaacgcctttaagacg	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRG**R**RRTGQ TKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKL CHKRGYDYVEMPKEKTRE EIEAMSADEKKQWEKEKKE WEETERNSRHRDEVTKDVR KIMSERDATEQQISHVIKIFN KQYRPKRFNNRILTKCKVE DCGVNTPLRKNVRDLLIKDI VRFLPLTQSEKAELEAHVL DKEGREKVKTFFRRHRINEF LRKQIYDIAFNQKLSGRTVF CKKHILEHEQHTKEERKVF RLAPSLKTKINNVLAVVKD EILPQFGINKVIMESNNFDIA ARTKGKKRLAKDEYGKGH KEAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPRK AGGINVFANLVACCRECNE KKKGRTPLESGVLPQPEITA HLENDLKKKILEDARNMNE LNLNSIKYMSYASIGWRHM RDRLRGLIGNEELPIERQSGI LTAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRRWMVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVFSGGSSGGSSGSE TPGTSESATPESSGGSSGGS S*KRPAATKKAGQAKKKK* (SEQ ID NO: 62)

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aacgtgcacaacatcggggtggctgtttacctggacgagaaggg caagaagaaggcctacgaaattaagaaccccagactgctgaaac atttcgtg**AGA**aagccacctgagatcaagggaaagaagctgta caccctgcggagaggcgactacgtgacccacgagagcgaggac tccgtgtacagaatcaaaaaactgggcacaagccctgtgctggaa gctgtggtcaagggctctatcggccgcatccggaacagcgccag cgccacaaaactgaccaaggtgtccaagcctcagaaggtgttttct ggaggatctagcggaggatcctctggcagcgagacaccaggaa caagcgagtcagcaacaccagagagcagtggcggcagcagcg gcggcagcagcaaaaggccggcggccacgaaaaaggccggc caggcaaaaaagaaaaag (SEQ ID NO: 61)	*With no C-terminal NLS (SEQ ID NO: 213)*
Truncated (16-775) I67R, D568R, W593R, E706R	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggc**AGA**agaagaacaggcca gacaaagaagaagagactgagagaactgaaaacctacctgaag aacatcggaggaaaatatgacgagtctatcggcacattcaaaaac gagcctttcatgacagtgtacaagctgtgtcacaagagaggctac gactacgtggaaatgcctaaggaaaagaccagagaagaaatcga ggccatgtctgccgacgagaaaaagcagtgggaaaaagagaaa aaggagtgggaggaaaccgagcggaactccagacaccgggat gaggtgaccaaggacgtgcggaagatcatgagcgagagagatg ccaccgagcagcagatcagccacgtgatcaagatcttcaacaag cagtaccggcccaagcggttcaataatagaatcctgacaaaatgta aggttgaggattgcggcgtgaacacacctctgagaaagaacgtca gagacctgctgatcaaggacatcgtgagattcctgcccctgaccc agagcgagaaagccgagctggaagctcatgtgctggataaggaa ggaagagagaaggtgaagaccttcttcagacggcacagaatcaa cgagttcctgcgcaagcagatctacgacatcgcctttaaccaaaag ctgagcggcagaaccgtgttctgcaagaagcacatcctggagca cgagcaacacaccaaagaggaacgcaaggtctttagactggccc ctagcctgaaaactaagattaacaacgtgctggctgtggttaagga cgagatcctgccccagttcggcatcaacaaggtcatcatggaatc caacaacttcgatattgccgctagaaccaagggcaagaagcggct ggccaaggatgagtacggcaagggccacaaggaggccaaaga gagcctgatcgaggtgctgctgcgggaaaccaagaaccagtgcg tgtactgcggaaaggtgatcaccgtgaccaccgccaataccgacc acatcttccccagaaaggcgggcggcatcaacgttttcgccaacc tggtggcctgttgtagagagtgcaacgagaagaagaagggcaga acacctctggaatccggcgtgcttccacagcctgagatcaccgcc cacctggaaaacgacctcaaaaagaagatccttgaggacgccag aaatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcct tatcggtaatgaagaactgcctatcgagcgccaaagcggcatcct gacagcttacttcagacggtggtggggatttcggaaggaacggg gaaatgacattcaccacgccctggacgctgtgatcctggccagca gaaaagactatacagatgacggcgtggatatgaccctgaagccc aaaagcagcaacggccgagaattcgaccccgagaaacatctgcc tgatacaatcagcttcaaaaga**AGA**aagtccaacaaaggaacc cctctgtatgataagaacccattgtctatcaagaattctctgatcaca agaagg**CGG**atggtgacagagatcgaacacggcaaggagg acaccatcatcagcgacgagtacagaaagaagctgaaggaggct ttcaaacactggggcaccagcaagggcaagtgcctgaccgacga	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTDHI FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKR**R**KS NKGTPLYDKNPLSIKNSLIT RR**R**MVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FV**R**KPPEIKGKKLYTLRRG DYVTHESEDSVYRIKKLGT SPVLEAVVKGSIGRIRNSAS ATKLTKVSKPQKVF (SEQ ID NO: 215)

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ggaagccagacagaccggcttctaccagagaaagaatggcaact tcatgagcctgcggtgcgtggagaagggcctgggctggggcca gatggtgaagatcaataacaacgcctttaagacgaacgtgcacaa catcggggtggctgtttacctggacgagaagggcaagaagaagg cctacgaaattaagaaccccagactgctgaaacatttcgtg**AGA** aagccacctgagatcaagggaaagaagctgtacaccctgcggag aggcgactacgtgacccacgagagcgaggactccgtgtacaga atcaaaaaactgggcacaagccctgtgctggaagctgtggtcaag ggctctatcggccgcatccggaacagcgccagcgccacaaaact gaccaaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 214)	*With N-terminal M residue (SEQ ID NO: 216)*
Truncated (16-775) I67R, D568R, W593R, E706R with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggc**AGA**ag aagaacaggccagacaaagaagaagagactgagagaactgaaa acctacctgaagaacatcggaggaaaatatgacgagtctatcggc acattcaaaaacgagcctttcatgacagtgtacaagctgtgtcaca agagaggctacgactacgtggaaatgcctaaggaaaagaccaga gaagaaatcgaggccatgtctgccgacgagaaaaagcagtggg aaaaagagaaaaaggagtgggaggaaaccgagcggaactcca gacaccgggatgaggtgaccaaggacgtgcggaagatcatgag cgagagagatgccaccgagcagcagatcagccacgtgatcaag atcttcaacaagcagtaccggcccaagcggttcaataatagaatcc tgacaaaatgtaaggttgaggattgcggcgtgaacacacctctga gaaagaacgtcagagacctgctgatcaaggacatcgtgagattcc tgcccctgacccagagcgagaaagccgagctggaagctcatgtg ctggataaggaaggaagagagaaggtgaagaccttcttcagacg gcacagaatcaacgagttcctgcgcaagcagatctacgacatcgc ctttaaccaaaagctgagcggcagaaccgtgttctgcaagaagca catcctggagcacgagcaacacaccaaagaggaacgcaaggtc tttagactggcccctagcctgaaaactaagattaacaacgtgctgg ctgtggttaaggacgagatcctgccccagttcggcatcaacaagg tcatcatggaatccaacaacttcgatattgccgctagaaccaaggg caagaagcggctggccaaggatgagtacggcaagggccacaa ggaggccaaagagagcctgatcgaggtgctgctgcgggaaacc aagaaccagtgcgtgtactgcggaaaggtgatcaccgtgaccac cgccaataccgaccacatcttccccagaaaggcgggcggcatca acgttttcgccaacctggtggcctgttgtagagagtgcaacgagaa gaagaagggcagaacacctctggaatccggcgtgcttccacagc ctgagatcaccgcccacctggaaaacgacctcaaaaagaagatc cttgaggacgccagaaatatgaacgagctgaatctgaacagcatc aagtacatgagctacgccagtatcggatggcggcacatgagagat agactgaggggccttatcggtaatgaagaactgcctatcgagcgc caaagcggcatcctgacagcttacttcagacggtggtggggatttc ggaaggaacggggaaatgacattcaccacgccctggacgctgtg atcctggccagcagaaaagactatacagatgacggcgtggatatg accctgaagcccaaaagcagcaacggccgagaattcgaccccg agaaacatctgcctgatacaatcagcttcaaaaga**AGA**aagtcc aacaaaggaaccctctgtatgataagaacccattgtctatcaaga attctctgatcacaagaagg**CGG**atggtgacagagatcgaaca cggcaaggaggacaccatcatcagcgacgagtacagaaagaag	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRG**R**RRTGQ TKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKL CHKRGYDYVEMPKEKTRE EIEAMSADEKKQWEKEKKE WEETERNSRHRDEVTKDVR KIMSERDATEQQISHVIKIFN KQYRPKRFNNRILTKCKVE DCGVNTPLRKNVRDLLIKDI VRFLPLTQSEKAELEAHVL DKEGREKVKTFFRRHRINEF LRKQIYDIAFNQKLSGRTVF CKKHILEHEQHTKEERKVF RLAPSLKTKINNVLAVVKD EILPQFGINKVIMESNNFDIA ARTKGKKRLAKDEYGKGH KEAKESLIEVLLRETKNQCV YCGKVITVTTANTDHIFPRK AGGINVFANLVACCRECNE KKKGRTPLESGVLPQPEITA HLENDLKKKILEDARNMNE LNLNSIKYMSYASIGWRHM RDRLRGLIGNEELPIERQSGI LTAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRR**R**MVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVF<u>SGGSSGGSSGSE TPGTSESATPESSGGSSGGS</u>

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ctgaaggaggctttcaaacactggggcaccagcaagggcaagtg cctgaccgacgaggaagccagacagaccggcttctaccagaga aagaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaagac gaacgtgcacaacatcggggtggctgtttacctggacgagaagg gcaagaagaaggcctacgaaattaagaaccccagactgctgaaa catttcgtg**AGA**aagccacctgagatcaagggaaagaagctgt acaccctgcggagaggcgactacgtgacccacgagagcgagga ctccgtgtacagaatcaaaaaactgggcacaagccctgtgctgga agctgtggtcaagggctctatcggccgcatccggaacagcgcca gcgccacaaaactgaccaaggtgtccaagcctcagaaggtgtttt ctggaggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagcagc ggcggcagcagcaaaaggccggcggccacgaaaaaggccgg ccaggcaaaaaagaaaaag (SEQ ID NO: 63)	S*KRPAATKKAGQAKKKK* (SEQ ID NO: 64) *With no C-terminal NLS (SEQ ID NO: 217)*
Truncated (16-775) reference CRISPR nickase (H397A)	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggcattagaagaacaggccagac aaagaagaagagactgagagaactgaaaacctacctgaagaaca tcggaggaaaatatgacgagtctatcggcacattcaaaaacgagc ctttcatgacagtgtacaagctgtgtcacaagagaggctacgacta cgtggaaatgcctaaggaaaagaccagagaagaaatcgaggcc atgtctgccgacgagaaaaagcagtgggaaaaagagaaaaagg agtgggaggaaaccgagcggaactccagacaccgggatgaggt gaccaaggacgtgcggaagatcatgagcgagagagatgccacc gagcagcagatcagccacgtgatcaagatcttcaacaagcagtac cggcccaagcggttcaataatagaatcctgacaaaatgtaaggttg aggattgcggcgtgaacacacctctgagaaagaacgtcagagac ctgctgatcaaggacatcgtgagattcctgcccctgacccagagc gagaaagccgagctggaagctcatgtgctggataaggaaggaa gagagaaggtgaagaccttcttcagacggcacagaatcaacgag ttcctgcgcaagcagatctacgacatcgcctttaaccaaaagctga gcggcagaaccgtgttctgcaagaagcacatcctggagcacgag caacacaccaaagaggaacgcaaggtctttagactggcccctag cctgaaaactaagattaacaacgtgctggctgtggttaaggacga gatcctgccccagttcggcatcaacaaggtcatcatggaatccaac aacttcgatattgccgctagaaccaagggcaagaagcggctggc caaggatgagtacggcaagggccacaaggaggccaaagagag cctgatcgaggtgctgctgcgggaaaccaagaaccagtgcgtgt actgcggaaaggtgatcaccgtgaccaccgccaataccgac**GC** **C**atcttccccagaaaggcgggcggcatcaacgttttcgccaacct ggtggcctgttgtagagagtgcaacgagaagaagaagggcaga acacctctggaatccggcgtgcttccacagcctgagatcaccgcc cacctggaaaacgacctcaaaaagaagatccttgaggacgccag aaatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcct tatcggtaatgaagaactgcctatcgagcgccaaagcggcatcct gacagcttacttcagacggtggtggggatttcggaaggaacggg gaaatgacattcaccacgccctggacgctgtgatcctggccagca gaaaagactatacagatgacggcgtggatatgaccctgaagccc aaaagcagcaacggccgagaattcgaccccgagaaacatctgcc tgatacaatcagcttcaaaagagataagtccaacaaaggaacccc	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRGIRRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTD**AI** FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKRDKS NKGTPLYDKNPLSIKNSLIT RRWMVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FVEKPPEIKGKKLYTLRRGD YVTHESEDSVYRIKKLGTSP

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	tctgtatgataagaacccattgtctatcaagaattctctgatcacaag aaggtggatggtgacagagatcgaacacggcaaggaggacacc atcatcagcgacgagtacagaaagaagctgaaggaggctttcaa acactggggcaccagcaagggcaagtgcctgaccgacgaggaa gccagacagaccggcttctaccagagaaagaatggcaacttcatg agcctgcggtgcgtggagaagggcctgggctggggccagatgg tgaagatcaataacaacgcctttaagacgaacgtgcacaacatcg gggtggctgtttacctggacgagaagggcaagaagaaggcctac gaaattaagaaccccagactgctgaaacatttcgtggaaaagcca cctgagatcaagggaaagaagctgtacaccctgcggagaggcg actacgtgacccacgagagcgaggactccgtgtacagaatcaaa aaactgggcacaagccctgtgctggaagctgtggtcaagggctct atcggccgcatccggaacagcgccagcgccacaaaactgacca aggtgtccaagcctcagaaggtgttt (SEQ ID NO: 218)	VLEAVVKGSIGRIRNSASAT KLTKVSKPQKVF (SEQ ID NO: 219) *With N-terminal M residue* *(SEQ ID NO: 220)*
Truncated (16-775) reference CRISPR nickase (H397A) with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggcattagaa gaacaggccagacaaagaagaagagactgagagaactgaaaac ctacctgaagaacatcggaggaaaatatgacgagtctatcggcac attcaaaaacgagcctttcatgacagtgtacaagctgtgtcacaag agaggctacgactacgtggaaatgcctaaggaaaagaccagaga agaaatcgaggccatgtctgccgacgagaaaaagcagtgggaa aaagagaaaaaggagtgggaggaaaccgagcggaactccaga caccgggatgaggtgaccaaggacgtgcggaagatcatgagcg agagagatgccaccgagcagcagatcagccacgtgatcaagatc ttcaacaagcagtaccggcccaagcggttcaataatagaatcctga caaaatgtaaggttgaggattgcggcgtgaacacacctctgagaa agaacgtcagagacctgctgatcaaggacatcgtgagattcctgc ccctgacccagagcgagaaagccgagctggaagctcatgtgctg gataaggaaggaagagagaaggtgaagaccttcttcagacggca cagaatcaacgagttcctgcgcaagcagatctacgacatcgccttt aaccaaaagctgagcggcagaaccgtgttctgcaagaagcacat cctggagcacgagcaacacaccaaagaggaacgcaaggtcttta gactggcccctagcctgaaaactaagattaacaacgtgctggctgt ggttaaggacgagatcctgccccagttcggcatcaacaaggtcat catggaatccaacaacttcgatattgccgctagaaccaagggcaa gaagcggctggccaaggatgagtacggcaagggccacaagga ggccaaagagagcctgatcgaggtgctgctgcgggaaaccaag aaccagtgcgtgtactgcggaaaggtgatcaccgtgaccaccgc caataccgac**GCC**atcttccccagaaaggcgggcggcatcaa cgttttcgccaacctggtggcctgttgtagagagtgcaacgagaag aagaagggcagaacacctctggaatccggcgtgcttccacagcc tgagatcaccgcccacctggaaaacgacctcaaaaagaagatcct tgaggacgccagaaatatgaacgagctgaatctgaacagcatcaa gtacatgagctacgccagtatcggatggcggcacatgagagata gactgaggggccttatcggtaatgaagaactgcctatcgagcgcc aaagcggcatcctgacagcttacttcagacggtggtggggatttcg gaaggaacggggaaatgacattcaccacgccctggacgctgtga tcctggccagcagaaaagactatacagatgacggcgtggatatga ccctgaagcccaaaagcagcaacggccgagaattcgaccccga gaaacatctgcctgatacaatcagcttcaaaagagataagtccaac	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRGIRRTGQT KKKRLRELKTYLKNIGGKY DESIGTFKNEPFMTVYKLC HKRGYDYVEMPKEKTREEI EAMSADEKKQWEKEKKEW EETERNSRHRDEVTKDVRK IMSERDATEQQISHVIKIFNK QYRPKRFNNRILTKCKVED CGVNTPLRKNVRDLLIKDIV RFLPLTQSEKAELEAHVLD KEGREKVKTFFRRHRINEFL RKQIYDIAFNQKLSGRTVFC KKHILEHEQHTKEERKVFR LAPSLKTKINNVLAVVKDEI LPQFGINKVIMESNNFDIAA RTKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCVY CGKVITVTTANTD**A**IFPRKA GGINVFANLVACCRECNEK KKGRTPLESGVLPQPEITAH LENDLKKKILEDARNMNEL NLNSIKYMSYASIGWRHMR DRLRGLIGNEELPIERQSGIL TAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKRDKSNKGTPLY DKNPLSIKNSLITRRWMVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFVEKPP EIKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEAV

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aaaggaacccctctgtatgataagaacccattgtctatcaagaattc tctgatcacaagaaggtggatggtgacagagatcgaacacggca aggaggacaccatcatcagcgacgagtacagaaagaagctgaa ggaggctttcaaacactggggcaccagcaagggcaagtgcctga ccgacgaggaagccagacagaccggcttctaccagagaaagaa tggcaacttcatgagcctgcggtgcgtggagaagggcctgggct ggggccagatggtgaagatcaataacaacgccttttaagacgaac gtgcacaacatcggggtggctgtttacctggacgagaagggcaa gaagaaggcctacgaaattaagaaccccagactgctgaaacattt cgtggaaaagccacctgagatcaagggaaagaagctgtacaccc tgcggagaggcgactacgtgacccacgagagcgaggactccgt gtacagaatcaaaaaactgggcacaagccctgtgctggaagctgt ggtcaagggctctatcggccgcatccggaacagcgccagcgcc acaaaactgaccaaggtgtccaagcctcagaaggtgttttctggag gatctagcggaggatcctctggcagcgagacaccaggaacaag cgagtcagcaacaccagagagcagtggcggcagcagcggcgg cagcagcaaaaggccggcggccacgaaaaaggccggccagg caaaaaagaaaaag (SEQ ID NO: 221)	VKGSIGRIRNSASATKLTKV SKPQKVFSGGSSGGSSGSET PGTSESATPESSGGSSGGSS (SEQ ID NO: 222)
I67R, D568R, E706R CRISPR nickase (H397A)	cggaactgcaagaaggacctgtacgtggaagtgacccctatgca gaaagagctggtgctgggcattgactacggcggaaagtacacag gcctggccgtggtggaccggagaaacaaccaggtgctgtacgcc agaaccgtcaagatgagagacgacgtggccgatatcctcgaagg ccggagggaacagcgaggc**AGA**agaagaacaggccagaca aagaagaagagactgagagaactgaaaacctacctgaagaacat cggaggaaaatatgacgagtctatcggcacattcaaaaacgagcc tttcatgacagtgtacaagctgtgtcacaagagaggctacgactac gtggaaatgcctaaggaaaagaccagagaagaaatcgaggccat gtctgccgacgagaaaaagcagtgggaaaaagagaaaaaggag tgggaggaaaccgagcggaactccagacaccgggatgaggtga ccaaggacgtgcggaagatcatgagcgagagagatgccaccga gcagcagatcagccacgtgatcaagatcttcaacaagcagtaccg gcccaagcggttcaataatagaatcctgacaaaatgtaaggttgag gattgcggcgtgaacacacctctgagaaagaacgtcagagacct gctgatcaaggacatcgtgagattcctgcccctgacccagagcga gaaagccgagctggaagctcatgtgctggataaggaaggaaga gagaaggtgaagaccttcttcagacggcacagaatcaacgagttc ctgcgcaagcagatctacgacatcgcctttaaccaaaagctgagc ggcagaaccgtgttctgcaagaagcacatcctggagcacgagca acacaccaaagaggaacgcaaggtctttagactggccctagcct gaaaactaagattaacaacgtgctggctgtggttaaggacgagat cctgccccagttcggcatcaacaaggtcatcatggaatccaacaa cttcgatattgccgctagaaccaagggcaagaagcggctggcca aggatgagtacggcaagggccacaaggaggccaaagagagcc tgatcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtact gcggaaaggtgatcaccgtgaccaccgccaataccgac**GCC**a tcttccccagaaaggcgggcggcatcaacgttttcgccaacctggt ggcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgcccac ctggaaaacgacctcaaaaagaagatccttgaggacgccagaaa tatgaacgagctgaatctgaacagcatcaagtacatgagctacgc cagtatcggatggcggcacatgagagatagactgaggggccttat cggtaatgaagaactgcctatcgagcgccaaagcggcatcctga cagcttacttcagacggtggtggggatttcggaaggaacggggga	RNCKKDLYVEVTPMQKEL VLGIDYGGKYTGLAVVDRR NNQVLYARTVKMRDDVAD ILEGRREQRGRRRTGQTKK KRLRELKTYLKNIGGKYDE SIGTFKNEPFMTVYKLCHK RGYDYVEMPKEKTREEIEA MSADEKKQWEKEKKEWEE TERNSRHRDEVTKDVRKIM SERDATEQQISHVIKIFNKQ YRPKRFNNRILTKCKVEDC GVNTPLRKNVRDLLIKDIVR FLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLR KQIYDIAFNQKLSGRTVFCK KHILEHEQHTKEERKVFRL APSLKTKINNVLAVVKDEIL PQFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHKE AKESLIEVLLRETKNQCVY CGKVITVTTANTDAIFPRKA GGINVFANLVACCRECNEK KKGRTPLESGVLPQPEITAH LENDLKKKILEDARNMNEL NLNSIKYMSYASIGWRHMR DRLRGLIGNEELPIERQSGIL TAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKRRKSNKGTPLY DKNPLSIKNSLITRRWMVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aatgacattcaccacgccctggacgctgtgatcctggccagcaga aaagactatacagatgacggcgtggatatgaccctgaagcccaaa agcagcaacggccgagaattcgaccccgagaaacatctgcctga tacaatcagcttcaaaaga**AGA**aagtccaacaaaggaaccccct ctgtatgataagaacccattgtctatcaagaattctctgatcacaaga aggtggatggtgacagagatcgaacacggcaaggaggacacca tcatcagcgacgagtacagaaagaagctgaaggaggctttcaaa cactggggcaccagcaagggcaagtgcctgaccgacgaggaa gccagacagaccggcttctaccagagaaagaatggcaacttcatg agcctgcggtgcgtggagaagggcctgggctggggccagatgg tgaagatcaataacaacgcctttaagacgaacgtgcacaacatcg gggtggctgtttacctggacgagaagggcaagaagaaggcctac gaaattaagaaccccagactgctgaaacatttcgtg**AGA**aagc cacctgagatcaagggaaagaagctgtacaccctgcggagagg cgactacgtgacccacgagagcgaggactccgtgtacagaatca aaaaactgggcacaagccctgtgctggaagctgtggtcaagggc tctatcggccgcatccggaacagcgccagcgccacaaaactgac caaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 223)	KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFVRKP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVF (SEQ ID NO: 224) *With N-terminal M residue (SEQ ID NO: 225)*
I67R, D568R, E706R CRISPR nickase (H397A) with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccggaactgcaagaaggacctgtacgtgga agtgacccctatgcagaaagagctggtgctgggcattgactacgg cggaaagtacacaggcctggccgtggtggaccggagaaacaac caggtgctgtacgccagaaccgtcaagatgagagacgacgtggc cgatatcctcgaaggccggagggaacagcgaggc**AGA**agaa gaacaggccagacaaagaagaagagactgagagaactgaaaac ctacctgaagaacatcggaggaaaatatgacgagtctatcggcac attcaaaaacgagcctttcatgacagtgtacaagctgtgtcacaag agaggctacgactacgtggaaatgcctaaggaaaagaccagaga agaaatcgaggccatgtctgccgacgagaaaaagcagtgggaa aaagagaaaaaggagtgggaggaaaccgagcggaactccaga caccgggatgaggtgaccaaggacgtgcggaagatcatgagcg agagagatgccaccgagcagcagatcagccacgtgatcaagatc ttcaacaagcagtaccggcccaagcggttcaataatagaatcctga caaaatgtaaggttgaggattgcggcgtgaacacacctctgagaa agaacgtcagagacctgctgatcaaggacatcgtgagattcctgc ccctgacccagagcgagaaagccgagctggaagctcatgtgctg gataaggaaggaagagagaaggtgaagaccttcttcagacggca cagaatcaacgagttcctgcgcaagcagatctacgacatcgccttt aaccaaaagctgagcggcagaaccgtgttctgcaagaagcacat cctggagcacgagcaacacaccaaagaggaacgcaaggtcttta gactggcccctagcctgaaaactaagattaacaacgtgctggctgt ggttaaggacgagatcctgccccagttcggcatcaacaaggtcat catggaatccaacaacttcgatattgccgctagaaccaagggcaa gaagcggctggccaaggatgagtacggcaagggccacaagga ggccaaagagagcctgatcgaggtgctgctgcgggaaaccaag aaccagtgcgtgtactgcggaaaggtgatcaccgtgaccaccgc caataccgac**GCC**atcttccccagaaaggcgggcggcatcaa cgttttcgccaacctggtggcctgttgtagagagtgcaacgagaag aagaagggcagaacacctctggaatccggcgtgcttccacagcc tgagatcaccgcccacctggaaaacgacctcaaaaagaagatcct tgaggacgccagaaatatgaacgagctgaatctgaacagcatcaa gtacatgagctacgccagtatcggatggcggcacatgagagata	M*KRTADGSEFESPKKKRKV*R NCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADI LEGRREQRGRRRTGQTKKK RLRELKTYLKNIGGKYDESI GTFKNEPFMTVYKLCHKRG YDYVEMPKEKTREEIEAMS ADEKKQWEKEKKEWEETE RNSRHRDEVTKDVRKIMSE RDATEQQISHVIKIFNKQYR PKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFL PLTQSEKAELEAHVLDKEG REKVKTFFRRHRINEFLRKQ IYDIAFNQKLSGRTVFCKKH ILEHEQHTKEERKVFRLAPS LKTKINNVLAVVKDEILPQF GINKVIMESNNFDIAARTKG KKRLAKDEYGKGHKEAKE SLIEVLLRETKNQCVYCGK VITVTTANTDAIFPRKAGGI NVFANLVACCRECNEKKK GRTPLESGVLPQPEITAHLE NDLKKKILEDARNMNELNL NSIKYMSYASIGWRHMRDR LRGLIGNEELPIERQSGILTA YFRRWWGFRKERGNDIHH ALDAVILASRKDYTDDGVD MTLKPKSSNGREFDPEKHL PDTISFKRRKSNKGTPLYDK NPLSIKNSLITRRWMVTEIE HGKEDTIISDEYRKKLKEAF KHWGTSKGKCLTDEEARQ

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gactgaggggccttatcggtaatgaagaactgcctatcgagcgcc aaagcggcatcctgacagcttacttcagacggtggtggggatttcg gaaggaacggggaaatgacattcaccacgccctggacgctgtga tcctggccagcagaaaagactatacagatgacggcgtggatatga ccctgaagcccaaaagcagcaacggccgagaattcgaccccga gaaacatctgcctgatacaatcagcttcaaaaga**AGA**aagtcca acaaaggaacccctctgtatgataagaacccattgtctatcaagaat tctctgatcacaagaaggtggatggtgacagagatcgaacacggc aaggaggacaccatcatcagcgacgagtacagaaagaagctga aggaggctttcaaacactggggcaccagcaagggcaagtgcctg accgacgaggaagccagacagaccggcttctaccagagaaaga atggcaacttcatgagcctgcggtgcgtggagaagggcctgggc tggggccagatggtgaagatcaataacaacgcctttaagacgaac gtgcacaacatcggggtggctgtttacctggacgagaagggcaa gaagaaggcctacgaaattaagaaccccagactgctgaaacattt cgtg**AGA**aagccacctgagatcaagggaaagaagctgtacac cctgcggagaggcgactacgtgacccacgagagcgaggactcc gtgtacagaatcaaaaaactgggcacaagccctgtgctggaagct gtggtcaagggctctatcggccgcatccggaacagcgccagcgc cacaaaactgaccaaggtgtccaagcctcagaaggtgttttctgga ggatctagcggaggatcctctggcagcgagacaccaggaacaa gcgagtcagcaacaccagagagcagtggcggcagcagcggcg gcagcagcaaaaggccggcggccacgaaaaaggccggccag gcaaaaaagaaaaag (SEQ ID NO: 226)	TGFYQRKNGNFMSLRCVEK GLGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKKA YEIKNPRLLKHFVRKPPEIK GKKLYTLRRGDYVTHESED SVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSK PQKVFSGGSSGGSSGSETPG TSESATPESSGGSSGGSS*KR PAATKKAGQAKKKK* (SEQ ID NO: 227) *With no C-terminal NLS (SEQ ID NO: 228)*
Truncated (16-775) I67R, D568R, E706R CRISPR nickase (H397A)	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggc**AGA**agaagaacaggcca gacaaagaagaagagactgagagaactgaaaacctacctgaag aacatcggaggaaaatatgacgagtctatcggcacattcaaaaac gagcctttcatgacagtgtacaagctgtgtcacaagagaggctac gactacgtggaaatgcctaaggaaaagaccagagaagaaatcga ggccatgtctgccgacgagaaaaagcagtgggaaaaagagaaa aaggagtgggaggaaaccgagcggaactccagacaccgggat gaggtgaccaaggacgtgcggaagatcatgagcgagagagatg ccaccgagcagcagatcagccacgtgatcaagatcttcaacaag cagtaccggcccaagcggttcaataatagaatcctgacaaaatgta aggttgaggattgcggcgtgaacacacctctgagaaagaacgtca gagacctgctgatcaaggacatcgtgagattcctgcccctgaccc agagcgagaaagccgagctggaagctcatgtgctggataaggaa ggaagagagaaggtgaagaccttcttcagacggcacagaatcaa cgagttcctgcgcaagcagatctacgacatcgcctttaaccaaaag ctgagcggcagaaccgtgttctgcaagaagcacatcctggagca cgagcaacacaccaaagaggaacgcaaggtctttagactggccc ctagcctgaaaactaagattaacaacgtgctggctgtggttaagga cgagatcctgccccagttcggcatcaacaaggtcatcatggaatc caacaacttcgatattgccgctagaaccaagggcaagaagcggct ggccaaggatgagtacggcaagggccacaaggaggccaaaga gagcctgatcgaggtgctgctgcgggaaaccaagaaccagtgcg tgtactgcggaaaggtgatcaccgtgaccaccgccaataccgac **GCC**atcttccccagaaaggcgggcggcatcaacgttttcgcca acctggtggcctgttgtagagagtgcaacgagaagaagaagggc agaacacctctggaatccggcgtgcttccacagcctgagatcacc	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTD**A**I FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKR**R**KS

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gcccacctggaaaacgacctcaaaaagaagatccttgaggacgc cagaaatatgaacgagctgaatctgaacagcatcaagtacatgag ctacgccagtatcggatggcggcacatgagagatagactgaggg gccttatcggtaatgaagaactgcctatcgagcgccaaagcggca tcctgacagcttacttcagacggtggtggggatttcggaaggaac ggggaaatgacattcaccacgccctggacgctgtgatcctggcca gcagaaaagactatacagatgacggcgtggatatgaccctgaag cccaaaagcagcaacggccgagaattcgaccccgagaaacatct gcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaagga acccctctgtatgataagaacccattgtctatcaagaattctctgatc acaagaaggtggatggtgacagagatcgaacacggcaaggagg acaccatcatcagcgacgagtacagaaagaagctgaaggaggct ttcaaacactggggcaccagcaagggcaagtgcctgaccgacga ggaagccagacagaccggcttctaccagagaaagaatggcaact tcatgagcctgcggtgcgtggagaagggcctgggctggggcca gatggtgaagatcaataacaacgcctttaagacgaacgtgcacaa catcggggtggctgtttacctggacgagaagggcaagaagaagg cctacgaaattaagaaccccagactgctgaaacatttcgtg**AGA** aagccacctgagatcaagggaaagaagctgtacaccctgcggag aggcgactacgtgacccacgagagcgaggactccgtgtacaga atcaaaaaactgggcacaagccctgtgctggaagctgtggtcaag ggctctatcggccgcatccggaacagcgccagcgccacaaaact gaccaaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 229)	NKGTPLYDKNPLSIKNSLIT RRWMVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FV**R**KPPEIKGKKLYTLRRG DYVTHESEDSVYRIKKLGT SPVLEAVVKGSIGRIRNSAS ATKLTKVSKPQKVF (SEQ ID NO: 230) *With N-terminal M residue (SEQ ID NO: 231)*
Truncated (16-775) I67R, D568R, E706R CRISPR nickase (H397A) with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggc**AGA**ag aagaacaggccagacaaagaagaagagactgagagaactgaaa acctacctgaagaacatcggaggaaaatatgacgagtctatcggc acattcaaaaacgagcctttcatgacagtgtacaagctgtgtcaca agagaggctacgactacgtggaaatgcctaaggaaaagaccaga gaagaaatcgaggccatgtctgccgacgagaaaaagcagtggg aaaaagagaaaaaggagtgggaggaaaccgagcggaactcca gacaccgggatgaggtgaccaaggacgtgcggaagatcatgag cgagagagatgccaccgagcagcagatcagccacgtgatcaag atcttcaacaagcagtaccggcccaagcggttcaataatagaatcc tgacaaaatgtaaggttgaggattgcggcgtgaacacacctctga gaaagaacgtcagagacctgctgatcaaggacatcgtgagattcc tgcccctgacccagagcgagaaagccgagctggaagctcatgtg ctggataaggaaggaagagagaaggtgaagaccttcttcagacg gcacagaatcaacgagttcctgcgcaagcagatctacgacatcgc ctttaaccaaaagctgagcggcagaaccgtgttctgcaagaagca catcctggagcacgagcaacacaccaaagaggaacgcaaggtc tttagactggcccctagcctgaaaactaagattaacaacgtgctgg ctgtggttaaggacgagatcctgccccagttcggcatcaacaagg tcatcatggaatccaacaacttcgatattgccgctagaaccaaggg caagaagcggctggccaaggatgagtacggcaagggccacaa ggaggccaaagagagcctgatcgaggtgctgctgcgggaaacc aagaaccagtgcgtgtactgcggaaaggtgatcaccgtgaccac cgccaataccgac**GCC**atcttccccagaaaggcgggcggcat caacgttttcgccaacctggtggcctgttgtagagagtgcaacgag	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRG**R**RRTGQ TKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKL CHKRGYDYVEMPKEKTRE EIEAMSADEKKQWEKEKKE WEETERNSRHRDEVTKDVR KIMSERDATEQQISHVIKIFN KQYRPKRFNNRILTKCKVE DCGVNTPLRKNVRDLLIKDI VRFLPLTQSEKAELEAHVL DKEGREKVKTFFRRHRINEF LRKQIYDIAFNQKLSGRTVF CKKHILEHEQHTKEERKVF RLAPSLKTKINNVLAVVKD EILPQFGINKVIMESNNFDIA ARTKGKKRLAKDEYGKGH KEAKESLIEVLLRETKNQCV YCGKVITVTTANTD**A**IFPRK AGGINVFANLVACCRECNE KKKGRTPLESGVLPQPEITA HLENDLKKKILEDARNMNE LNLNSIKYMSYASIGWRHM RDRLRGLIGNEELPIERQSGI LTAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aagaagaagggcagaacacctctggaatccggcgtgcttccaca gcctgagatcaccgcccacctggaaaacgacctcaaaaagaaga tccttgaggacgccagaaatatgaacgagctgaatctgaacagca tcaagtacatgagctacgccagtatcggatggcggcacatgagag atagactgaggggccttatcggtaatgaagaactgcctatcgagc gccaaagcggcatcctgacagcttacttcagacggtggtggggat ttcggaaggaacggggaaatgacattcaccacgccctggacgct gtgatcctggccagcagaaaagactatacagatgacggcgtggat atgaccctgaagcccaaaagcagcaacggccgagaattcgaccc cgagaaacatctgcctgatacaatcagcttcaaaaga**AGA**aag tccaacaaaggaaccctctgtatgataagaacccattgtctatcaa gaattctctgatcacaagaaggtggatggtgacagagatcgaaca cggcaaggaggacaccatcatcagcgacgagtacagaaagaag ctgaaggaggctttcaaacactggggcaccagcaagggcaagtg cctgaccgacgaggaagccagacagaccggcttctaccagaga aagaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaagac gaacgtgcacaacatcggggtggctgtttacctggacgagaagg gcaagaagaaggcctacgaaattaagaaccccagactgctgaaa catttcgtg**AGA**aagccacctgagatcaagggaaagaagctgt acaccctgcggagaggcgactacgtgacccacgagagcgagga ctccgtgtacagaatcaaaaaactgggcacaagccctgtgctgga agctgtggtcaagggctctatcggccgcatccggaacagcgcca gcgccacaaaactgaccaaggtgtccaagcctcagaaggtgtttt ctggaggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagcagc ggcggcagcagcaaaaggccggcggccacgaaaaaggccgg ccaggcaaaaaagaaaaag (SEQ ID NO: 232)	HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRRWMVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVFSGGSSGGSSGSE TPGTSESATPESSGGSSGGS S (SEQ ID NO: 233)
I67R, D568R, W593R, E706R CRISPR nickase (H397A)	cggaactgcaagaaggacctgtacgtggaagtgacccctatgca gaaagagctggtgctgggcattgactacggcggaaagtacacag gcctggccgtggtggaccggagaaacaaccaggtgctgtacgcc agaaccgtcaagatgagagacgacgtggccgatatcctcgaagg ccggagggaacagcgaggc**AGA**agaagaacaggccagaca aagaagaagagactgagagaactgaaaacctacctgaagaacat cggaggaaaatatgacgagtctatcggcacattcaaaaacgagcc tttcatgacagtgtacaagctgtgtcacaagagaggctacgactac gtggaaatgcctaaggaaaagaccagagaagaaatcgaggccat gtctgccgacgagaaaaagcagtgggaaaaagagaaaaaggag tgggaggaaaccgagcggaactccagacaccgggatgaggtga ccaaggacgtgcggaagatcatgagcgagagagatgccaccga gcagcagatcagccacgtgatcaagatcttcaacaagcagtaccg gcccaagcggttcaataatagaatcctgacaaaatgtaaggttgag gattgcggcgtgaacacacctctgagaaagaacgtcagagacct gctgatcaaggacatcgtgagattcctgcccctgacccagagcga gaaagccgagctggaagctcatgtgctggataaggaaggaaga gagaaggtgaagaccttcttcagacggcacagaatcaacgagttc ctgcgcaagcagatctacgacatcgcctttaaccaaaagctgagc ggcagaaccgtgttctgcaagaagcacatcctggagcacgagca acacaccaaagaggaacgcaaggtctttagactggcccctagcct gaaaactaagattaacaacgtgctggctgtggttaaggacgagat cctgccccagttcggcatcaacaaggtcatcatggaatccaacaa cttcgatattgccgctagaaccaagggcaagaagcggctggcca aggatgagtacggcaagggccacaaggaggccaaagagagcc	RNCKKDLYVEVTPMQKEL VLGIDYGGKYTGLAVVDRR NNQVLYARTVKMRDDVAD ILEGRREQRG**R**RRTGQTKK KRLRELKTYLKNIGGKYDE SIGTFKNEPFMTVYKLCHK RGYDYVEMPKEKTREEIEA MSADEKKQWEKEKKEWEE TERNSRHRDEVTKDVRKIM SERDATEQQISHVIKIFNKQ YRPKRFNNRILTKCKVEDC GVNTPLRKNVRDLLIKDIVR FLPLTQSEKAELEAHVLDK EGREKVKTFFRRHRINEFLR KQIYDIAFNQKLSGRTVFCK KHILEHEQHTKEERKVFRL APSLKTKINNVLAVVKDEIL PQFGINKVIMESNNFDIAAR TKGKKRLAKDEYGKGHKE AKESLIEVLLRETKNQCVY CGKVITVTTANTD**A**IFPRKA GGINVFANLVACCRECNEK KKGRTPLESGVLPQPEITAH LENDLKKKILEDARNMNEL NLNSIKYMSYASIGWRHMR

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	tgatcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtact gcggaaaggtgatcaccgtgaccaccgccaataccgac**GCC**a tcttccccagaaaggcgggcggcatcaacgttttcgccaacctggt ggcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgcccac ctggaaaacgacctcaaaaagaagatccttgaggacgccagaaa tatgaacgagctgaatctgaacagcatcaagtacatgagctacgc cagtatcggatggcggcacatgagagatagactgaggggccttat cggtaatgaagaactgcctatcgagcgccaaagcggcatcctga cagcttacttcagacggtggtggggatttcggaaggaacgggga aatgacattcaccacgccctggacgctgtgatcctggccagcaga aaagactatacagatgacggcgtggatatgaccctgaagcccaaa agcagcaacggccgagaattcgaccccgagaaacatctgcctga tacaatcagcttcaaaaga**AGA**aagtccaacaaaggaacccct ctgtatgataagaacccattgtctatcaagaattctctgatcacaaga agg**CGG**atggtgacagagatcgaacacggcaaggaggacac catcatcagcgacgagtacagaaagaagctgaaggaggctttca aacactggggcaccagcaagggcaagtgcctgaccgacgagga agccagacagaccggcttctaccagagaaagaatggcaacttcat gagcctgcggtgcgtggagaagggcctgggctggggccagatg gtgaagatcaataacaacgcctttaagacgaacgtgcacaacatc ggggtggctgtttacctggacgagaagggcaagaagaaggccta cgaaattaagaaccccagactgctgaaacatttcgtg**AGA**aag ccacctgagatcaagggaaagaagctgtacaccctgcggagag gcgactacgtgacccacgagagcgaggactccgtgtacagaatc aaaaaactgggcacaagccctgtgctggaagctgtggtcaaggg ctctatcggccgcatccggaacagcgccagcgccacaaaactga ccaaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 234)	DRLRGLIGNEELPIERQSGIL TAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRR**R**MVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVF (SEQ ID NO: 235) *With N-terminal M residue (SEQ ID NO: 236)*
I67R, D568R, W593R, E706R CRISPR nickase (H397A) with NLS and linker sequences	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccggaactgcaagaaggacctgtacgtgga agtgacccctatgcagaaagagctggtgctgggcattgactacgg cggaaagtacacaggcctggccgtggtggaccggagaaacaac caggtgctgtacgccagaaccgtcaagatgagagacgacgtggc cgatatcctcgaaggccggagggaacagcgaggc**AGA**agaa gaacaggccagacaaagaagaagagactgagagaactgaaaac ctacctgaagaacatcggaggaaaatatgacgagtctatcggcac attcaaaaacgagcctttcatgacagtgtacaagctgtgtcacaag agaggctacgactacgtggaaatgcctaaggaaaagaccagaga agaaatcgaggccatgtctgccgacgagaaaaagcagtgggaa aaagagaaaaaggagtgggaggaaaccgagcggaactccaga caccgggatgaggtgaccaaggacgtgcggaagatcatgagcg agagagatgccaccgagcagcagatcagccacgtgatcaagatc ttcaacaagcagtaccggcccaagcggttcaataatagaatcctga caaaatgtaaggttgaggattgcggcgtgaacacacctctgagaa agaacgtcagagacctgctgatcaaggacatcgtgagattcctgc ccctgacccagagcgagaaagccgagctggaagctcatgtgctg gataaggaaggaagagagaaggtgaagaccttcttcagacggca cagaatcaacgagttcctgcgcaagcagatctacgacatcgccttt aaccaaaagctgagcggcagaaccgtgttctgcaagaagcacat cctggagcacgagcaacacaccaaagaggaacgcaaggtcttta gactggcccctagcctgaaaactaagattaacaacgtgctggctgt ggttaaggacgagatcctgccccagttcggcatcaacaaggtcat	M*KRTADGSEFESPKKKRKV*R NCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADI LEGRREQRG**R**RRTGQTKKK RLRELKTYLKNIGGKYDESI GTFKNEPFMTVYKLCHKRG YDYVEMPKEKTREEIEAMS ADEKKQWEKEKKEWEETE RNSRHRDEVTKDVRKIMSE RDATEQQISHVIKIFNKQYR PKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFL PLTQSEKAELEAHVLDKEG REKVKTFFRRHRINEFLRKQ IYDIAFNQKLSGRTVFCKKH ILEHEQHTKEERKVFRLAPS LKTKINNVLAVVKDEILPQF GINKVIMESNNFDIAARTKG KKRLAKDEYGKGHKEAKE SLIEVLLRETKNQCVYCGK VITVTTANTD**A**IFPRKAGGI NVFANLVACCRECNEKKK GRTPLESGVLPQPEITAHLE

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	catggaatccaacaacttcgatattgccgctagaaccaagggcaa gaagcggctggccaaggatgagtacggcaagggccacaagga ggccaaagagagcctgatcgaggtgctgctgcgggaaaccaag aaccagtgcgtgtactgcggaaaggtgatcaccgtgaccaccgc caataccgac**GCC**atcttccccagaaaggcgggcggcatcaa cgttttcgccaacctggtggcctgttgtagagagtgcaacgagaag aagaagggcagaacacctctggaatccggcgtgcttccacagcc tgagatcaccgcccacctggaaaacgacctcaaaaagaagatcct tgaggacgccagaaatatgaacgagctgaatctgaacagcatcaa gtacatgagctacgccagtatcggatggcggcacatgagagata gactgaggggccttatcggtaatgaagaactgcctatcgagcgcc aaagcggcatcctgacagcttacttcagacggtggtggggatttcg gaaggaacggggaaatgacattcaccacgccctggacgctgtga tcctggccagcagaaaagactatacagatgacggcgtggatatga ccctgaagcccaaaagcagcaacggccgagaattcgaccccga gaaacatctgcctgatacaatcagcttcaaaaga**AGA**aagtcca acaaaggaacccctctgtatgataagaacccattgtctatcaagaat tctctgatcacaagaagg**CGG**atggtgacagagatcgaacac ggcaaggaggacaccatcatcagcgacgagtacagaaagaagc tgaaggaggctttcaaacactggggcaccagcaagggcaagtgc ctgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcctg ggctggggccagatggtgaagatcaataacaacgcctttaagacg aacgtgcacaacatcggggtggctgtttacctggacgagaaggg caagaagaaggcctacgaaattaagaaccccagactgctgaaac atttcgtg**AGA**aagccacctgagatcaagggaaagaagctgta caccctgcggagaggcgactacgtgacccacgagagcgaggac tccgtgtacagaatcaaaaaactgggcacaagccctgtgctggaa gctgtggtcaagggctctatcggccgcatccggaacagcgccag cgccacaaaactgaccaaggtgtccaagcctcagaaggtgttttct ggaggatctagcggaggatcctctggcagcgagacaccaggaa caagcgagtcagcaacaccagagagcagtggcggcagcagcg gcggcagcagcaaaaggccggcggccacgaaaaaggccggc caggcaaaaaagaaaaag (SEQ ID NO: 237)	NDLKKKILEDARNMNELNL NSIKYMSYASIGWRHMRDR LRGLIGNEELPIERQSGILTA YFRRWWGFRKERGNDIHH ALDAVILASRKDYTDDGVD MTLKPKSSNGREFDPEKHL PDTISFKR**R**KSNKGTPLYDK NPLSIKNSLITRR**R**MVTEIEH GKEDTIISDEYRKKLKEAFK HWGTSKGKCLTDEEARQT GFYQRKNGNFMSLRCVEK GLGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKKA YEIKNPRLLKHFV**R**KPPEIK GKKLYTLRRGDYVTHESED SVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSK PQKVFSGGSSGGSSGSETPG TSESATPESSGGSSGGSS (SEQ ID NO: 238)
Truncated (16-775) I67R, D568R, W593R, E706R CRISPR nickase (H397A)	cagaaagagctggtgctgggcattgactacggcggaaagtacac aggcctggccgtggtggaccggagaaacaaccaggtgctgtac gccagaaccgtcaagatgagagacgacgtggccgatatcctcga aggccggagggaacagcgaggc**AGA**agaagaacaggcca gacaaagaagaagagactgagagaactgaaaacctacctgaag aacatcggaggaaaatatgacgagtctatcggcacattcaaaaac gagcctttcatgacagtgtacaagctgtgtcacaagagaggctac gactacgtggaaatgcctaaggaaaagaccagagaagaaatcga ggccatgtctgccgacgagaaaaagcagtgggaaaaagagaaa aaggagtgggaggaaaccgagcggaactccagacaccgggat gaggtgaccaaggacgtgcggaagatcatgagcgagagagatg ccaccgagcagcagatcagccacgtgatcaagatcttcaacaag cagtaccggcccaagcggttcaataatagaatcctgacaaaatgta aggttgaggattgcggcgtgaacacacctctgagaaagaacgtca gagacctgctgatcaaggacatcgtgagattcctgcccctgaccc agagcgagaaagccgagctggaagctcatgtgctggataaggaa ggaagagagaaggtgaagaccttcttcagacggcacagaatcaa cgagttcctgcgcaagcagatctacgacatcgcctttaaccaaaag ctgagcggcagaaccgtgttctgcaagaagcacatcctggagca	QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGG KYDESIGTFKNEPFMTVYK LCHKRGYDYVEMPKEKTR EEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIF NKQYRPKRFNNRILTKCKV EDCGVNTPLRKNVRDLLIK DIVRFLPLTQSEKAELEAHV LDKEGREKVKTFFRRHRIN EFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERK VFRLAPSLKTKINNVLAVV KDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGK GHKEAKESLIEVLLRETKN

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	cgagcaacacaccaaagaggaacgcaaggtctttagactggccc ctagcctgaaaactaagattaacaacgtgctggctgtggttaagga cgagatcctgccccagttcggcatcaacaaggtcatcatggaatc caacaacttcgatattgccgctagaaccaagggcaagaagcggct ggccaaggatgagtacggcaagggccacaaggaggccaaaga gagcctgatcgaggtgctgctgcgggaaaccaagaaccagtgcg tgtactgcggaaaggtgatcaccgtgaccaccgccaataccgac **GCC**atcttccccagaaaggcgggcggcatcaacgttttcgcca acctggtggcctgttgtagagagtgcaacgagaagaagaagggc agaacacctctggaatccggcgtgcttccacagcctgagatcacc gcccacctggaaaacgacctcaaaaagaagatccttgaggacgc cagaaatatgaacgagctgaatctgaacagcatcaagtacatgag ctacgccagtatcggatggcggcacatgagagatagactgaggg gccttatcggtaatgaagaactgcctatcgagcgccaaagcggca tcctgacagcttacttcagacggtggtggggatttcggaaggaac ggggaaatgacattcaccacgccctggacgctgtgatcctggcca gcagaaaagactatacagatgacggcgtggatatgaccctgaag cccaaaagcagcaacggccgagaattcgaccccgagaaacatct gcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaagga acccctctgtatgataagaacccattgtctatcaagaattctctgatc acaagaagg**CGG**atggtgacagagatcgaacacggcaagga ggacaccatcatcagcgacgagtacagaaagaagctgaaggag gctttcaaacactggggcaccagcaagggcaagtgcctgaccga cgaggaagccagacagaccggcttctaccagagaaagaatggc aacttcatgagcctgcggtgcgtggagaagggcctgggctgggg ccagatggtgaagatcaataacaacgcctttaagacgaacgtgca caacatcggggtggctgtttacctggacgagaagggcaagaaga aggcctacgaaattaagaaccccagactgctgaaacatttcgtg**A** **GA**aagccacctgagatcaagggaaagaagctgtacaccctgc ggagaggcgactacgtgacccacgagagcgaggactccgtgta cagaatcaaaaaactgggcacaagccctgtgctggaagctgtggt caagggctctatcggccgcatccggaacagcgccagcgccaca aaactgaccaaggtgtccaagcctcagaaggtgttt (SEQ ID NO: 239)	QCVYCGKVITVTTANTD**A**I FPRKAGGINVFANLVACCR ECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDAR NMNELNLNSIKYMSYASIG WRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK ERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNG REFDPEKHLPDTISFKR**R**KS NKGTPLYDKNPLSIKNSLIT RR**R**MVTEIEHGKEDTIISDE YRKKLKEAFKHWGTSKGK CLTDEEARQTGFYQRKNGN FMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYL DEKGKKKAYEIKNPRLLKH FV**R**KPPEIKGKKLYTLRRG DYVTHESEDSVYRIKKLGT SPVLEAVVKGSIGRIRNSAS ATKLTKVSKPQKVF (SEQ ID NO: 240) *With N-terminal M residue (SEQ ID NO: 241)*
Truncated (16-775) I67R, D568R, W593R, E706R CRISPR nickase (H397A)	atgaaacggacagccgacggaagcgagttcgagtcaccaaaga agaagcggaaagtccagaaagagctggtgctgggcattgactac ggcggaaagtacacaggcctggccgtggtggaccggagaaaca accaggtgctgtacgccagaaccgtcaagatgagagacgacgtg gccgatatcctcgaaggccggagggaacagcgaggc**AGA**ag aagaacaggccagacaaagaagaagagactgagagaactgaaa acctacctgaagaacatcggaggaaaatatgacgagtctatcggc acattcaaaaacgagcctttcatgacagtgtacaagctgtgtcaca agagaggctacgactacgtggaaatgcctaaggaaaagaccaga gaagaaatcgaggccatgtctgccgacgagaaaaagcagtggg aaaaagagaaaaaggagtgggaggaaaccgagcggaactcca gacaccgggatgaggtgaccaaggacgtgcggaagatcatgag cgagagagatgccaccgagcagcagatcagccacgtgatcaag atcttcaacaagcagtaccggcccaagcggttcaataatagaatcc tgacaaaatgtaaggttgaggattgcggcgtgaacacacctctga gaaagaacgtcagagacctgctgatcaaggacatcgtgagattcc tgcccctgacccagagcgagaaagccgagctggaagctcatgtg ctggataaggaaggaagagagaaggtgaagaccttcttcagacg gcacagaatcaacgagttcctgcgcaagcagatctacgacatcgc	M*KRTADGSEFESPKKKRKV*Q KELVLGIDYGGKYTGLAVV DRRNNQVLYARTVKMRDD VADILEGRREQRG**R**RRTGQ TKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKL CHKRGYDYVEMPKEKTRE EIEAMSADEKKQWEKEKKE WEETERNSRHRDEVTKDVR KIMSERDATEQQISHVIKIFN KQYRPKRFNNRILTKCKVE DCGVNTPLRKNVRDLLIKDI VRFLPLTQSEKAELEAHVL DKEGREKVKTFFRRHRINEF LRKQIYDIAFNQKLSGRTVF CKKHILEHEQHTKEERKVF RLAPSLKTKINNVLAVVKD EILPQFGINKVIMESNNFDIA ARTKGKKRLAKDEYGKGH

Nuclease polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ctttaaccaaaagctgagcggcagaaccgtgttctgcaagaagca catcctggagcacgagcaacacaccaaagaggaacgcaaggtc tttagactggcccctagcctgaaaactaagattaacaacgtgctgg ctgtggttaaggacgagatcctgccccagttcggcatcaacaagg tcatcatggaatccaacaacttcgatattgccgctagaaccaaggg caagaagcggctggccaaggatgagtacggcaagggccacaa ggaggccaaagagagcctgatcgaggtgctgctgcgggaaacc aagaaccagtgcgtgtactgcggaaaggtgatcaccgtgaccac cgccaataccgac**GCC**atcttccccagaaaggcgggcggcat caacgttttcgccaacctggtggcctgttgtagagagtgcaacgag aagaagaagggcagaacacctctggaatccggcgtgcttccaca gcctgagatcaccgcccacctggaaaacgacctcaaaaagaaga tccttgaggacgccagaaatatgaacgagctgaatctgaacagca tcaagtacatgagctacgccagtatcggatggcggcacatgagag atagactgaggggccttatcggtaatgaagaactgcctatcgagc gccaaagcggcatcctgacagcttacttcagacggtggtggggat ttcggaaggaacggggaaatgacattcaccacgccctggacgct gtgatcctggccagcagaaaagactatacagatgacggcgtggat atgaccctgaagcccaaaagcagcaacggccgagaattcgaccc cgagaaacatctgcctgatacaatcagcttcaaaaga**AGA**aag tccaacaaaggaacccctctgtatgataagaacccattgtctatcaa gaattctctgatcacaagaagg**CGG**atggtgacagagatcgaa cacggcaaggaggacaccatcatcagcgacgagtacagaaaga agctgaaggaggctttcaaacactggggcaccagcaagggcaa gtgcctgaccgacgaggaagccagacagaccggcttctaccaga gaaagaatggcaacttcatgagcctgcggtgcgtggagaagggc ctgggctggggccagatggtgaagatcaataacaacgcctttaag acgaacgtgcacaacatcggggtggctgtttacctggacgagaa gggcaagaagaaggcctacgaaattaagaaccccagactgctga aacatttcgtg**AGA**aagccacctgagatcaagggaaagaagct gtacaccctgcggagaggcgactacgtgacccacgagagcgag gactccgtgtacagaatcaaaaaactgggcacaagccctgtgctg gaagctgtggtcaagggctctatcggccgcatccggaacagcgc cagcgccacaaaactgaccaaggtgtccaagcctcagaaggtgtt ttctggaggatctagcggaggatcctctggcagcgagacaccag gaacaagcgagtcagcaacaccagagagcagtggcggcagca gcggcggcagcagcaaaaggccggcggccacgaaaaaggcc ggccaggcaaaaaagaaaaag (SEQ ID NO: 242)	KEAKESLIEVLLRETKNQCV YCGKVITVTTANTD**A**IFPRK AGGINVFANLVACCRECNE KKKGRTPLESGVLPQPEITA HLENDLKKKILEDARNMNE LNLNSIKYMSYASIGWRHM RDRLRGLIGNEELPIERQSGI LTAYFRRWWGFRKERGNDI HHALDAVILASRKDYTDDG VDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLY DKNPLSIKNSLITRR**R**MVTE IEHGKEDTIISDEYRKKLKE AFKHWGTSKGKCLTDEEA RQTGFYQRKNGNFMSLRC VEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGK KKAYEIKNPRLLKHFV**R**KP PEIKGKKLYTLRRGDYVTH ESEDSVYRIKKLGTSPVLEA VVKGSIGRIRNSASATKLTK VSKPQKVFSGGSSGGSSGSE TPGTSESATPESSGGSSGGS S (SEQ ID NO: 243)

Example 6: Fusion of CRISPR Nucleases and CRISPR Nickases to a Reverse Transcriptase

In this Example, a reverse transcriptase polypeptide was fused to the C-terminus of the reference CRISPR nuclease of SEQ ID NO: 1, the variant CRISPR nuclease of SEQ ID NO: 6, a truncated version of the CRISPR nuclease of SEQ ID NO: 6, an H397A nickase variant of the CRISPR nuclease of SEQ ID NO: 6, and an H397A nickase variant of the truncated version of the CRISPR nuclease of SEQ ID NO: 6. See **Tables 1** and **11** above.

Sequences encoding the reference CRISPR nuclease-reverse transcriptase fusion polypeptide or the I67R, D568R, E706R variant CRISPR nuclease-reverse transcriptase fusion

polypeptide were 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: 6, 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 15** below. The variant MMLV reverse transcriptase was a human codon-optimized DNA sequence. Research-grade plasmids were received from GenScript.

Table 15. Reverse Transcriptase (RT) Sequences

Component	Nucleotide Sequence	Amino Acid Sequence
Variant MMLV	atgaccctaaatatagaagatgagtatcggctacatgagacctcaaaag agccagatgtttctctagggtccacatggctgtctgattttcctcaggcct gggcggaaaccgggggcatgggactggcagttcgccaagctcctct gatcatacctctgaaagcaacctctacccccgtgtccataaaacaatac cccatgtcacaagaagccagactggggatcaagccccacatacagag actgttggaccagggaatactggtaccctgccagtccccctggaacac gcccctgctacccgttaagaaaccagggactaatgattataggcctgtc caggatctgagagaagtcaacaagcgggtggaagacatccacccca ccgtgcccaaccttacaacctcttgagcgggctcccaccgtcccacc agtggtacactgtgcttgatttaaaggatgccttttctgcctgagactcc accccaccagtcagcctctcttcgcctttgagtggagagatccagagat gggaatctcaggacaattgacctggaccagactcccacagggtttcaa aaacagtcccaccctgtttaatgaggcactgcacagagacctagcaga cttccggatccagcacccagacttgatcctgctacagtacgtggatgac ttactgctggccgccacttctgagctagactgccaacaaggtactcggg ccctgttacaaaccctagggaacctcgggtatcgggcctcggccaag aaagcccaaatttgccagaaacaggtcaagtatctggggtatcttctaa aagagggtcagagatggctgactgaggccagaaaagagactgtgat ggggcagcctactccgaagaccctcgacaactaagggagttcctag ggaaggcaggcttctgtcgcctcttcatccctgggtttgcagaaatggc agccccctgtaccctctcaccaaaccggggactctgtttaattggggc ccagaccaacaaaaggcctatcaagaaatcaagcaagctcttctaact gccccagccctggggttgccagatttgactaagccctttgaactctttgt cgacgagaagcagggctacgccaaaggtgtcctaacgcaaaaactg ggaccttggcgtcggccggtggcctacctgtccaaaaagctagaccc agtagcagctgggtggccccctttgcctacggatggtagcagccattgc cgtactgacaaaggatgcaggcaagctaaccatgggacagccactag tcattctggccccccatgcagtagaggcactagtcaaacaacccccg accgctggctttccaacgcccggatgactcactatcaggccttgcttttg gacacggaccgggtccagttcggaccggtggtagccctgaacccgg ctacgctgctcccactgcctgaggaagggctgcaacacaactgccttg atatcctggccgaagcccacggaacccgacccgacctaacggacca gccgctcccagacgccgaccacacctggtacacggatggaagcagt ctcttacaagagggacagcgtaaggcgggagctgcggtgaccaccg agaccgaggtaatctgggctaaagccctgccagccgggacatccgct cagcgggctgaactgatagcactcacccaggccctaaagatggcaga	MTLNIEDEYRLHETSKEPDV SLGSTWLSDFPQAWAETGG MGLAVRQAPLIIPLKATSTPV SIKQYPMSQEARLGIKPHIQR LLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVN KRVEDIHPTVPNPYNLLSGL PPSHQWYTVLDLKDAFFCLR LHPTSQPLFAFEWRDPEMGI SGQLTWTRLPQGFKNSPTLF NEALHRDLADFRIQHPDLILL QYVDDLLLAATSELDCQQG TRALLQTLGNLGYRASAKK AQICQKQVKYLGYLLKEGQ RWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNW GPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYL SKKLDPVAAGWPPCLRMVA AIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQF GPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQ PLPDADHTWYTDGSSLLQE GQRKAGAAVTTETEVIWAK ALPAGTSAQRAELIALTQAL KMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGK EIKNKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARGNR MADQAARKAAITETPDTSTL LIENSSP (SEQ ID NO: 66)

	aggtaagaagctaaatgtttatactgatagccgttatgcttttgctactgc ccatatccatggagaaatatacagaaggcgtgggtggctcacatcaga aggcaaagagatcaaaaataaagacgagatcttggccctactaaaag ccctctttctgcccaaaagacttagcataatccattgtccaggacatcaa aagggacacagcgccgaggctagaggcaaccggatggctgaccaa gcggcccgaaaggcagccatcacagagactccagacacctctaccct cctcatagaaaattcatcaccc (SEQ ID NO: 65)	*With N-terminus M residue (SEQ ID NO: 176)*

An H397A nickase mutation (*see* **Example 5**) was installed into the reference CRISPR nuclease-reverse transcriptase fusion polypeptide plasmid and the I67R, D568R, E706R CRISPR nuclease-reverse transcriptase fusion polypeptide plasmid. A 15 amino acid N-terminal deletion (16-775) was subsequently introduced into the I67R, D568R, E706R CRISPR nuclease-reverse transcriptase plasmid and the I67R, D568R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide plasmid. Finally, a W593R substitution was installed in the I67R, D568R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide. The nucleotide and amino acid sequences of the CRISPR nuclease-reverse transcriptase fusion polypeptides and the CRISPR nickase-reverse transcriptase fusion polypeptides are shown in **Table 16**. The codon encoding substituted residues are capitalized, bold, and underlined in the nucleotide sequence, and the substituted residues are shown in bold and underlined in the amino acid sequence. Research-grade plasmids were received from Genscript.

Table 16. Fusion Polypeptides

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
Reference CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggcatta gaagaacaggccagacaaagaagaagagactgagagaactga aaacctacctgaagaacatcggaggaaaatatgacgagtctatc ggcacattcaaaaacgagcctttcatgacagtgtacaagctgtgt cacaagagaggctacgactacgtggaaatgcctaaggaaaaga ccagagaagaaatcgaggccatgtctgccgacgagaaaaagc agtgggaaaaagagaaaaaggagtgggaggaaaccgagcgg aactccagacaccgggatgaggtgaccaaggacgtgcggaag atcatgagcgagagagatgccaccgagcagcagatcagccac gtgatcaagatcttcaacaagcagtaccggcccaagcggttcaat aatagaatcctgacaaaatgtaaggttgaggattgcggcgtgaac acacctctgagaaagaacgtcagagacctgctgatcaaggacat cgtgagattcctgcccctgacccagagcgagaaagccgagctg gaagctcatgtgctggataaggaaggaagagagaaggtgaaga ccttcttcagacggcacagaatcaacgagttcctgcgcaagcag	MKRTADGSEFESPKKKRKV RNCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADIL EGRREQRGIRRTGQTKKKRL RELKTYLKNIGGKYDESIGT FKNEPFMTVYKLCHKRGYD YVEMPKEKTREEIEAMSADE KKQWEKEKKEWEETERNSR HRDEVTKDVRKIMSERDAT EQQISHVIKIFNKQYRPKRFN NRILTKCKVEDCGVNTPLRK NVRDLLIKDIVRFLPLTQSEK AELEAHVLDKEGREKVKTFF RRHRINEFLRKQIYDIAFNQK LSGRTVFCKKHILEHEQHTK EERKVFRLAPSLKTKINNVL AVVKDEILPQFGINKVIMES NNFDIAARTKGKKRLAKDE YGKGHKEAKESLIEVLLRET

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	atctacgacatcgcctttaaccaaaagctgagcggcagaaccgt	KNQCVYCGKVITVTTANTD
	gttctgcaagaagcacatcctggagcacgagcaacacaccaaa	HIFPRKAGGINVFANLVACC
	gaggaacgcaaggtctttagactggcccctagcctgaaaactaa	RECNEKKKGRTPLESGVLPQ
	gattaacaacgtgctggctgtggttaaggacgagatcctgcccca	PEITAHLENDLKKKILEDAR
	gttcggcatcaacaaggtcatcatggaatccaacaacttcgatatt	NMNELNLNSIKYMSYASIG
	gccgctagaaccaagggcaagaagcggctggccaaggatgag	WRHMRDRLRGLIGNEELPIE
	tacggcaagggccacaaggaggccaaagagagcctgatcgag	RQSGILTAYFRRWWGFRKE
	gtgctgctgcgggaaaccaagaaccagtgcgtgtactgcggaa	RGNDIHHALDAVILASRKDY
	aggtgatcaccgtgaccaccgccaataccgaccacatcttcccc	TDDGVDMTLKPKSSNGREF
	agaaaggcgggcggcatcaacgttttcgccaacctggtggcctg	DPEKHLPDTISFKRDKSNKG
	ttgtagagagtgcaacgagaagaagaagggcagaacacctctg	TPLYDKNPLSIKNSLITRRW
	gaatccggcgtgcttccacagcctgagatcaccgcccacctgga	MVTEIEHGKEDTIISDEYRKK
	aaacgacctcaaaaagaagatccttgaggacgccagaaatatga	LKEAFKHWGTSKGKCLTDE
	acgagctgaatctgaacagcatcaagtacatgagctacgccagt	EARQTGFYQRKNGNFMSLR
	atcggatggcggcacatgagagatagactgaggggccttatcg	CVEKGLGWGQMVKINNNAF
	gtaatgaagaactgcctatcgagcgccaaagcggcatcctgaca	KTNVHNIGVAVYLDEKGKK
	gcttacttcagacggtggtggggatttcggaaggaacggggaaa	KAYEIKNPRLLKHFVEKPPEI
	tgacattcaccacgccctggacgctgtgatcctggccagcagaa	KGKKLYTLRRGDYVTHESE
	aagactatacagatgacggcgtggatatgaccctgaagcccaaa	DSVYRIKKLGTSPVLEAVVK
	agcagcaacggccgagaattcgaccccgagaaacatctgcctg	GSIGRIRNSASATKLTKVSKP
	atacaatcagcttcaaaagagataagtccaacaaaggaacccct	QKVFSGGSSGGSSGSETPGT
	ctgtatgataagaacccattgtctatcaagaattctctgatcacaag	SESATPESSGGSSGGSSTLNI
	aaggtggatggtgacagagatcgaacacggcaaggaggacac	EDEYRLHETSKEPDVSLGST
	catcatcagcgacgagtacagaaagaagctgaaggaggctttca	WLSDFPQAWAETGGMGLA
	aacactggggcaccagcaagggcaagtgcctgaccgacgagg	VRQAPLIIPLKATSTPVSIKQ
	aagccagacagaccggcttctaccagagaaagaatggcaactt	YPMSQEARLGIKPHIQRLLD
	catgagcctgcggtgcgtggagaagggcctgggctggggcca	QGILVPCQSPWNTPLLPVKK
	gatggtgaagatcaataacaacgcctttaagacgaacgtgcaca	PGTNDYRPVQDLREVNKRV
	acatcggggtggctgtttacctggacgagaagggcaagaagaa	EDIHPTVPNPYNLLSGLPPSH
	ggcctacgaaattaagaaccccagactgctgaaacatttcgtgga	QWYTVLDLKDAFFCLRLHP
	aaagccacctgagatcaagggaaagaagctgtacaccctgcgg	TSQPLFAFEWRDPEMGISGQ
	agaggcgactacgtgacccacgagagcgaggactccgtgtaca	LTWTRLPQGFKNSPTLFNEA
	gaatcaaaaaactgggcacaagccctgtgctggaagctgtggtc	LHRDLADFRIQHPDLILLQY
	aagggctctatcggccgcatccggaacagcgccagcgccaca	VDDLLLAATSELDCQQGTR
	aaactgaccaaggtgtccaagcctcagaaggtgttttctggagga	ALLQTLGNLGYRASAKKAQI
	tctagcggaggatcctctggcagcgagacaccaggaacaagcg	CQKQVKYLGYLLKEGQRWL
	agtcagcaacaccagagagcagtggcggcagcagcggcggc	TEARKETVMGQPTPKTPRQL
	agcagcaccctaaatatagaagatgagtatcggctacatgagac	REFLGKAGFCRLFIPGFAEM
	ctcaaaagagccagatgtttctctagggtccacatggctgtctgat	AAPLYPLTKPGTLFNWGPDQ
	tttcctcaggcctgggcggaaaccgggggcatgggactggcag	QKAYQEIKQALLTAPALGLP
	ttcgccaagctcctctgatcatacctctgaaagcaacctctacccc	DLTKPFELFVDEKQGYAKG
	cgtgtccataaaacaatacccccatgtcacaagaagccagactgg	VLTQKLGPWRRPVAYLSKK
	ggatcaagccccacatacagagactgttggaccagggaatactg	LDPVAAGWPPCLRMVAAIA
	gtaccctgccagtccccctggaacacgcccctgctacccgttaa	VLTKDAGKLTMGQPLVILAP
	gaaaccagggactaatgattataggcctgtccaggatctgagag	HAVEALVKQPPDRWLSNAR
	aagtcaacaagcgggtggaagacatccaccccaccgtgcccaa	MTHYQALLLDTDRVQFGPV
	cccttacaacctcttgagcgggctcccaccgtcccaccagtggta	VALNPATLLPLPEEGLQHNC
	cactgtgcttgatttaaaggatgccttttctgcctgagactccacc	LDILAEAHGTRPDLTDQPLP
	ccaccagtcagcctctcttcgcctttgagtggagagatccagaga	DADHTWYTDGSSLLQEGQR
	tgggaatctcaggacaattgacctggaccagactcccacagggt	KAGAAVTTETEVIWAKALP
	ttcaaaaacagtcccaccctgtttaatgaggcactgcacagagac	AGTSAQRAELIALTQALKM
	ctagcagacttccggatccagcacccagacttgatcctgctacag	AEGKKLNVYTDSRYAFATA
	tacgtggatgacttactgctggccgccacttctgagctagactgc	HIHGEIYRRRGWLTSEGKEI

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	caacaaggtactcgggccctgttacaaaccctagggaacctcgg gtatcgggcctcggccaagaaagcccaaatttgccagaaacag gtcaagtatctggggtatcttctaaaagagggtcagagatggctg actgaggccagaaaagagactgtgatggggcagcctactccga agacccctcgacaactaagggagttcctagggaaggcaggctt ctgtcgcctcttcatccctgggtttgcagaaatggcagcccccctg taccctctcaccaaaccggggactctgtttaattggggcccagac caacaaaaggcctatcaagaaatcaagcaagctcttctaactgcc ccagccctggggttgccagatttgactaagccctttgaactctttgt cgacgagaagcagggctacgccaaaggtgtcctaacgcaaaa actgggaccttggcgtcggccggtggcctacctgtccaaaaagc tagacccagtagcagctgggtggccccttgcctacggatggta gcagccattgccgtactgacaaaggatgcaggcaagctaaccat gggacagccactagtcattctggccccccatgcagtagaggcac tagtcaaacaaccccccgaccgctggctttccaacgcccggatg actcactatcaggccttgcttttggacacggaccgggtccagttc ggaccggtggtagccctgaacccggctacgctgctcccactgc ctgaggaagggctgcaacacaactgccttgatatcctggccgaa gcccacggaacccgacccgacctaacggaccagccgctccca gacgccgaccacacctggtacacggatggaagcagtctcttaca agagggacagcgtaaggcgggagctgcggtgaccaccgaga ccgaggtaatctgggctaaagccctgccagccgggacatccgc tcagcgggctgaactgatagcactcacccaggccctaaagatgg cagaaggtaagaagctaaatgtttatactgatagccgttatgctttt gctactgcccatatccatggagaaatatacagaaggcgtgggtg gctcacatcagaaggcaaagagatcaaaaataaagacgagatct tggccctactaaaagccctctttctgcccaaaagacttagcataat ccattgtccaggacatcaaaagggacacagcgccgaggctaga ggcaaccggatggctgaccaagcggcccgaaaggcagccatc acagagactccagacacctctaccctcctcatagaaaattcatca cccaaaaggccggcggccacgaaaaaggccggccaggcaaa aaagaaaaag (SEQ ID NO: 67)	KNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLI ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 68)
I67R, D568R, E706R CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggc**AG** **A**agaagaacaggccagacaaagaagaagagactgagagaac tgaaaacctacctgaagaacatcggaggaaaatatgacgagtct atcggcacattcaaaaacgagcctttcatgacagtgtacaagctg tgtcacaagagaggctacgactacgtggaaatgcctaaggaaaa gaccagagaagaaatcgaggccatgtctgccgacgagaaaaa gcagtgggaaaaagagaaaaaggagtgggaggaaaccgagc ggaactccagacaccgggatgaggtgaccaaggacgtgcgga agatcatgagcgagagagatgccaccgagcagcagatcagcc acgtgatcaagatcttcaacaagcagtaccggcccaagcggttc aataatagaatcctgacaaaatgtaaggttgaggattgcggcgtg aacacacctctgagaaagaacgtcagagacctgctgatcaagg acatcgtgagattcctgcccctgacccagagcgagaaagccga gctggaagctcatgtgctggataaggaaggaagagagaaggtg aagaccttcttcagacggcacagaatcaacgagttcctgcgcaa gcagatctacgacatcgcctttaaccaaaagctgagcggcagaa	MKRTADGSEFESPKKKRKV RNCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADIL EGRREQRG**R**RRTGQTKKKR LRELKTYLKNIGGKYDESIG TFKNEPFMTVYKLCHKRGY DYVEMPKEKTREEIEAMSA DEKKQWEKEKKEWEETERN SRHRDEVTKDVRKIMSERD ATEQQISHVIKIFNKQYRPKR FNNRILTKCKVEDCGVNTPL RKNVRDLLIKDIVRFLPLTQS EKAELEAHVLDKEGREKVK TFFRRHRINEFLRKQIYDIAF NQKLSGRTVFCKKHILEHEQ HTKEERKVFRLAPSLKTKIN NVLAVVKDEILPQFGINKVI MESNNFDIAARTKGKKRLA KDEYGKGHKEAKESLIEVLL RETKNQCVYCGKVITVTTA

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ccgtgttctgcaagaagcacatcctggagcacgagcaacacac	NTDHIFPRKAGGINVFANLV
	caaagaggaacgcaaggtctttagactggcccctagcctgaaaa	ACCRECNEKKKGRTPLESGV
	ctaagattaacaacgtgctggctgtggttaaggacgagatcctgc	LPQPEITAHLENDLKKKILED
	cccagttcggcatcaacaaggtcatcatggaatccaacaacttcg	ARNMNELNLNSIKYMSYASI
	atattgccgctagaaccaagggcaagaagcggctggccaagga	GWRHMRDRLRGLIGNEELPI
	tgagtacggcaagggccacaaggaggccaaagagagcctgat	ERQSGILTAYFRRWWGFRK
	cgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgc	ERGNDIHHALDAVILASRKD
	ggaaaggtgatcaccgtgaccaccgccaataccgaccacatctt	YTDDGVDMTLKPKSSNGRE
	ccccagaaaggcgggcggcatcaacgttttcgccaacctggtg	FDPEKHLPDTISFKR**R**KSNK
	gcctgttgtagagagtgcaacgagaagaagaagggcagaaca	GTPLYDKNPLSIKNSLITRRW
	cctctggaatccggcgtgcttccacagcctgagatcaccgccca	MVTEIEHGKEDTIISDEYRKK
	cctggaaaacgacctcaaaaagaagatccttgaggacgccaga	LKEAFKHWGTSKGKCLTDE
	aatatgaacgagctgaatctgaacagcatcaagtacatgagctac	EARQTGFYQRKNGNFMSLR
	gccagtatcggatggcggcacatgagagatagactgaggggcc	CVEKGLGWGQMVKINNNAF
	ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc	KTNVHNIGVAVYLDEKGKK
	ctgacagcttacttcagacggtggtggggatttcggaaggaacg	KAYEIKNPRLLKHFV**R**KPPEI
	gggaaatgacattcaccacgccctggacgctgtgatcctggcca	KGKKLYTLRRGDYVTHESE
	gcagaaaagactatacagatgacggcgtggatatgaccctgaa	DSVYRIKKLGTSPVLEAVVK
	gcccaaaagcagcaacggccgagaattcgaccccgagaaaca	GSIGRIRNSASATKLTKVSKP
	tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa	QKVFSGGSSGGSSGSETPGT
	ggaacccctctgtatgataagaacccattgtctatcaagaattctct	SESATPESSGGSSGGSSTLNI
	gatcacaagaaggtggatggtgacagagatcgaacacggcaa	EDEYRLHETSKEPDVSLGST
	ggaggacaccatcatcagcgacgagtacagaaagaagctgaa	WLSDFPQAWAETGGMGLA
	ggaggctttcaaacactggggcaccagcaagggcaagtgcctg	VRQAPLIIPLKATSTPVSIKQ
	accgacgaggaagccagacagaccggcttctaccagagaaag	YPMSQEARLGIKPHIQRLLD
	aatggcaacttcatgagcctgcggtgcgtggagaagggcctgg	QGILVPCQSPWNTPLLPVKK
	gctggggccagatggtgaagatcaataacaacgcctttaagacg	PGTNDYRPVQDLREVNKRV
	aacgtgcacaacatcggggtggctgtttacctggacgagaagg	EDIHPTVPNPYNLLSGLPPSH
	gcaagaagaaggcctacgaaattaagaaccccagactgctgaa	QWYTVLDLKDAFFCLRLHP
	acatttcgtg**AGA**aagccacctgagatcaagggaaagaagct	TSQPLFAFEWRDPEMGISGQ
	gtacaccctgcggagaggcgactacgtgacccacgagagcga	LTWTRLPQGFKNSPTLFNEA
	ggactccgtgtacagaatcaaaaaactgggcacaagccctgtgc	LHRDLADFRIQHPDLILLQY
	tggaagctgtggtcaagggctctatcggccgcatccggaacagc	VDDLLLAATSELDCQQGTR
	gccagcgccacaaaactgaccaaggtgtccaagcctcagaagg	ALLQTLGNLGYRASAKKAQI
	tgttttctggaggatctagcggaggatcctctggcagcgagacac	CQKQVKYLGYLLKEGQRWL
	caggaacaagcgagtcagcaacaccagagagcagtggcggc	TEARKETVMGQPTPKTPRQL
	agcagcggcggcagcagcaccctaaatatagaagatgagtatc	REFLGKAGFCRLFIPGFAEM
	ggctacatgagacctcaaaagagccagatgtttctctagggtcca	AAPLYPLTKPGTLFNWGPDQ
	catggctgtctgattttcctcaggcctgggcggaaaccgggggc	QKAYQEIKQALLTAPALGLP
	atgggactggcagttcgccaagctcctctgatcatacctctgaaa	DLTKPFELFVDEKQGYAKG
	gcaacctctacccccgtgtccataaaacaataccccatgtcacaa	VLTQKLGPWRRPVAYLSKK
	gaagccagactggggatcaagccccacatacagagactgttgg	LDPVAAGWPPCLRMVAAIA
	accagggaatactggtaccctgccagtccccctggaacacgcc	VLTKDAGKLTMGQPLVILAP
	cctgctacccgttaagaaaccagggactaatgattataggcctgt	HAVEALVKQPPDRWLSNAR
	ccaggatctgagagaagtcaacaagcgggtggaagacatccac	MTHYQALLLDTDRVQFGPV
	cccaccgtgcccaacccttacaacctcttgagcgggctcccacc	VALNPATLLPLPEEGLQHNC
	gtcccaccagtggtacactgtgcttgatttaaaggatgccttttct	LDILAEAHGTRPDLTDQPLP
	gcctgagactccaccccaccagtcagcctctcttcgcctttgagt	DADHTWYTDGSSLLQEGQR
	ggagagatccagagatgggaatctcaggacaattgacctggac	KAGAAVTTETEVIWAKALP
	cagactcccacagggtttcaaaaacagtcccaccctgtttaatga	AGTSAQRAELIALTQALKM
	ggcactgcacagagacctagcagacttccggatccagcaccca	AEGKKLNVYTDSRYAFATA
	gacttgatcctgctacagtacgtggatgacttactgctggccgcc	HIHGEIYRRRGWLTSEGKEI
	acttctgagctagactgccaacaaggtactcgggccctgttacaa	KNKDEILALLKALFLPKRLSI

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gcccttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 69)	IHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLI ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 70)
Truncated (16-775) I67R, D568R, E706R CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggc**A** **GA**agaagaacaggccagacaaagaagaagagactgagaga actgaaaacctacctgaagaacatcggaggaaaatatgacgagt ctatcggcacattcaaaaacgagcctttcatgacagtgtacaagct gtgtcacaagagaggctacgactacgtggaaatgcctaaggaa aagaccagagaagaaatcgaggccatgtctgccgacgagaaa aagcagtgggaaaaagagaaaaaggagtgggaggaaaccga gcggaactccagacaccgggatgaggtgaccaaggacgtgcg gaagatcatgagcgagagagatgccaccgagcagcagatcag ccacgtgatcaagatcttcaacaagcagtaccggcccaagcggt tcaataatagaatcctgacaaaatgtaaggttgaggattgcggcg tgaacacacctctgagaaagaacgtcagagacctgctgatcaag gacatcgtgagattcctgccccctgacccagagcgagaaagccg agctggaagctcatgtgctggataaggaaggaagagagaaggt gaagaccttcttcagacggcacagaatcaacgagttcctgcgca agcagatctacgacatcgcctttaaccaaaagctgagcggcaga accgtgttctgcaagaagcacatcctggagcacgagcaacaca ccaaagaggaacgcaaggtctttagactggccctagcctgaaa	MKRTADGSEFESPKKKRKV QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKLC HKRGYDYVEMPKEKTREEIE AMSADEKKQWEKEKKEWE ETERNSRHRDEVTKDVRKIM SERDATEQQISHVIKIFNKQY RPKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFLP LTQSEKAELEAHVLDKEGRE KVKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHIL EHEQHTKEERKVFRLAPSLK TKINNVLAVVKDEILPQFGIN KVIMESNNFDIAARTKGKKR LAKDEYGKGHKEAKESLIEV LLRETKNQCVYCGKVITVTT ANTDHIFPRKAGGINVFANL VACCRECNEKKKGRTPLESG

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	actaagattaacaacgtgctggctgtggttaaggacgagatcctg	VLPQPEITAHLENDLKKKILE
	ccccagttcggcatcaacaaggtcatcatggaatccaacaacttc	DARNMNELNLNSIKYMSYA
	gatattgccgctagaaccaagggcaagaagcggctggccaag	SIGWRHMRDRLRGLIGNEEL
	gatgagtacggcaagggccacaaggaggccaaagagagcctg	PIERQSGILTAYFRRWWGFR
	atcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactg	KERGNDIHHALDAVILASRK
	cggaaaggtgatcaccgtgaccaccgccaataccgaccacatct	DYTDDGVDMTLKPKSSNGR
	tccccagaaaggcgggcggcatcaacgttttcgccaacctggtg	EFDPEKHLPDTISFKR**R**KSNK
	gcctgttgtagagagtgcaacgagaagaagaagggcagaaca	GTPLYDKNPLSIKNSLITRRW
	cctctggaatccggcgtgcttccacagcctgagatcaccgccca	MVTEIEHGKEDTIISDEYRKK
	cctggaaaacgacctcaaaaagaagatccttgaggacgccaga	LKEAFKHWGTSKGKCLTDE
	aatatgaacgagctgaatctgaacagcatcaagtacatgagctac	EARQTGFYQRKNGNFMSLR
	gccagtatcggatggcggcacatgagagatagactgaggggcc	CVEKGLGWGQMVKINNNAF
	ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc	KTNVHNIGVAVYLDEKGKK
	ctgacagcttacttcagacggtggtggggatttcggaaggaacg	KAYEIKNPRLLKHFV**R**KPPEI
	gggaaatgacattcaccacgccctggacgctgtgatcctggcca	KGKKLYTLRRGDYVTHESE
	gcagaaaagactatacagatgacggcgtggatatgaccctgaa	DSVYRIKKLGTSPVLEAVVK
	gcccaaaagcagcaacggccgagaattcgaccccgagaaaca	GSIGRIRNSASATKLTKVSKP
	tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa	QKVFSGGSSGGSSGSETPGT
	ggaacccctctgtatgataagaacccattgtctatcaagaattctct	SESATPESSGGSSGGSSTLNI
	gatcacaagaaggtggatggtgacagagatcgaacacggcaa	EDEYRLHETSKEPDVSLGST
	ggaggacaccatcatcagcgacgagtacagaaagaagctgaa	WLSDFPQAWAETGGMGLA
	ggaggctttcaaacactggggcaccagcaagggcaagtgcctg	VRQAPLIIPLKATSTPVSIKQ
	accgacgaggaagccagacagaccggcttctaccagagaaag	YPMSQEARLGIKPHIQRLLD
	aatggcaacttcatgagcctgcggtgcgtggagaagggcctgg	QGILVPCQSPWNTPLLPVKK
	gctggggccagatggtgaagatcaataacaacgcctttaagacg	PGTNDYRPVQDLREVNKRV
	aacgtgcacaacatcggggtggctgtttacctggacgagaagg	EDIHPTVPNPYNLLSGLPPSH
	gcaagaagaaggcctacgaaattaagaaccccagactgctgaa	QWYTVLDLKDAFFCLRLHP
	acatttcgtg**AGA**aagccacctgagatcaagggaaagaagct	TSQPLFAFEWRDPEMGISGQ
	gtacaccctgcggagaggcgactacgtgacccacgagagcga	LTWTRLPQGFKNSPTLFNEA
	ggactccgtgtacagaatcaaaaaactgggcacaagccctgtgc	LHRDLADFRIQHPDLILLQY
	tggaagctgtggtcaagggctctatcggccgcatccggaacagc	VDDLLLAATSELDCQQGTR
	gccagcgccacaaaactgaccaaggtgtccaagcctcagaagg	ALLQTLGNLGYRASAKKAQI
	gttttctggaggatctagcggaggatcctctggcagcgagacac	CQKQVKYLGYLLKEGQRWL
	caggaacaagcgagtcagcaacaccagagagcagtggcggc	TEARKETVMGQPTPKTPRQL
	agcagcggcggcagcagcaccctaaatatagaagatgagtatc	REFLGKAGFCRLFIPGFAEM
	ggctacatgagacctcaaaagagccagatgtttctctagggtcca	AAPLYPLTKPGTLFNWGPDQ
	catggctgtctgattttcctcaggcctgggcggaaaccgggggc	QKAYQEIKQALLTAPALGLP
	atgggactggcagttcgccaagctcctctgatcatacctctgaaa	DLTKPFELFVDEKQGYAKG
	gcaacctctacccccgtgtccataaaacaataccccatgtcacaa	VLTQKLGPWRRPVAYLSKK
	gaagccagactggggatcaagccccacatacagagactgttgg	LDPVAAGWPPCLRMVAAIA
	accagggaatactggtaccctgccagtccccctggaacacgcc	VLTKDAGKLTMGQPLVILAP
	cctgctacccgttaagaaaccagggactaatgattataggcctgt	HAVEALVKQPPDRWLSNAR
	ccaggatctgagagaagtcaacaagcgggtggaagacatccac	MTHYQALLLDTDRVQFGPV
	cccaccgtgcccaaccccttacaacctcttgagcgggctcccacc	VALNPATLLPLPEEGLQHNC
	gtcccaccagtggtacactgtgcttgatttaaaggatgccttttct	LDILAEAHGTRPDLTDQPLP
	gcctgagactccaccccaccagtcagcctctcttcgcctttgagt	DADHTWYTDGSSLLQEGQR
	ggagagatccagagatgggaatctcaggacaattgacctggac	KAGAAVTTETEVIWAKALP
	cagactcccacagggtttcaaaaacagtcccaccctgtttaatga	AGTSAQRAELIALTQALKM
	ggcactgcacagagacctagcagacttccggatccagcaccca	AEGKKLNVYTDSRYAFATA
	gacttgatcctgctacagtacgtggatgacttactgctggccgcc	HIHGEIYRRRGWLTSEGKEI
	acttctgagctagactgccaacaaggtactcgggccctgttacaa	KNKDEILALLKALFLPKRLSI
	accctagggaacctcgggtatcgggcctcggccaagaaagccc	IHCPGHQKGHSAEARGNRM
	aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga	ADQAARKAAITETPDTSTLLI

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagcccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gcccttt gaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 71)	ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 72)
I67R, D568R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggcaga agaagaacaggccagacaaagaagaagagactgagagaactg aaaacctacctgaagaacatcggaggaaaatatgacgagtctat cggcacattcaaaaacgagcctttcatgacagtgtacaagctgtg tcacaagagaggctacgactacgtggaaatgcctaaggaaaag accagagaagaaatcgaggccatgtctgccgacgagaaaaag cagtgggaaaaagagaaaaaggagtgggaggaaaccgagcg gaactccagacaccgggatgaggtgaccaaggacgtgcggaa gatcatgagcgagagagatgccaccgagcagcagatcagcca cgtgatcaagatcttcaacaagcagtaccggcccaagcggttca ataatagaatcctgacaaaatgtaaggttgaggattgcggcgtga acacacctctgagaaagaacgtcagagacctgctgatcaagga catcgtgagattcctgcccctgacccagagcgagaaagccgag ctggaagctcatgtgctggataaggaaggaagagagaaggtga agaccttcttcagacggcacagaatcaacgagttcctgcgcaag cagatctacgacatcgcctttaaccaaaagctgagcggcagaac cgtgttctgcaagaagcacatcctggagcacgagcaacacacc aaagaggaacgcaaggtctttagactggcccctagcctgaaaac taagattaacaacgtgctggctgtggttaaggacgagatcctgcc	MKRTADGSEFESPKKKRKV RNCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADIL EGRREQRG**R**RRTGQTKKKR LRELKTYLKNIGGKYDESIG TFKNEPFMTVYKLCHKRGY DYVEMPKEKTREEIEAMSA DEKKQWEKEKKEWEETERN SRHRDEVTKDVRKIMSERD ATEQQISHVIKIFNKQYRPKR FNNRILTKCKVEDCGVNTPL RKNVRDLLIKDIVRFLPLTQS EKAELEAHVLDKEGREKVK TFFRRHRINEFLRKQIYDIAF NQKLSGRTVFCKKHILEHEQ HTKEERKVFRLAPSLKTKIN NVLAVVKDEILPQFGINKVI MESNNFDIAARTKGKKRLA KDEYGKGHKEAKESLIEVLL RETKNQCVYCGKVITVTTA NTD**A**IFPRKAGGINVFANLV ACCRECNEKKKGRTPLESGV LPQPEITAHLENDLKKKILED

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ccagttcggcatcaacaaggtcatcatggaatccaacaacttcga	ARNMNELNLNSIKYMSYASI
	tattgccgctagaaccaagggcaagaagcggctggccaaggat	GWRHMRDRLRGLIGNEELPI
	gagtacggcaagggccacaaggaggccaaagagagcctgatc	ERQSGILTAYFRRWWGFRK
	gaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgcg	ERGNDIHHALDAVILASRKD
	gaaaggtgatcaccgtgaccaccgccaataccgac**GCC**atct	YTDDGVDMTLKPKSSNGRE
	tccccagaaaggcgggcggcatcaacgttttcgccaacctggtg	FDPEKHLPDTISFKR**R**KSNK
	gcctgttgtagagagtgcaacgagaagaagaagggcagaaca	GTPLYDKNPLSIKNSLITRRW
	cctctggaatccggcgtgcttccacagcctgagatcaccgccca	MVTEIEHGKEDTIISDEYRKK
	cctggaaaacgacctcaaaaagaagatccttgaggacgccaga	LKEAFKHWGTSKGKCLTDE
	aatatgaacgagctgaatctgaacagcatcaagtacatgagctac	EARQTGFYQRKNGNFMSLR
	gccagtatcggatggcggcacatgagagatagactgaggggcc	CVEKGLGWGQMVKINNNAF
	ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc	KTNVHNIGVAVYLDEKGKK
	ctgacagcttacttcagacggtggtggggatttcggaaggaacg	KAYEIKNPRLLKHFV**R**KPPEI
	gggaaatgacattcaccacgccctggacgctgtgatcctggcca	KGKKLYTLRRGDYVTHESE
	gcagaaaagactatacagatgacggcgtggatatgaccctgaa	DSVYRIKKLGTSPVLEAVVK
	gcccaaaagcagcaacggccgagaattcgaccccgagaaaca	GSIGRIRNSASATKLTKVSKP
	tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa	QKVFSGGSSGGSSGSETPGT
	ggaacccctctgtatgataagaacccattgtctatcaagaattctct	SESATPESSGGSSGGSSTLNI
	gatcacaagaaggtggatggtgacagagatcgaacacggcaa	EDEYRLHETSKEPDVSLGST
	ggaggacaccatcatcagcgacgagtacagaaagaagctgaa	WLSDFPQAWAETGGMGLA
	ggaggctttcaaacactggggcaccagcaagggcaagtgcctg	VRQAPLIIPLKATSTPVSIKQ
	accgacgaggaagccagacagaccggcttctaccagagaaag	YPMSQEARLGIKPHIQRLLD
	aatggcaacttcatgagcctgcggtgcgtggagaagggcctgg	QGILVPCQSPWNTPLLPVKK
	gctggggccagatggtgaagatcaataacaacgcctttaagacg	PGTNDYRPVQDLREVNKRV
	aacgtgcacaacatcggggtggctgtttacctggacgagaagg	EDIHPTVPNPYNLLSGLPPSH
	gcaagaagaaggcctacgaaattaagaaccccagactgctgaa	QWYTVLDLKDAFFCLRLHP
	acatttcgtg**AGA**aagccacctgagatcaagggaaagaagct	TSQPLFAFEWRDPEMGISGQ
	gtacaccctgcggagaggcgactacgtgacccacgagagcga	LTWTRLPQGFKNSPTLFNEA
	ggactccgtgtacagaatcaaaaaactgggcacaagccctgtgc	LHRDLADFRIQHPDLILLQY
	tggaagctgtggtcaagggctctatcggccgcatccggaacagc	VDDLLLAATSELDCQQGTR
	gccagcgccacaaaactgaccaaggtgtccaagcctcagaagg	ALLQTLGNLGYRASAKKAQI
	tgttttctggaggatctagcggaggatcctctggcagcgagacac	CQKQVKYLGYLLKEGQRWL
	caggaacaagcgagtcagcaacaccagagagcagtggcggc	TEARKETVMGQPTPKTPRQL
	agcagcggcggcagcagcaccctaaatatagaagatgagtatc	REFLGKAGFCRLFIPGFAEM
	ggctacatgagacctcaaaagagccagatgtttctctagggtcca	AAPLYPLTKPGTLFNWGPDQ
	catggctgtctgattttcctcaggcctgggcggaaaccgggggc	QKAYQEIKQALLTAPALGLP
	atgggactggcagttcgccaagctcctctgatcatacctctgaaa	DLTKPFELFVDEKQGYAKG
	gcaacctctacccccgtgtccataaaacaataccccatgtcacaa	VLTQKLGPWRRPVAYLSKK
	gaagccagactggggatcaagccccacatacagagactgttgg	LDPVAAGWPPCLRMVAAIA
	accagggaatactggtaccctgccagtccccctggaacacgcc	VLTKDAGKLTMGQPLVILAP
	cctgctacccgttaagaaaccagggactaatgattataggcctgt	HAVEALVKQPPDRWLSNAR
	ccaggatctgagagaagtcaacaagcgggtggaagacatccac	MTHYQALLLDTDRVQFGPV
	cccaccgtgcccaaccccttacaacctcttgagcgggctcccacc	VALNPATLLPLPEEGLQHNC
	gtcccaccagtggtacactgtgcttgatttaaaggatgccttttct	LDILAEAHGTRPDLTDQPLP
	gcctgagactccaccccaccagtcagcctctcttcgcctttgagt	DADHTWYTDGSSLLQEGQR
	ggagagatccagagatgggaatctcaggacaattgacctggac	KAGAAVTTETEVIWAKALP
	cagactcccacagggtttcaaaaacagtcccaccctgtttaatga	AGTSAQRAELIALTQALKM
	ggcactgcacagagacctagcagacttccggatccagcaccca	AEGKKLNVYTDSRYAFATA
	gacttgatcctgctacagtacgtggatgacttactgctggccgcc	HIHGEIYRRRGWLTSEGKEI
	acttctgagctagactgccaacaaggtactcgggccctgttacaa	KNKDEILALLKALFLPKRLSI
	accctagggaacctcgggtatcgggcctcggccaagaaagccc	IHCPGHQKGHSAEARGNRM
	aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga	ADQAARKAAITETPDTSTLLI
	gggtcagagatggctgactgaggccagaaaagagactgtgatg	

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagcccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gcccttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 73)	ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 74)
Truncated (16-775) I67R, D568R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggcag aagaagaacaggccagacaaagaagaagagactgagagaact gaaaacctacctgaagaacatcggaggaaaatatgacgagtcta tcggcacattcaaaaacgagcctttcatgacagtgtacaagctgt gtcacaagagaggctacgactacgtggaaatgcctaaggaaaa gaccagagaagaaatcgaggccatgtctgccgacgagaaaaa gcagtgggaaaaagagaaaaaggagtgggaggaaaccgagc ggaactccagacaccgggatgaggtgaccaaggacgtgcgga agatcatgagcgagagagatgccaccgagcagcagatcagcc acgtgatcaagatcttcaacaagcagtaccggcccaagcggttc aataatagaatcctgacaaaatgtaaggttgaggattgcggcgtg aacacacctctgagaaagaacgtcagagacctgctgatcaagg acatcgtgagattcctgcccctgacccagagcgagaaagccga gctggaagctcatgtgctggataaggaaggaagagagaaggtg aagaccttcttcagacggcacagaatcaacgagttcctgcgcaa gcagatctacgacatcgcctttaaccaaaagctgagcggcagaa ccgtgttctgcaagaagcacatcctggagcacgagcaacacac caaagaggaacgcaaggtctttagactggcccctagcctgaaaa ctaagattaacaacgtgctggctgtggttaaggacgagatcctgc cccagttcggcatcaacaaggtcatcatggaatccaacaacttcg atattgccgctagaaccaagggcaagaagcggctggccaagga	MKRTADGSEFESPKKKRKV QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKLC HKRGYDYVEMPKEKTREEIE AMSADEKKQWEKEKKEWE ETERNSRHRDEVTKDVRKIM SERDATEQQISHVIKIFNKQY RPKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFLP LTQSEKAELEAHVLDKEGRE KVKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHIL EHEQHTKEERKVFRLAPSLK TKINNVLAVVKDEILPQFGIN KVIMESNNFDIAARTKGKKR LAKDEYGKGHKEAKESLIEV LLRETKNQCVYCGKVITVTT ANTD**A**IFPRKAGGINVFANL VACCRECNEKKKGRTPLESG VLPQPEITAHLENDLKKKILE DARNMNELNLNSIKYMSYA SIGWRHMRDRLRGLIGNEEL

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	tgagtacggcaagggccacaaggaggccaaagagagcctgat cgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgc ggaaaggtgatcaccgtgaccaccgccaataccgac**GCC**at cttccccagaaaggcgggcggcatcaacgttttcgccaacctgg tggcctgttgtagagagtgcaacgagaagaagaagggcagaac acctctggaatccggcgtgcttccacagcctgagatcaccgccc acctggaaaacgacctcaaaaagaagatccttgaggacgccag aaatatgaacgagctgaatctgaacagcatcaagtacatgagcta cgccagtatcggatggcggcacatgagagatagactgaggggc cttatcggtaatgaagaactgcctatcgagcgccaaagcggcat cctgacagcttacttcagacggtggtggggatttcggaaggaac ggggaaatgacattcaccacgccctggacgctgtgatcctggcc agcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa ggaacccctctgtatgataagaacccattgtctatcaagaattctct gatcacaagaaggtggatggtgacagagatcgaacacggcaa ggaggacaccatcatcagcgacgagtacagaaagaagctgaa ggaggctttcaaacactggggcaccagcaagggcaagtgcctg accgacgaggaagccagacagaccggcttctaccagagaaag aatggcaacttcatgagcctgcggtgcgtggagaagggcctgg gctggggccagatggtgaagatcaataacaacgcctttaagacg aacgtgcacaacatcggggtggctgtttacctggacgagaagg gcaagaagaaggcctacgaaattaagaaccccagactgctgaa acatttcgtg**AGA**aagccacctgagatcaagggaaagaagct gtacaccctgcggagaggcgactacgtgacccacgagagcga ggactccgtgtacagaatcaaaaaactgggcacaagccctgtgc tggaagctgtggtcaagggctctatcggccgcatccggaacagc gccagcgccacaaaactgaccaaggtgtccaagcctcagaagg tgttttctggaggatctagcggaggatcctctggcagcgagacac caggaacaagcgagtcagcaacaccagagagcagtggcggc agcagcggcggcagcagcaccctaaatatagaagatgagtatc ggctacatgagacctcaaaagagccagatgtttctctagggtcca catggctgtctgattttcctcaggcctgggcggaaaccgggggc atgggactggcagttcgccaagctcctctgatcatacctctgaaa gcaacctctacccccgtgtccataaaacaataccccatgtcacaa gaagccagactggggatcaagccccacatacagagactgttgg accagggaatactggtaccctgccagtccccctggaacacgcc cctgctacccgttaagaaaccagggactaatgattataggcctgt ccaggatctgagagaagtcaacaagcgggtggaagacatccac cccaccgtgcccaaccccttacaacctcttgagcgggctcccacc gtcccaccagtggtacactgtgcttgatttaaaggatgccttttct gcctgagactccaccccaccagtcagcctctcttcgcctttgagt ggagagatccagagatgggaatctcaggacaattgacctggac cagactcccacagggtttcaaaaacagtcccaccctgtttaatga ggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgttacaa accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga	PIERQSGILTAYFRRWWGFR KERGNDIHHALDAVILASRK DYTDDGVDMTLKPKSSNGR EFDPEKHLPDTISFKR**R**KSNK GTPLYDKNPLSIKNSLITRRW MVTEIEHGKEDTIISDEYRKK LKEAFKHWGTSKGKCLTDE EARQTGFYQRKNGNFMSLR CVEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGKK KAYEIKNPRLLKHFV**R**KPPEI KGKKLYTLRRGDYVTHESE DSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKP QKVFSGGSSGGSSGSETPGT SESATPESSGGSSGGSSTLNI EDEYRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRV EDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHP TSQPLFAFEWRDPEMGISGQ LTWTRLPQGFKNSPTLFNEA LHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWL TEARKETVMGQPTPKTPRQL REFLGKAGFCRLFIPGFAEM AAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNC LDILAEAHGTRPDLTDQPLP DADHTWYTDGSSLLQEGQR KAGAAVTTETEVIWAKALP AGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATA HIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLI ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 76)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aatggcagccccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gccctttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggccccc atgcagtagaggcactagtcaaacaacccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 75)	
I67R, D568R, W593R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggcaga agaagaacaggccagacaaagaagaagagactgagagaactg aaaacctacctgaagaacatcggaggaaaatatgacgagtctat cggcacattcaaaaacgagcctttcatgacagtgtacaagctgtg tcacaagagaggctacgactacgtggaaatgcctaaggaaaag accagagaagaaatcgaggccatgtctgccgacgagaaaaag cagtgggaaaaagagaaaaaggagtgggaggaaaccgagcg gaactccagacaccgggatgaggtgaccaaggacgtgcggaa gatcatgagcgagagagatgccaccgagcagcagatcagcca cgtgatcaagatcttcaacaagcagtaccggcccaagcggttca ataatagaatcctgacaaaatgtaaggttgaggattgcggcgtga acacacctctgagaaagaacgtcagagacctgctgatcaagga catcgtgagattcctgcccctgacccagagcgagaaagccgag ctggaagctcatgtgctggataaggaaggaagagagaaggtga agaccttcttcagacggcacagaatcaacgagttcctgcgcaag cagatctacgacatcgcctttaaccaaaagctgagcggcagaac cgtgttctgcaagaagcacatcctggagcacgagcaacacacc aaagaggaacgcaaggtctttagactggccctagcctgaaaac taagattaacaacgtgctggctgtggttaaggacgagatcctgcc ccagttcggcatcaacaaggtcatcatggaatccaacaacttcga tattgccgctagaaccaagggcaagaagcggctggccaaggat gagtacggcaagggccacaaggaggccaaagagagcctgatc	MKRTADGSEFESPKKKRKV RNCKKDLYVEVTPMQKELV LGIDYGGKYTGLAVVDRRN NQVLYARTVKMRDDVADIL EGRREQRG**R**RRTGQTKKKR LRELKTYLKNIGGKYDESIG TFKNEPFMTVYKLCHKRGY DYVEMPKEKTREEIEAMSA DEKKQWEKEKKEWEETERN SRHRDEVTKDVRKIMSERD ATEQQISHVIKIFNKQYRPKR FNNRILTKCKVEDCGVNTPL RKNVRDLLIKDIVRFLPLTQS EKAELEAHVLDKEGREKVK TFFRRHRINEFLRKQIYDIAF NQKLSGRTVFCKKHILEHEQ HTKEERKVFRLAPSLKTKIN NVLAVVKDEILPQFGINKVI MESNNFDIAARTKGKKRLA KDEYGKGHKEAKESLIEVLL RETKNQCVYCGKVITVTTA NTD**A**IFPRKAGGINVFANLV ACCRECNEKKKGRTPLESGV LPQPEITAHLENDLKKKILED ARNMNELNLNSIKYMSYASI GWRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRK

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgcg gaaaggtgatcaccgtgaccaccgccaataccgac**GCC**atct tccccagaaaggcgggcggcatcaacgttttcgccaacctggtg gcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgccca cctggaaaacgacctcaaaaagaagatccttgaggacgccaga aatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcc ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc ctgacagcttacttcagacggtggtggggatttcggaaggaacg gggaaatgacattcaccacgccctggacgctgtgatcctggcca gcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa ggaacccctctgtatgataagaacccattgtctatcaagaattctct gatcacaagaagg**CGG**atggtgacagagatcgaacacggc aaggaggacaccatcatcagcgacgagtacagaaagaagctg aaggaggctttcaaacactggggcaccagcaagggcaagtgcc tgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaaga cgaacgtgcacaacatcggggtggctgtttacctggacgagaag ggcaagaagaaggcctacgaaattaagaaccccagactgctga aacatttcgtg**AGA**aagccacctgagatcaagggaaagaag ctgtacaccctgcggagaggcgactacgtgacccacgagagcg aggactccgtgtacagaatcaaaaaactgggcacaagccctgtg ctggaagctgtggtcaagggctctatcggccgcatccggaacag cgccagcgccacaaaactgaccaaggtgtccaagcctcagaag gtgttttctggaggatctagcggaggatcctctggcagcgagaca ccaggaacaagcgagtcagcaacaccagagagcagtggcgg cagcagcggcggcagcagcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctagggtcc acatggctgtctgattttcctcaggcctgggcggaaaccggggg catgggactggcagttcgccaagctcctctgatcatacctctgaa agcaacctctacccccgtgtccataaaacaatacccccatgtcaca agaagccagactggggatcaagccccacatacagagactgttg gaccagggaatactggtaccctgccagtccccctggaacacgc ccctgctacccgttaagaaaccagggactaatgattataggcctg tccaggatctgagagaagtcaacaagcgggtggaagacatcca ccccaccgtgcccaaccttacaacctcttgagcgggctcccac cgtcccaccagtggtacactgtgcttgatttaaaggatgccttttct gcctgagactccaccccaccagtcagcctctcttcgcctttgagt ggagagatccagagatgggaatctcaggacaattgacctggac cagactcccacagggtttcaaaaacagtcccaccctgtttaatga ggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgttacaa accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagccccccctgtaccctctcaccaaaccggggactctgtt	ERGNDIHHALDAVILASRKD YTDDGVDMTLKPKSSNGRE FDPEKHLPDTISFKR**R**KSNK GTPLYDKNPLSIKNSLITRR**R** MVTEIEHGKEDTIISDEYRKK LKEAFKHWGTSKGKCLTDE EARQTGFYQRKNGNFMSLR CVEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGKK KAYEIKNPRLLKHFV**R**KPPEI KGKKLYTLRRGDYVTHESE DSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKP QKVFSGGSSGGSSGSETPGT SESATPESSGGSSGGSSTLNI EDEYRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRV EDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHP TSQPLFAFEWRDPEMGISGQ LTWTRLPQGFKNSPTLFNEA LHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWL TEARKETVMGQPTPKTPRQL REFLGKAGFCRLFIPGFAEM AAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNC LDILAEAHGTRPDLTDQPLP DADHTWYTDGSSLLQEGQR KAGAAVTTETEVIWAKALP AGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATA HIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLI ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 78)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gccctttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 77)	
Truncated (16-775) I67R, D568R, W593R, E706R CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggc**A** **GA**agaagaacaggccagacaaagaagaagagactgagaga actgaaaacctacctgaagaacatcggaggaaaatatgacgagt ctatcggcacattcaaaaacgagcctttcatgacagtgtacaagct gtgtcacaagagaggctacgactacgtggaaatgcctaaggaa aagaccagagaagaaatcgaggccatgtctgccgacgagaaa aagcagtgggaaaaagagaaaaaggagtgggaggaaaccga gcggaactccagacaccgggatgaggtgaccaaggacgtgcg gaagatcatgagcgagagagatgccaccgagcagcagatcag ccacgtgatcaagatcttcaacaagcagtaccggcccaagcggt tcaataatagaatcctgacaaaatgtaaggttgaggattgcggcg tgaacacacctctgagaaagaacgtcagagacctgctgatcaag gacatcgtgagattcctgcccctgacccagagcgagaaagccg agctggaagctcatgtgctggataaggaaggaagagagaaggt gaagaccttcttcagacggcacagaatcaacgagttcctgcgca agcagatctacgacatcgcctttaaccaaaagctgagcggcaga accgtgttctgcaagaagcacatcctggagcacgagcaacaca ccaaagaggaacgcaaggtctttagactggcccctagcctgaaa actaagattaacaacgtgctggctgtggttaaggacgagatcctg ccccagttcggcatcaacaaggtcatcatggaatccaacaacttc gatattgccgctagaaccaagggcaagaagcggctggccaag gatgagtacggcaagggccacaaggaggccaaagagagcctg atcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactg	MKRTADGSEFESPKKKRKV QKELVLGIDYGGKYTGLAV VDRRNNQVLYARTVKMRD DVADILEGRREQRG**R**RRTG QTKKKRLRELKTYLKNIGGK YDESIGTFKNEPFMTVYKLC HKRGYDYVEMPKEKTREEIE AMSADEKKQWEKEKKEWE ETERNSRHRDEVTKDVRKIM SERDATEQQISHVIKIFNKQY RPKRFNNRILTKCKVEDCGV NTPLRKNVRDLLIKDIVRFLP LTQSEKAELEAHVLDKEGRE KVKTFFRRHRINEFLRKQIY DIAFNQKLSGRTVFCKKHIL EHEQHTKEERKVFRLAPSLK TKINNVLAVVKDEILPQFGIN KVIMESNNFDIAARTKGKKR LAKDEYGKGHKEAKESLIEV LLRETKNQCVYCGKVITVTT ANTD**A**IFPRKAGGINVFANL VACCRECNEKKKGRTPLESG VLPQPEITAHLENDLKKKILE DARNMNELNLNSIKYMSYA SIGWRHMRDRLRGLIGNEEL PIERQSGILTAYFRRWWGFR KERGNDIHHALDAVILASRK

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	cggaaaggtgatcaccgtgaccaccgccaataccgac**GCC**a tcttccccagaaaggcgggcggcatcaacgttttcgccaacctg gtggcctgttgtagagagtgcaacgagaagaagaagggcagaa cacctctggaatccggcgtgcttccacagcctgagatcaccgcc cacctggaaaacgacctcaaaaagaagatccttgaggacgcca gaaatatgaacgagctgaatctgaacagcatcaagtacatgagct acgccagtatcggatggcggcacatgagagatagactgagggg ccttatcggtaatgaagaactgcctatcgagcgccaaagcggca tcctgacagcttacttcagacggtggtggggatttcggaaggaac ggggaaatgacattcaccacgccctggacgctgtgatcctggcc agcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa ggaacccctctgtatgataagaacccattgtctatcaagaattctct gatcacaagaagg**CGG**atggtgacagagatcgaacacggc aaggaggacaccatcatcagcgacgagtacagaaagaagctg aaggaggctttcaaacactggggcaccagcaagggcaagtgcc tgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaaga cgaacgtgcacaacatcggggtggctgtttacctggacgagaag ggcaagaagaaggcctacgaaattaagaaccccagactgctga aacatttcgtg**AGA**aagccacctgagatcaagggaaagaag ctgtacaccctgcggagaggcgactacgtgacccacgagagcg aggactccgtgtacagaatcaaaaaactgggcacaagccctgtg ctggaagctgtggtcaagggctctatcggccgcatccggaacag cgccagcgccacaaaactgaccaaggtgtccaagcctcagaag gtgttttctggaggatctagcggaggatcctctggcagcgagaca ccaggaacaagcgagtcagcaacaccagagagcagtggcgg cagcagcggcggcagcagcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctagggtcc acatggctgtctgattttcctcaggcctgggcggaaaccggggg catgggactggcagttcgccaagctcctctgatcatacctctgaa agcaacctctacccccgtgtccataaaacaataccccatgtcaca agaagccagactggggatcaagccccacatacagagactgttg gaccagggaatactggtaccctgccagtccccctggaacacgc ccctgctacccgttaagaaaccagggactaatgattataggcctg tccaggatctgagagaagtcaacaagcgggtggaagacatcca ccccaccgtgcccaaccccttacaacctcttgagcgggctcccac cgtcccaccagtggtacactgtgcttgatttaaaggatgccttttct gcctgagactccaccccaccagtcagcctctcttcgcctttgagt ggagagatccagagatgggaatctcaggacaattgacctggac cagactcccacagggtttcaaaaacagtcccaccctgtttaatga ggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgttacaa accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagccccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc	DYTDDGVDMTLKPKSSNGR EFDPEKHLPDTISFKR**R**KSNK GTPLYDKNPLSIKNSLITRR**R** MVTEIEHGKEDTIISDEYRKK LKEAFKHWGTSKGKCLTDE EARQTGFYQRKNGNFMSLR CVEKGLGWGQMVKINNNAF KTNVHNIGVAVYLDEKGKK KAYEIKNPRLLKHFV**R**KPPEI KGKKLYTLRRGDYVTHESE DSVYRIKKLGTSPVLEAVVK GSIGRIRNSASATKLTKVSKP QKVFSGGSSGGSSGSETPGT SESATPESSGGSSGGSSTLNI EDEYRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLA VRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLD QGILVPCQSPWNTPLLPVKK PGTNDYRPVQDLREVNKRV EDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHP TSQPLFAFEWRDPEMGISGQ LTWTRLPQGFKNSPTLFNEA LHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWL TEARKETVMGQPTPKTPRQL REFLGKAGFCRLFIPGFAEM AAPLYPLTKPGTLFNWGPDQ QKAYQEIKQALLTAPALGLP DLTKPFELFVDEKQGYAKG VLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIA VLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNC LDILAEAHGTRPDLTDQPLP DADHTWYTDGSSLLQEGQR KAGAAVTTETEVIWAKALP AGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATA HIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSI IHCPGHQKGHSAEARGNRM ADQAARKAAITETPDTSTLLI ENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 90)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	aagctcttctaactgccccagccctggggttgccagatttgactaa gccctttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 89)	

This Example describes how the variant CRISPR nuclease-reverse transcriptase fusion polypeptide and variant CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide constructs were cloned. The constructs of SEQ ID NOs: 70, 72, 74, 76, 78, and 90 were used in **Example 7** to install edits in human target genes. The construct of SEQ ID NO: 68 is used in **Example 10**.

Example 7: RNA-Templated Editing of Human Genes in HEK293T Cells Using CRISPR Nuclease-Reverse Transcriptase and CRISPR Nickase-Reverse Transcriptase Fusion Polypeptides

This Example shows genetic modification of human genes utilizing CRISPR nuclease-reverse transcriptase fusion polypeptides of SEQ ID NOs: 70 and 72 from **Table 16** and CRISPR nickase-reverse transcriptase fusion polypeptides of SEQ ID NOs: 74, 76, 78, and 90 from **Table 16** constructed in **Example 6**. Specifically, these 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 17** 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 31-nucleotides in length. Two different PBS sequences, varied in length, were tested: 8-nucleotides (editing templates 1) and 10-nucleotides (editing template 2). 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 targeting AAVS1-T3 and EMX1-T7 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 **Table 17** and **Table 18** 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 18**.

Table 17. Sequences of Editing Template RNA Components

Editing Template RNAs	Spacer	Scaffold	RTT	PBS	Linker	3' Extension
1	UUGUGAGAAUGGUGCGUCCU (SEQ ID NO: 108)	GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGCGAUAU (SEQ ID NO: 112)	GGCGGCCACGACCUGGUGAACACgauccuCG (SEQ ID NO: 116)	CACCAUUC	AAGAUCAU	CGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 120)
2	CUUUAGCAAGGGACUAUUCA (SEQ ID NO: 109)	GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGCG	AAAAGCUUCAUCCCUGCUUCAagggacAAUA (SEQ ID NO: 117)	GUCCCUUGCU (SEQ ID NO: 177)	CACAGCAU	CGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 120)

Editing Template RNAs	Spacer	Scaffold	RTT	PBS	Linker	3' Extension
		AUAU (SEQ ID NO: 112)				
3	UUGUGAGAAUGGUGCGUCCU (SEQ ID NO: 108)	GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUACGGUGAAGGGAUAUACGCUUGGGUUUGUCCAGCCUGAGCCUCUAUGCCAGAAAUGGCGCCUUCAUCGUGGGUUAGGACA (SEQ ID NO: 114)	GGCGGCCACGACCUGGUGAACACgauccuCG (SEQ ID NO: 116)	CACCAUUC	AAGAUCAU	CGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 120)
4	CUUUAGCAAGGGACUAUUCA (SEQ ID NO: 109)	GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUACGGUGAAGGGAUAUACGCUUGGGUUUGUCCAGCCUGAGCCUCUAUGCCAGAAAUGGCGCCUUCAUCGUGGGUUAGGACA (SEQ ID NO: 114)	AAAAGCUUCAUCCCUGCUUCAagggacAAUA (SEQ ID NO: 117)	GUCCCUUGCU (SEQ ID NO: 177)	CACAGCAU	CGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 120)

**6-nucleotide substitution in lowercase*

Table 18. Full-length Editing Template RNA Sequences

Editing Template RNA	Target	Sequence
1	AAVS1-T3	UUGUGAGAAUGGUGCGUCCU**GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGACGAUAUU**GGCGGCCACGACCUGGUGAACACgauccuCGCACCAUUCAAGAUCAUCGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 124)
2	EMX1-T7	CUUUAGCAAGGGACUAUUCA**GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGACGAUAUU**AAAAGCUUCAUCCCUGCUUCAagggacAAUAGUCCCUUGCUCACAGCAUCGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 125)
3	AAVS1-T3	UUGUGAGAAUGGUGCGUCCU**GUUACAGUUAAGGGAAACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGCGAUAU**GGCGGCCACGACCUGGUGAACACgauccuCGCA

		CCAUUCAAGAUCAUCGCGGUUCUAUCUAGUUACGCGUUAAACCAA CUAGAA (SEQ ID NO: 126)
4	EMX1-T7	CUUUAGCAAGGGACUAUUCA**GUUACAGUUAAGGGAAACCUUAAU UGUAAAACGCCUAUAUGGUAAAGUGAUGUACGUUUGGGUAUAU AUCGCCAGCCUGAACCUCUACGCCAGAAAUGGCAGCUUUAUCAU GGGUUAGGCGAUAU**AAAAGCUUCAUCCCUGCUUCAagggacAAUAG UCCCUUGCUCACAGCAUCGCGGUUCUAUCUAGUUACGCGUUAAAC CAACUAGAA (SEQ ID NO: 127)
5	AAVS1-T3	UUGUGAGAAUGGUGCGUCCU**GUUACAGUUAAGGGAAACCUUAAU UGUAAAACGCCUAUACGGUGAAGGGAUAUACGCUUGGGUUUGU CCAGCCUGAGCCUCUAUGCCAGAAAUGGCGCCUUCAUCGUGGG UUAGGACA**GGCGGCCACGACCUGGUGAACACgauccuCGCACCAUUC AAGAUCAUCGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAA (SEQ ID NO: 128)
6	EMX1-T7	CUUUAGCAAGGGACUAUUCA**GUUACAGUUAAGGGAAACCUUAAU UGUAAAACGCCUAUACGGUGAAGGGAUAUACGCUUGGGUUUGU CCAGCCUGAGCCUCUAUGCCAGAAAUGGCGCCUUCAUCGUGGG UUAGGACA**AAAAGCUUCAUCCCUGCUUCAagggacAAUAGUCCCUUG CUCACAGCAUCGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAG AA (SEQ ID NO: 129)

**Scaffold sequence is shown in bold and the 6-nucleotide substitution is shown in lowercase*

HEK293T cells were transfected with plasmids expressing the CRISPR nuclease polypeptides and the editing template RNAs provided herein following the methods provided in **Example 1** above. Gene editing efficiencies were determined by NGS following the methods also provided in **Example 1**. The NGS results were analyzed also following the methods provided in **Example 1** above.

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, 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 19** and **Table 20**. The data presented therein is the average of three technical replicates.

Table 19. Editing Efficiencies with Variant CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptides

CRISPR Nuclease-Reverse Transcriptase Variant	Editing Template RNA	Scaffold	Target	Average % Indel	Average % 6-nt substitution installation
I67R, D568R, E706R	5	Scaffold 2	AAVS1-T3	28.45	4.97

I67R, D568R, E706R	6	Scaffold 2	EMX1-T7	20.08	0.16
Truncated (16-775) I67R, D568R, E706R	5	Scaffold 2	AAVS1-T3	25.88	6.08
Truncated (16-775) I67R, D568R, E706R	6	Scaffold 2	EMX1-T7	20.24	0.17

Table 20. Editing Efficiencies with Variant CRISPR Nickase (H397A)-Reverse Transcriptase Fusion Polypeptides

CRISPR Nickase (H397A)-Reverse Transcriptase Variant	Editing Template RNA	Scaffold	Target	Average % Indel	Average % 6-nt substitution installation
I67R, D568R, E706R	1	Truncated Reference	AAVS1-T3	0.23	1.64
I67R, D568R, E706R	2	Truncated Reference	EMX1-T7	0.39	0.11
I67R, D568R, E706R	5	Scaffold 2	AAVS1-T3	0.29	5.46
I67R, D568R, E706R	6	Scaffold 2	EMX1-T7	0.36	0.12
Truncated (16-775) I67R, D568R, E706R	5	Scaffold 2	AAVS1-T3	0.35	6.50
Truncated (16-775) I67R, D568R, E706R	6	Scaffold 2	EMX1-T7	0.39	0.14
I67R, D568R, W593R, E706R	5	Scaffold 2	AAVS1-T3	0.32	6.08
I67R, D568R, W593R, E706R	6	Scaffold 2	EMX1-T7	0.39	0.19
Truncated (16-775) I67R, D568R, W593R, E706R	5	Scaffold 2	AAVS1-T3	0.42	9.76
Truncated (16-775) I67R, D568R, W593R, E706R	6	Scaffold 2	EMX1-T7	0.72	0.92
None	5	Scaffold 2	AAVS1-T3	0.31	0.04
None	6	Scaffold 2	EMX1-T7	0.35	0.00

As shown in **Table 19**, the variant CRISPR nuclease-reverse transcriptase fusion polypeptides introduced substitutions encoded by the tested editing template RNAs at AAVS1 and EMX1 target loci. At the AAVS1 target locus, the average percentage of NGS reads comprising indels ranged from 25.88% to 28.45%, while the average percentage of NGS reads comprising encoding edits ranged from 4.97% to 6.08% (**Table 19**). At the EMX1 target locus, the average percentage of NGS reads comprising indels ranged from 20.08% to 20.24%, while the average percentage of NGS reads comprising encoding edits ranged from 0.16% to 0.17% (**Table 19**). At both targets, the truncated (16-775) I67R, D568R, E706R variant CRISPR

nuclease reverse-transcriptase fusions polypeptide displayed both the highest encoded edit installation.

As shown in **Table 20**, the variant CRISPR nickase (H397A)-reverse transcriptase fusion polypeptides introduced substitutions encoded by the tested editing template RNAs at AAVS1 and EMX1 target loci. Across the variant CRISPR nickase-reverse transcriptase fusion polypeptides, the average percentage of NGS reads comprising indels ranged from 0.23% to 0.42%, whereas the average percentage of NGS reads comprising encoding edits ranged from 1.64% to 9.76% at the AAVS1 target locus and between 0.11% and 0.92% at the EMX1 target locus. Editing template RNA 5 demonstrated the highest encoded edit installation when paired with the Truncated (16-775) I67R, D568R, W593R, E706R variant CRISPR nickase (H397A)-reverse transcriptase fusion.

The background indel rate was determined by testing controls comprising editing template guide plasmids transfected in the absence of plasmids encoding the RNA-guided fusion protein. *See*, **Table 20**. The indel levels observed for each of the RNA-guide nickase (H397A)-reverse transcriptase fusion polypeptides were approximately within error of these background indel rates. Additionally, the percentage of reads observed with the 6-nucleotide substitution installed was 0.04% for the AAVS1 target locus and <0.01% for the EMX1 target locus. This confirmed that the H397A substitution converted the variant CRISPR nuclease fusion polypeptides to CRISPR nickase fusion polypeptides and demonstrated that 6-nucleotide substitutions observed in the presence of the editing template were above the expected background rate.

This Example thus shows that the variant CRISPR nuclease-reverse transcriptase fusion polypeptides and the variant CRISPR nickase (H397A)-reverse transcriptase fusion polypeptides described here incorporated substitutions encoded by editing template RNAs into human genes. In addition, the I67R, D568R, E706R, and H397A nickase variant demonstrated increased editing with the addition of the W593R substitution or a truncated (16-775) CRISPR nuclease. Combining the substitution with the truncation in the Truncated (16-775) I67R, D568R, W593R, E706R variant CRISPR nickase (H397A)-reverse transcriptase fusion (SEQ ID NO: 90) further improved edit installation as shown above.

Example 8: Design of Variant CRISPR Nuclease Polypeptides, CRISPR Nuclease-Reverse Transcriptase Polypeptide Fusions, and CRISPR Nickase-Reverse Transcriptase Polypeptide Fusions

In this example, substitutions and/or sequence deletions are installed into the CRISPR

nuclease described in **Example 1**, CRISPR nuclease-reverse transcriptase polypeptide fusions described in **Example 6**, and CRISPR nickase (H397A)-reverse transcriptase polypeptide fusions described in **Example 6**.

Table 21 shows exemplary CRISPR nuclease-reverse transcriptase fusion polypeptide nucleotide and amino acid sequences. **Table 22** shows exemplary CRISPR nickase (H397A)-reverse transcriptase polypeptide nucleotide and amino acid sequences. The nucleotide and amino acid sequences CRISPR nuclease or nickase (H397A) portion of the fusion polypeptides are shown in **Table 14**. The codons encoding substituted residues are capitalized, bold, and underlined in the nucleotide sequence, and the substituted residue is shown in bold and underlined in the amino acid sequence.

Table 21. CRISPR Nuclease-Reverse Transcriptase Fusion Polypeptides

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
I67R, D568R, W593R, E706R CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggc**AG** **A**agaagaacaggccagacaaagaagaagagactgagagaac tgaaaacctacctgaagaacatcggaggaaaatatgacgagtct atcggcacattcaaaaacgagcctttcatgacagtgtacaagctg tgtcacaagagaggctacgactacgtggaaatgcctaaggaaaa gaccagagaagaaatcgaggccatgtctgccgacgagaaaaa gcagtgggaaaaagagaaaaaggagtgggaggaaaccgagc ggaactccagacaccgggatgaggtgaccaaggacgtgcgga agatcatgagcgagagagatgccaccgagcagcagatcagcc acgtgatcaagatcttcaacaagcagtaccggcccaagcggttc aataatagaatcctgacaaaatgtaaggttgaggattgcggcgtg aacacacctctgagaaagaacgtcagagacctgctgatcaagg acatcgtgagattcctgcccctgacccagagcgagaaagccga gctggaagctcatgtgctggataaggaaggaagagagaaggtg aagaccttcttcagacggcacagaatcaacgagttcctgcgcaa gcagatctacgacatcgcctttaaccaaaagctgagcggcagaa ccgtgttctgcaagaagcacatcctggagcacgagcaacacac caaagaggaacgcaaggtctttagactggcccctagcctgaaaa ctaagattaacaacgtgctggctgtggttaaggacgagatcctgc cccagttcggcatcaacaaggtcatcatggaatccaacaacttcg atattgccgctagaaccaagggcaagaagcggctggccaagga tgagtacggcaagggccacaaggaggccaaagagagcctgat cgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgc ggaaaggtgatcaccgtgaccaccgccaataccgaccacatctt ccccagaaaggcgggcggcatcaacgttttcgccaacctggtg gcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgccca cctggaaaacgacctcaaaaagaagatccttgaggacgccaga aatatgaacgagctgaatctgaacagcatcaagtacatgagctac	MKRTADGSEFESPKKKRKVRN CKKDLYVEVTPMQKELVLGID YGGKYTGLAVVDRRNNQVLY ARTVKMRDDVADILEGRREQR G**R**RRTGQTKKKRLRELKTYLK NIGGKYDESIGTFKNEPFMTVY KLCHKRGYDYVEMPKEKTREE IEAMSADEKKQWEKEKKEWEE TERNSRHRDEVTKDVRKIMSER DATEQQISHVIKIFNKQYRPKRF NNRILTKCKVEDCGVNTPLRK NVRDLLIKDIVRFLPLTQSEKAE LEAHVLDKEGREKVKTFFRRH RINEFLRKQIYDIAFNQKLSGRT VFCKKHILEHEQHTKEERKVFR LAPSLKTKINNVLAVVKDEILP QFGINKVIMESNNFDIAARTKG KKRLAKDEYGKGHKEAKESLI EVLLRETKNQCVYCGKVITVTT ANTDHIFPRKAGGINVFANLVA CCRECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDARNM NELNLNSIKYMSYASIGWRHM RDRLRGLIGNEELPIERQSGILT AYFRRWWGFRKERGNDIHHAL DAVILASRKDYTDDGVDMTLK PKSSNGREFDPEKHLPDTISFKR **R**KSNKGTPLYDKNPLSIKNSLIT RR**R**MVTEIEHGKEDTIISDEYR KKLKEAFKHWGTSKGKCLTDE EARQTGFYQRKNGNFMSLRCV EKGLGWGQMVKINNNAFKTN VHNIGVAVYLDEKGKKKAYEI KNPRLLKHFV**R**KPPEIKGKKLY

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gccagtatcggatggcggcacatgagagatagactgaggggcc ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc ctgacagcttacttcagacggtggtggggatttcggaaggaacg gggaaatgacattcaccacgccctggacgctgtgatcctggcca gcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa ggaacccctctgtatgataagaacccattgtctatcaagaattctct gatcacaagaagg**CGG**atggtgacagagatcgaacacggc aaggaggacaccatcatcagcgacgagtacagaaagaagctg aaggaggctttcaaacactggggcaccagcaagggcaagtgcc tgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaaga cgaacgtgcacaacatcggggtggctgtttacctggacgagaag ggcaagaagaaggcctacgaaattaagaaccccagactgctga aacatttcgtg**AGA**aagccacctgagatcaagggaaagaag ctgtacaccctgcggagaggcgactacgtgacccacgagagcg aggactccgtgtacagaatcaaaaaactgggcacaagccctgtg ctggaagctgtggtcaagggctctatcggccgcatccggaacag cgccagcgccacaaaactgaccaaggtgtccaagcctcagaag gtgttttctggaggatctagcggaggatcctctggcagcgagaca ccaggaacaagcgagtcagcaacaccagagagcagtggcgg cagcagcggcggcagcagcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctagggtcc acatggctgtctgattttcctcaggcctgggcggaaaccggggg catgggactggcagttcgccaagctcctctgatcatacctctgaa agcaacctctacccccgtgtccataaaacaatacccccatgtcaca agaagccagactggggatcaagccccacatacagagactgttg gaccagggaatactggtaccctgccagtccccctggaacacgc ccctgctacccgttaagaaaccagggactaatgattataggcctg tccaggatctgagagaagtcaacaagcgggtggaagacatcca ccccaccgtgcccaaccttacaacctcttgagcgggctcccac cgtcccaccagtggtacactgtgcttgatttaaaggatgccttttct gcctgagactccaccccaccagtcagcctctcttcgcctttgagt ggagagatccagagatgggaatctcaggacaattgacctggac cagactcccacagggtttcaaaaacagtcccaccctgtttaatga ggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgttacaa accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagaccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagccccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gcccttttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc	TLRRGDYVTHESEDSVYRIKKL GTSPVLEAVVKGSIGRIRNSAS ATKLTKVSKPQKVFSGGSSGGS SGSETPGTSESATPESSGGSSGG SSTLNIEDEYRLHETSKEPDVSL GSTWLSDFPQAWAETGGMGL AVRQAPLIIPLKATSTPVSIKQY PMSQEARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKPGTND YRPVQDLREVNKRVEDIHPTVP NPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWR DPEMGISGQLTWTRLPQGFKNS PTLFNEALHRDLADFRIQHPDLI LLQYVDDLLLAATSELDCQQG TRALLQTLGNLGYRASAKKAQ ICQKQVKYLGYLLKEGQRWLT EARKETVMGQPTPKTPRQLREF LGKAGFCRLFIPGFAEMAAPLY PLTKPGTLFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPFELF VDEKQGYAKGVLTQKLGPWR RPVAYLSKKLDPVAAGWPPCL RMVAAIAVLTKDAGKLTMGQP LVILAPHAVEALVKQPPDRWLS NARMTHYQALLLDTDRVQFGP VVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDAD HTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQR AELIALTQALKMAEGKKLNVY TDSRYAFATAHIHGEIYRRRGW LTSEGKEIKNKDEILALLKALFL PKRLSIIHCPGHQKGHSAEARG NRMADQAARKAAITETPDTST LLIENSSPKRPAATKKAGQAKK KK (SEQ ID NO: 80)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	atgcagtagaggcactagtcaaacaacccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 79)	
Truncated (16-775) reference CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggcatt agaagaacaggccagacaaagaagaagagactgagagaactg aaaacctacctgaagaacatcggaggaaaatatgacgagtctat cggcacattcaaaaacgagcctttcatgacagtgtacaagctgtg tcacaagagaggctacgactacgtggaaatgcctaaggaaaag accagagaagaaatcgaggccatgtctgccgacgagaaaaag cagtgggaaaaagagaaaaaggagtgggaggaaaccgagcg gaactccagacaccgggatgaggtgaccaaggacgtgcggaa gatcatgagcgagagagatgccaccgagcagcagatcagcca cgtgatcaagatcttcaacaagcagtaccggcccaagcggttca ataatagaatcctgacaaaatgtaaggttgaggattgcggcgtga acacacctctgagaaagaacgtcagagacctgctgatcaagga catcgtgagattcctgcccctgacccagagcgagaaagccgag ctggaagctcatgtgctggataaggaaggaagagagaaggtga agaccttcttcagacggcacagaatcaacgagttcctgcgcaag cagatctacgacatcgcctttaaccaaaagctgagcggcagaac cgtgttctgcaagaagcacatcctggagcacgagcaacacacc aaagaggaacgcaaggtctttagactggccctagcctgaaaac taagattaacaacgtgctggctgtggttaaggacgagatcctgcc ccagttcggcatcaacaaggtcatcatggaatccaacaacttcga tattgccgctagaaccaagggcaagaagcggctggccaaggat gagtacggcaagggccacaaggaggccaaagagagcctgatc gaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgcg gaaaggtgatcaccgtgaccaccgccaataccgaccacatcttc cccagaaaggcgggcggcatcaacgttttcgccaacctggtgg cctgttgtagagagtgcaacgagaagaagaagggcagaacacc tctggaatccggcgtgcttccacagcctgagatcaccgcccacct ggaaaacgacctcaaaaagaagatccttgaggacgccagaaat atgaacgagctgaatctgaacagcatcaagtacatgagctacgc cagtatcggatggcggcacatgagagatagactgaggggcctt atcggtaatgaagaactgcctatcgagcgccaaagcggcatcct	MKRTADGSEFESPKKKRKVQK ELVLGIDYGGKYTGLAVVDRR NNQVLYARTVKMRDDVADILE GRREQRGIRRTGQTKKKRLREL KTYLKNIGGKYDESIGTFKNEP FMTVYKLCHKRGYDYVEMPK EKTREEIEAMSADEKKQWEKE KKEWEETERNSRHRDEVTKDV RKIMSERDATEQQISHVIKIFNK QYRPKRFNNRILTKCKVEDCG VNTPLRKNVRDLLIKDIVRFLPL TQSEKAELEAHVLDKEGREKV KTFFRRHRINEFLRKQIYDIAFN QKLSGRTVFCKKHILEHEQHTK EERKVFRLAPSLKTKINNVLAV VKDEILPQFGINKVIMESNNFDI AARTKGKKRLAKDEYGKGHK EAKESLIEVLLRETKNQCVYCG KVITVTTANTDHIFPRKAGGIN VFANLVACCRECNEKKKGRTP LESGVLPQPEITAHLENDLKKKI LEDARNMNELNLNSIKYMSYA SIGWRHMRDRLRGLIGNEELPI ERQSGILTAYFRRWWGFRKER GNDIHHALDAVILASRKDYTD DGVDMTLKPKSSNGREFDPEK HLPDTISFKRDKSNKGTPLYDK NPLSIKNSLITRRWMVTEIEHG KEDTIISDEYRKKLKEAFKHWG TSKGKCLTDEEARQTGFYQRK NGNFMSLRCVEKGLGWGQMV KINNNAFKTNVHNIGVAVYLD EKGKKKAYEIKNPRLLKHFVE KPPEIKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEAVVK

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gacagcttacttcagacggtggtggggatttcggaaggaacggg gaaatgacattcaccacgccctggacgctgtgatcctggccagc agaaaagactatacagatgacggcgtggatatgaccctgaagcc caaaagcagcaacggccgagaattcgaccccgagaaacatctg cctgatacaatcagcttcaaaagagataagtccaacaaaggaac ccctctgtatgataagaacccattgtctatcaagaattctctgatca caagaaggtggatggtgacagagatcgaacacggcaaggagg acaccatcatcagcgacgagtacagaaagaagctgaaggagg ctttcaaacactggggcaccagcaagggcaagtgcctgaccga cgaggaagccagacagaccggcttctaccagagaaagaatgg caacttcatgagcctgcggtgcgtggagaagggcctgggctgg ggccagatggtgaagatcaataacaacgcctttaagacgaacgt gcacaacatcggggtggctgtttacctggacgagaagggcaag aagaaggcctacgaaattaagaaccccagactgctgaaacattt cgtggaaaagccacctgagatcaagggaaagaagctgtacacc ctgcggagaggcgactacgtgacccacgagagcgaggactcc gtgtacagaatcaaaaaactgggcacaagccctgtgctggaagc tgtggtcaagggctctatcggccgcatccggaacagcgccagc gccacaaaactgaccaaggtgtccaagcctcagaaggtgttttct ggaggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagcag cggcggcagcagcaccctaaatatagaagatgagtatcggctac atgagacctcaaaagagccagatgtttctctagggtccacatggc tgtctgattttcctcaggcctgggcggaaaccgggggcatggga ctggcagttcgccaagctcctctgatcatacctctgaaagcaacct ctaccccgtgtccataaaacaataccccatgtcacaagaagcc agactggggatcaagccccacatacagagactgttggaccagg gaatactggtaccctgccagtcccctggaacacgcccctgcta cccgttaagaaaccagggactaatgattataggcctgtccaggat ctgagagaagtcaacaagcgggtggaagacatccaccccacc gtgcccaaccttacaacctcttgagcgggctcccaccgtccca ccagtggtacactgtgcttgatttaaaggatgccttttctgcctga gactccaccccaccagtcagcctctcttcgcctttgagtggagag atccagagatgggaatctcaggacaattgacctggaccagactc ccacagggtttcaaaaacagtcccaccctgtttaatgaggcactg cacagagacctagcagacttccggatccagcacccagacttgat cctgctacagtacgtggatgacttactgctggccgccacttctga gctagactgccaacaaggtactcgggccctgttacaaaccctag ggaacctcgggtatcgggcctcggccaagaaagcccaaatttg ccagaaacaggtcaagtatctggggtatcttctaaaagagggtca gagatggctgactgaggccagaaaagagactgtgatggggca gcctactccgaagacccctcgacaactaagggagttcctaggga aggcaggcttctgtcgcctcttcatccctgggtttgcagaaatggc agccccccctgtaccctctcaccaaaccggggactctgtttaattg gggcccagaccaacaaaaggcctatcaagaaatcaagcaagct cttctaactgccccagccctggggttgccagatttgactaagccct ttgaactctttgtcgacgagaagcagggctacgccaaaggtgtcc taacgcaaaaactgggaccttggcgtcggccggtggcctacctg tccaaaaagctagacccagtagcagctgggtggcccccttgcct acggatggtagcagccattgccgtactgacaaaggatgcaggc aagctaaccatgggacagccactagtcattctggcccccatgc agtagaggcactagtcaaacaaccccccgaccgctggctttcca acgcccggatgactcactatcaggccttgcttttggacacggacc	GSIGRIRNSASATKLTKVSKPQ KVFSGGSSGGSSGSETPGTSESA TPESSGGSSGGSSTLNIEDEYRL HETSKEPDVSLGSTWLSDFPQA WAETGGMGLAVRQAPLIIPLK ATSTPVSIKQYPMSQEARLGIKP HIQRLLDQGILVPCQSPWNTPL LPVKKPGTNDYRPVQDLREVN KRVEDIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLRLHPT SQPLFAFEWRDPEMGISGQLTW TRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLA ATSELDCQQGTRALLQTLGNL GYRASAKKAQICQKQVKYLGY LLKEGQRWLTEARKETVMGQP TPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFNW GPDQQKAYQEIKQALLTAPAL GLPDLTKPFELFVDEKQGYAK GVLTQKLGPWRRPVAYLSKKL DPVAAGWPPCLRMVAAIAVLT KDAGKLTMGQPLVILAPHAVE ALVKQPPDRWLSNARMTHYQ ALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHG TRPDLTDQPLPDADHTWYTDG SSLLQEGQRKAGAAVTTETEVI WAKALPAGTSAQRAELIALTQ ALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSIIH CPGHQKGHSAEARGNRMADQ AARKAAITETPDTSTLLIENSSP KRPAATKKAGQAKKKK (SEQ ID NO: 82)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gggtccagttcggaccggtggtagccctgaacccggctacgct gctcccactgcctgaggaagggctgcaacacaactgccttgata tcctggccgaagcccacggaacccgacccgacctaacggacc agccgctcccagacgccgaccacacctggtacacggatggaa gcagtctcttacaagagggacagcgtaaggcgggagctgcggt gaccaccgagaccgaggtaatctgggctaaagccctgccagcc gggacatccgctcagcgggctgaactgatagcactcacccagg ccctaaagatggcagaaggtaagaagctaaatgtttatactgata gccgttatgcttttgctactgcccatatccatggagaaatatacaga aggcgtgggtggctcacatcagaaggcaaagagatcaaaaata aagacgagatcttggccctactaaaagccctctttctgcccaaaa gacttagcataatccattgtccaggacatcaaaagggacacagc gccgaggctagaggcaaccggatggctgaccaagcggcccga aaggcagccatcacagagactccagacacctctaccctcctcat agaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 81)	
Truncated (16-775) I67R, D568R, W593R, E706R CRISPR nuclease-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggc**A** **GA**agaagaacaggccagacaaagaagaagagactgagaga actgaaaacctacctgaagaacatcggaggaaaatatgacgagt ctatcggcacattcaaaaacgagcctttcatgacagtgtacaagct gtgtcacaagagaggctacgactacgtggaaatgcctaaggaa aagaccagagaagaaatcgaggccatgtctgccgacgagaaa aagcagtgggaaaaagagaaaaaggagtgggaggaaaccga gcggaactccagacaccgggatgaggtgaccaaggacgtgcg gaagatcatgagcgagagagatgccaccgagcagcagatcag ccacgtgatcaagatcttcaacaagcagtaccggcccaagcggt tcaataatagaatcctgacaaaatgtaaggttgaggattgcggcg tgaacacacctctgagaaagaacgtcagagacctgctgatcaag gacatcgtgagattcctgcccctgacccagagcgagaaagccg agctggaagctcatgtgctggataaggaaggaagagagaaggt gaagaccttcttcagacggcacagaatcaacgagttcctgcgca agcagatctacgacatcgcctttaaccaaaagctgagcggcaga accgtgttctgcaagaagcacatcctggagcacgagcaacaca ccaaagaggaacgcaaggtctttagactggcccctagcctgaaa actaagattaacaacgtgctggctgtggttaaggacgagatcctg ccccagttcggcatcaacaaggtcatcatggaatccaacaacttc gatattgccgctagaaccaagggcaagaagcggctggccaag gatgagtacggcaagggccacaaggaggccaaagagagcctg atcgaggtgctgctgcgggaaaccaagaaccagtgcgtgtactg cggaaaggtgatcaccgtgaccaccgccaataccgaccacatct tccccagaaaggcgggcggcatcaacgttttcgccaacctggtg gcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgccca cctggaaaacgacctcaaaaagaagatccttgaggacgccaga aatatgaacgagctgaatctgaacagcatcaagtacatgagctac gccagtatcggatggcggcacatgagagatagactgaggggcc ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc ctgacagcttacttcagacggtggtggggatttcggaaggaacg gggaaatgacattcaccacgccctggacgctgtgatcctggcca	MKRTADGSEFESPKKKRKVQK ELVLGIDYGGKYTGLAVVDRR NNQVLYARTVKMRDDVADILE GRREQRG**R**RRTGQTKKKRLRE LKTYLKNIGGKYDESIGTFKNE PFMTVYKLCHKRGYDYVEMP KEKTREEIEAMSADEKKQWEK EKKEWEETERNSRHRDEVTKD VRKIMSERDATEQQISHVIKIFN KQYRPKRFNNRILTKCKVEDC GVNTPLRKNVRDLLIKDIVRFL PLTQSEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIYDIAF NQKLSGRTVFCKKHILEHEQHT KEERKVFRLAPSLKTKINNVLA VVKDEILPQFGINKVIMESNNF DIAARTKGKKRLAKDEYGKGH KEAKESLIEVLLRETKNQCVYC GKVITVTTANTDHIFPRKAGGI NVFANLVACCRECNEKKKGRT PLESGVLPQPEITAHLENDLKK KILEDARNMNELNLNSIKYMSY ASIGWRHMRDRLRGLIGNEELP IERQSGILTAYFRRWWGFRKER GNDIHHALDAVILASRKDYTD DGVDMTLKPKSSNGREFDPEK HLPDTISFKR**R**KSNKGTPLYDK NPLSIKNSLITRR**R**MVTEIEHGK EDTIISDEYRKKLKEAFKHWGT SKGKCLTDEEARQTGFYQRKN GNFMSLRCVEKGLGWGQMVK INNNAFKTNVHNIGVAVYLDE KGKKKAYEIKNPRLLKHFV**R**K PPEIKGKKLYTLRRGDYVTHES EDSVYRIKKLGTSPVLEAVVKG SIGRIRNSASATKLTKVSKPQK VFSGGSSGGSSGSETPGTSESAT

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaaga**AGA**aagtccaacaaa ggaacccctctgtatgataagaacccattgtctatcaagaattctct gatcacaagaagg**CGG**atggtgacagagatcgaacacggc aaggaggacaccatcatcagcgacgagtacagaaagaagctg aaggaggctttcaaacactggggcaccagcaagggcaagtgcc tgaccgacgaggaagccagacagaccggcttctaccagagaa agaatggcaacttcatgagcctgcggtgcgtggagaagggcct gggctggggccagatggtgaagatcaataacaacgcctttaaga cgaacgtgcacaacatcggggtggctgtttacctggacgagaag ggcaagaagaaggcctacgaaattaagaaccccagactgctga aacatttcgtg**AGA**aagccacctgagatcaagggaaagaag ctgtacaccctgcggagaggcgactacgtgacccacgagagcg aggactccgtgtacagaatcaaaaaactgggcacaagccctgtg ctggaagctgtggtcaagggctctatcggccgcatccggaacag cgccagcgccacaaaactgaccaaggtgtccaagcctcagaag gtgttttctggaggatctagcggaggatcctctggcagcgagaca ccaggaacaagcgagtcagcaacaccagagagcagtggcgg cagcagcggcggcagcagcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctagggtcc acatggctgtctgattttcctcaggcctgggcggaaaccggggg catgggactggcagttcgccaagctcctctgatcatacctctgaa agcaacctctacccccgtgtccataaaacaataccccatgtcaca agaagccagactggggatcaagccccacatacagagactgttg gaccagggaatactggtaccctgccagtccccctggaacacgc ccctgctacccgttaagaaaccagggactaatgattataggcctg tccaggatctgagagaagtcaacaagcgggtggaagacatcca ccccaccgtgcccaacccttacaacctcttgagcgggctcccac cgtcccaccagtggtacactgtgcttgatttaaaggatgccttttct gcctgagactccaccccaccagtcagcctctcttcgcctttgagt ggagagatccagagatgggaatctcaggacaattgacctggac cagactcccacagggtttcaaaaacagtcccaccctgtttaatga ggcactgcacagagacctagcagacttccggatccagcaccca gacttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgttacaa accctagggaacctcgggtatcgggcctcggccaagaaagccc aaatttgccagaaacaggtcaagtatctggggtatcttctaaaaga gggtcagagatggctgactgaggccagaaaagagactgtgatg gggcagcctactccgaagacccctcgacaactaagggagttcct agggaaggcaggcttctgtcgcctcttcatccctgggtttgcaga aatggcagccccccctgtaccctctcaccaaaccggggactctgtt taattggggcccagaccaacaaaaggcctatcaagaaatcaagc aagctcttctaactgccccagccctggggttgccagatttgactaa gccctttgaactctttgtcgacgagaagcagggctacgccaaag gtgtcctaacgcaaaaactgggaccttggcgtcggccggtggcc tacctgtccaaaaagctagacccagtagcagctgggtggccccc ttgcctacggatggtagcagccattgccgtactgacaaaggatgc aggcaagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaaccccccgaccgctggctt tccaacgcccggatgactcactatcaggccttgcttttggacacg gaccgggtccagttcggaccggtggtagccctgaacccggcta cgctgctcccactgcctgaggaagggctgcaacacaactgcctt	PESSGGSSGGSSTLNIEDEYRLH ETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKAT STPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQ WYTVLDLKDAFFCLRLHPTSQP LFAFEWRDPEMGISGQLTWTR LPQGFKNSPTLFNEALHRDLAD FRIQHPDLILLQYVDDLLLAATS ELDCQQGTRALLQTLGNLGYR ASAKKAQICQKQVKYLGYLLK EGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFA EMAAPLYPLTKPGTLFNWGPD QQKAYQEIKQALLTAPALGLPD LTKPFELFVDEKQGYAKGVLT QKLGPWRRPVAYLSKKLDPVA AGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVK QPPDRWLSNARMTHYQALLLD TDRVQFGPVVALNPATLLPLPE EGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQE GQRKAGAAVTTETEVIWAKAL PAGTSAQRAELIALTQALKMAE GKKLNVYTDSRYAFATAHIHG EIYRRRGWLTSEGKEIKNKDEI LALLKALFLPKRLSIIHCPGHQK GHSAEARGNRMADQAARKAAI TETPDTSTLLIENSSPKRPAATK KAGQAKKKK (SEQ ID NO: 84)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gatatcctggccgaagcccacggaacccgacccgacctaacgg accagccgctcccagacgccgaccacacctggtacacggatgg aagcagtctcttacaagagggacagcgtaaggcgggagctgcg gtgaccaccgagaccgaggtaatctgggctaaagccctgccag ccgggacatccgctcagcgggctgaactgatagcactcaccca ggccctaaagatggcagaaggtaagaagctaaatgtttatactga tagccgttatgcttttgctactgcccatatccatggagaaatataca gaaggcgtgggtggctcacatcagaaggcaaagagatcaaaaa taaagacgagatcttggccctactaaaagccctctttctgcccaaa agacttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggcccg aaaggcagccatcacagagactccagacacctctaccctcctca tagaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 83)	

Table 22. CRISPR Nickase (H397A)-Reverse Transcriptase Fusion Polypeptides

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
Reference CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccggaactgcaagaaggacctgtacgtg gaagtgacccctatgcagaaagagctggtgctgggcattgacta cggcggaaagtacacaggcctggccgtggtggaccggagaaa caaccaggtgctgtacgccagaaccgtcaagatgagagacgac gtggccgatatcctcgaaggccggagggaacagcgaggcatta gaagaacaggccagacaaagaagaagagactgagagaactga aaacctacctgaagaacatcggaggaaaatatgacgagtctatc ggcacattcaaaaacgagcctttcatgacagtgtacaagctgtgt cacaagagaggctacgactacgtggaaatgcctaaggaaaaga ccagagaagaaatcgaggccatgtctgccgacgagaaaaagc agtgggaaaaagagaaaaaggagtgggaggaaaccgagcgg aactccagacaccgggatgaggtgaccaaggacgtgcggaag atcatgagcgagagagatgccaccgagcagcagatcagccac gtgatcaagatcttcaacaagcagtaccggcccaagcggttcaat aatagaatcctgacaaaatgtaaggttgaggattgcggcgtgaac acacctctgagaaagaacgtcagagacctgctgatcaaggacat cgtgagattcctgcccctgacccagagcgagaaagccgagctg gaagctcatgtgctggataaggaaggaagagagaaggtgaaga ccttcttcagacggcacagaatcaacgagttcctgcgcaagcag atctacgacatcgcctttaaccaaaagctgagcggcagaaccgt gttctgcaagaagcacatcctggagcacgagcaacacaccaaa gaggaacgcaaggtctttagactggcccctagcctgaaaactaa gattaacaacgtgctggctgtggttaaggacgagatcctgcccca gttcggcatcaacaaggtcatcatggaatccaacaacttcgatatt gccgctagaaccaagggcaagaagcggctggccaaggatgag tacggcaagggccacaaggaggccaaagagagcctgatcgag gtgctgctgcgggaaaccaagaaccagtgcgtgtactgcggaa aggtgatcaccgtgaccaccgccaataccgac**GCC**atcttcc ccagaaaggcgggcggcatcaacgttttcgccaacctggtggc ctgttgtagagagtgcaacgagaagaagaagggcagaacacct ctggaatccggcgtgcttccacagcctgagatcaccgcccacct	MKRTADGSEFESPKKKRKVR NCKKDLYVEVTPMQKELVLGI DYGGKYTGLAVVDRRNNQVL YARTVKMRDDVADILEGRRE QRGIRRTGQTKKKRLRELKTY LKNIGGKYDESIGTFKNEPFMT VYKLCHKRGYDYVEMPKEKT REEIEAMSADEKKQWEKEKK EWEETERNSRHRDEVTKDVR KIMSERDATEQQISHVIKIFNK QYRPKRFNNRILTKCKVEDCG VNTPLRKNVRDLLIKDIVRFLP LTQSEKAELEAHVLDKEGREK VKTFFRRHRINEFLRKQIYDIA FNQKLSGRTVFCKKHILEHEQ HTKEERKVFRLAPSLKTKINN VLAVVKDEILPQFGINKVIMES NNFDIAARTKGKKRLAKDEY GKGHKEAKESLIEVLLRETKN QCVYCGKVITVTTANTD**A**IFP RKAGGINVFANLVACCRECNE KKKGRTPLESGVLPQPEITAHL ENDLKKKILEDARNMNELNLN SIKYMSYASIGWRHMRDRLRG LIGNEELPIERQSGILTAYFRR WWGFRKERGNDIHHALDAVI LASRKDYTDDGVDMTLKPKS SNGREFDPEKHLPDTISFKRDK SNKGTPLYDKNPLSIKNSLITR RWMVTEIEHGKEDTIISDEYR KKLKEAFKHWGTSKGKCLTD EEARQTGFYQRKNGNFMSLR

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	ggaaaacgacctcaaaaagaagatccttgaggacgccagaaat atgaacgagctgaatctgaacagcatcaagtacatgagctacgc cagtatcggatggcggcacatgagagatagactgaggggccttt atcggtaatgaagaactgcctatcgagcgccaaagcggcatcct gacagcttacttcagacggtggtggggatttcggaaggaacgggg gaaatgacattcaccacgccctggacgctgtgatcctggccagc agaaaagactatacagatgacggcgtggatatgaccctgaagcc caaaagcagcaacggccgagaattcgaccccgagaaacatctg cctgatacaatcagcttcaaaagagataagtccaacaaaggaac ccctctgtatgataagaacccattgtctatcaagaattctctgatca caagaaggtggatggtgacagagatcgaacacggcaaggagg acaccatcatcagcgacgagtacagaaagaagctgaaggagg ctttcaaacactggggcaccagcaagggcaagtgcctgaccga cgaggaagccagacagaccggcttctaccagagaaagaatgg caacttcatgagcctgcggtgcgtggagaagggcctgggctgg ggccagatggtgaagatcaataacaacgcctttaagacgaacgt gcacaacatcggggtggctgtttacctggacgagaagggcaag aagaaggcctacgaaattaagaaccccagactgctgaaacattt cgtggaaaagccacctgagatcaagggaaagaagctgtacacc ctgcggagaggcgactacgtgacccacgagagcgaggactcc gtgtacagaatcaaaaaactgggcacaagccctgtgctggaagc tgtggtcaagggctctatcggccgcatccggaacagcgccagc gccacaaaactgaccaaggtgtccaagcctcagaaggtgttttct ggaggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagcag cggcggcagcagcaccctaaatatagaagatgagtatcggctac atgagacctcaaaagagccagatgtttctctagggtccacatggc tgtctgattttcctcaggcctgggcggaaaccgggggcatggga ctggcagttcgccaagctcctctgatcatacctctgaaagcaacct ctacccccgtgtccataaaacaataccccatgtcacaagaagcc agactggggatcaagccccacatacagagactgttggaccagg gaatactggtaccctgccagtccccctggaacacgcccctgcta cccgttaagaaaccagggactaatgattataggcctgtccaggat ctgagagaagtcaacaagcgggtggaagacatccaccccacc gtgcccaacccttacaacctcttgagcgggctcccaccgtccca ccagtggtacactgtgcttgatttaaaggatgccttttttctgcctga gactccaccccaccagtcagcctctcttcgcctttgagtggagag atccagagatgggaatctcaggacaattgacctggaccagactc ccacagggtttcaaaaacagtcccaccctgtttaatgaggcactg cacagagacctagcagacttccggatccagcacccagacttgat cctgctacagtacgtggatgacttactgctggccgccacttctga gctagactgccaacaaggtactcgggccctgttacaaaccctag ggaacctcgggtatcgggcctcggccaagaaagcccaaatttg ccagaaacaggtcaagtatctggggtatcttctaaaagagggtca gagatggctgactgaggccagaaaagagactgtgatggggca gcctactccgaagacccctcgacaactaagggagttcctaggga aggcaggcttctgtcgcctcttcatccctgggtttgcagaaatggc agcccccctgtaccctctcaccaaaccggggactctgtttaattg gggcccagaccaacaaaaggcctatcaagaaatcaagcaagct cttctaactgccccagccctggggttgccagatttgactaagccct ttgaactctttgtcgacgagaagcagggctacgccaaaggtgtcc taacgcaaaaactgggaccttggcgtcggccggtggcctacctg tccaaaaagctagacccagtagcagctgggtggcccccttgcct	CVEKGLGWGQMVKINNNAFK TNVHNIGVAVYLDEKGKKKA YEIKNPRLLKHFVEKPPEIKGK KLYTLRRGDYVTHESEDSVYR IKKLGTSPVLEAVVKGSIGRIR NSASATKLTKVSKPQKVFSGG SSGGSSGSETPGTSESATPESSG GSSGGSSTLNIEDEYRLHETSK EPDVSLGSTWLSDFPQAWAET GGMGLAVRQAPLIIPLKATSTP VSIKQYPMSQEARLGIKPHIQR LLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHPTS QPLFAFEWRDPEMGISGQLTW TRLPQGFKNSPTLFNEALHRD LADFRIQHPDLILLQYVDDLLL AATSELDCQQGTRALLQTLGN LGYRASAKKAQICQKQVKYL GYLLKEGQRWLTEARKETVM GQPTPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTKPGT LFNWGPDQQKAYQEIKQALLT APALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVA AIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNA RMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDA DHTWYTDGSSLLQEGQRKAG AAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLN VYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGHS AEARGNRMADQAARKAAITE TPDTSTLLIENSSPKRPAATKK AGQAKKKK (SEQ ID NO: 86)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	acggatggtagcagccattgccgtactgacaaaggatgcaggc aagctaaccatgggacagccactagtcattctggccccccatgc agtagaggcactagtcaaacaacccccgaccgctggctttcca acgcccggatgactcactatcaggccttgcttttggacacggacc gggtccagttcggaccggtggtagccctgaacccggctacgct gctcccactgcctgaggaagggctgcaacacaactgccttgata tcctggccgaagcccacggaacccgacccgacctaacggacc agccgctcccagacgccgaccacacctggtacacggatggaa gcagtctcttacaagagggacagcgtaaggcgggagctgcggt gaccaccgagaccgaggtaatctgggctaaagccctgccagcc gggacatccgctcagcgggctgaactgatagcactcacccagg ccctaaagatggcagaaggtaagaagctaaatgtttatactgata gccgttatgcttttgctactgcccatatccatggagaaatatacaga aggcgtgggtggctcacatcagaaggcaaagagatcaaaaata aagacgagatcttggccctactaaaagccctctttctgcccaaaa gacttagcataatccattgtccaggacatcaaaagggacacagc gccgaggctagaggcaaccggatggctgaccaagcggcccga aaggcagccatcacagagactccagacacctctaccctcctcat agaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 85)	
Truncated (16-775) reference CRISPR nickase (H397A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaaag aagaagcggaaagtccagaaagagctggtgctgggcattgact acggcggaaagtacacaggcctggccgtggtggaccggagaa acaaccaggtgctgtacgccagaaccgtcaagatgagagacga cgtggccgatatcctcgaaggccggagggaacagcgaggcatt agaagaacaggccagacaaagaagaagagactgagagaactg aaaacctacctgaagaacatcggaggaaaatatgacgagtctat cggcacattcaaaaacgagcctttcatgacagtgtacaagctgtg tcacaagagaggctacgactacgtggaaatgcctaaggaaaag accagagaagaaatcgaggccatgtctgccgacgagaaaaag cagtgggaaaaagagaaaaaggagtgggaggaaaccgagcg gaactccagacaccgggatgaggtgaccaaggacgtgcggaa gatcatgagcgagagagatgccaccgagcagcagatcagcca cgtgatcaagatcttcaacaagcagtaccggcccaagcggttca ataatagaatcctgacaaaatgtaaggttgaggattgcggcgtga acacacctctgagaaagaacgtcagagacctgctgatcaagga catcgtgagattcctgcccctgacccagagcgagaaagccgag ctggaagctcatgtgctggataaggaaggaagagagaaggtga agaccttcttcagacggcacagaatcaacgagttcctgcgcaag cagatctacgacatcgcctttaaccaaaagctgagcggcagaac cgtgttctgcaagaagcacatcctggagcacgagcaacacacc aaagaggaacgcaaggtctttagactggcccctagcctgaaaac taagattaacaacgtgctggctgtggttaaggacgagatcctgcc ccagttcggcatcaacaaggtcatcatggaatccaacaacttcga tattgccgctagaaccaagggcaagaagcggctggccaaggat gagtacggcaagggccacaaggaggccaaagagagcctgatc gaggtgctgctgcgggaaaccaagaaccagtgcgtgtactgcg gaaaggtgatcaccgtgaccaccgccaataccgac**GCC**atct tccccagaaaggcgggcggcatcaacgttttcgccaacctggtg gcctgttgtagagagtgcaacgagaagaagaagggcagaaca cctctggaatccggcgtgcttccacagcctgagatcaccgccca cctggaaaacgacctcaaaaagaagatccttgaggacgccaga aatatgaacgagctgaatctgaacagcatcaagtacatgagctac	MKRTADGSEFESPKKKRKVQ KELVLGIDYGGKYTGLAVVD RRNNQVLYARTVKMRDDVA DILEGRREQRGIRRTGQTKKK RLRELKTYLKNIGGKYDESIGT FKNEPFMTVYKLCHKRGYDY VEMPKEKTREEIEAMSADEKK QWEKEKKEWEETERNSRHRD EVTKDVRKIMSERDATEQQIS HVIKIFNKQYRPKRFNNRILTK CKVEDCGVNTPLRKNVRDLLI KDIVRFLPLTQSEKAELEAHVL DKEGREKVKTFFRRHRINEFL RKQIYDIAFNQKLSGRTVFCK KHILEHEQHTKEERKVFRLAPS LKTKINNVLAVVKDEILPQFGI NKVIMESNNFDIAARTKGKKR LAKDEYGKGHKEAKESLIEVL LRETKNQCVYCGKVITVTTAN TD**A**IFPRKAGGINVFANLVAC CRECNEKKKGRTPLESGVLPQ PEITAHLENDLKKKILEDARN MNELNLNSIKYMSYASIGWRH MRDRLRGLIGNEELPIERQSGI LTAYFRRWWGFRKERGNDIH HALDAVILASRKDYTDDGVD MTLKPKSSNGREFDPEKHLPD TISFKRDKSNKGTPLYDKNPLS IKNSLITRRWMVTEIEHGKEDT IISDEYRKKLKEAFKHWGTSK GKCLTDEEARQTGFYQRKNG NFMSLRCVEKGLGWGQMVKI NNNAFKTNVHNIGVAVYLDE

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	gccagtatcggatggcggcacatgagagatagactgaggggcc ttatcggtaatgaagaactgcctatcgagcgccaaagcggcatc ctgacagcttacttcagacggtggtggggatttcggaaggaacg gggaaatgacattcaccacgccctggacgctgtgatcctggcca gcagaaaagactatacagatgacggcgtggatatgaccctgaa gcccaaaagcagcaacggccgagaattcgaccccgagaaaca tctgcctgatacaatcagcttcaaaagagataagtccaacaaagg aaccctctgtatgataagaacccattgtctatcaagaattctctga tcacaagaaggtggatggtgacagagatcgaacacggcaagg aggacaccatcatcagcgacgagtacagaaagaagctgaagg aggctttcaaacactggggcaccagcaagggcaagtgcctgac cgacgaggaagccagacagaccggcttctaccagagaaagaa tggcaacttcatgagcctgcggtgcgtggagaagggcctgggct ggggccagatggtgaagatcaataacaacgcctttaagacgaac gtgcacaacatcggggtggctgtttacctggacgagaagggca agaagaaggcctacgaaattaagaaccccagactgctgaaacat ttcgtggaaaagccacctgagatcaagggaaagaagctgtacac cctgcggagaggcgactacgtgacccacgagagcgaggactc cgtgtacagaatcaaaaaactgggcacaagccctgtgctggaag ctgtggtcaagggctctatcggccgcatccggaacagcgccag cgccacaaaactgaccaaggtgtccaagcctcagaaggtgttttc tggaggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagcag cggcggcagcagcaccctaaatatagaagatgagtatcggctac atgagacctcaaaagagccagatgtttctctagggtccacatggc tgtctgattttcctcaggcctgggcggaaaccgggggcatggga ctggcagttcgccaagctcctctgatcatacctctgaaagcaacct ctacccccgtgtccataaaacaatacccatgtcacaagaagcc agactggggatcaagccccacatacagagactgttggaccagg gaatactggtaccctgccagtccccctggaacacgcccctgcta cccgttaagaaaccagggactaatgattataggcctgtccaggat ctgagagaagtcaacaagcgggtggaagacatccacccacc gtgcccaaccctacaacctcttgagcgggctcccaccgtccca ccagtggtacactgtgcttgatttaaaggatgccttttctgcctga gactccacccccaccagtcagcctctcttcgcctttgagtggagag atccagagatgggaatctcaggacaattgacctggaccagactc ccacagggtttcaaaaacagtcccaccctgtttaatgaggcactg cacagagacctagcagacttccggatccagcacccagacttgat cctgctacagtacgtggatgacttactgctggccgccacttctga gctagactgccaacaaggtactcgggccctgttacaaaccctag ggaacctcgggtatcgggcctcggccaagaaagcccaaatttg ccagaaacaggtcaagtatctggggtatcttctaaaagagggtca gagatggctgactgaggccagaaaagagactgtgatggggca gcctactccgaagaccctcgacaactaagggagttcctaggga aggcaggcttctgtcgcctcttcatccctgggtttgcagaaatggc agccccctgtaccctctcaccaaaccggggactctgtttaattg gggcccagaccaacaaaaggcctatcaagaaatcaagcaagct cttctaactgccccagccctggggttgccagatttgactaagccct ttgaactctttgtcgacgagaagcagggctacgccaaaggtgtcc taacgcaaaaactgggaccttggcgtcggccggtggcctacctg tccaaaaagctagacccagtagcagctgggtggccccccttgcct acggatggtagcagccattgccgtactgacaaaggatgcaggc aagctaaccatgggacagccactagtcattctggcccccatgc	KGKKKAYEIKNPRLLKHFVEK PPEIKGKKLYTLRRGDYVTHE SEDSVYRIKKLGTSPVLEAVV KGSIGRIRNSASATKLTKVSKP QKVFSGGSSGGSSGSETPGTSE SATPESSGGSSGGSSTLNIEDE YRLHETSKEPDVSLGSTWLSD FPQAWAETGGMGLAVRQAPL IIPLKATSTPVSIKQYPMSQEAR LGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQ DLREVNKRVEDIHPTVPNPYN LLSGLPPSHQWYTVLDLKDAF FCLRLHPTSQPLFAFEWRDPE MGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCQQGT RALLQTLGNLGYRASAKKAQI CQKQVKYLGYLLKEGQRWLT EARKETVMGQPTPKTPRQLRE FLGKAGFCRLFIPGFAEMAAPL YPLTKPGTLFNWGPDQQKAY QEIKQALLTAPALGLPDLTKPF ELFVDEKQGYAKGVLTQKLG PWRRPVAYLSKKLDPVAAGW PPCLRMVAAIAVLTKDAGKLT MGQPLVILAPHAVEALVKQPP DRWLSNARMTHYQALLLDTD RVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQE GQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATAHI HGEIYRRRGWLTSEGKEIKNK DEILALLKALFLPKRLSIIHCPG HQKGHSAEARGNRMADQAA RKAAITETPDTSTLLIENSSPKR PAATKKAGQAKKKK (SEQ ID NO: 88)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
	agtagaggcactagtcaaacaacccccgaccgctggctttcca acgcccggatgactcactatcaggccttgcttttggacacggacc gggtccagttcggaccggtggtagccctgaacccggctacgct gctcccactgcctgaggaagggctgcaacacaactgccttgata tcctggccgaagcccacggaacccgacccgacctaacggacc agccgctcccagacgccgaccacacctggtacacggatggaa gcagtctcttacaagagggacagcgtaaggcgggagctgcggt gaccaccgagaccgaggtaatctgggctaaagccctgccagcc gggacatccgctcagcgggctgaactgatagcactcacccagg ccctaaagatggcagaaggtaagaagctaaatgtttatactgata gccgttatgcttttgctactgcccatatccatggagaaatatacaga aggcgtgggtggctcacatcagaaggcaaagagatcaaaaata aagacgagatcttggccctactaaaagccctctttctgcccaaaa gacttagcataatccattgtccaggacatcaaaagggacacagc gccgaggctagaggcaaccggatggctgaccaagcggcccga aaggcagccatcacagagactccagacacctctaccctcctcat agaaaattcatcacccaaaaggccggcggccacgaaaaaggc cggccaggcaaaaaagaaaaag (SEQ ID NO: 87)	

The CRISPR nuclease polypeptides described in **Table 14** are tested as described in **Example 9** to introduce indels into human target genes. The fusion polypeptides described in **Tables 21 and 22** are to be tested as described in **Example 10** to install edits in human target genes.

Example 9: Editing of Human Genes in HEK293T Cells Using CRISPR Nuclease Polypeptides

This Example shows genetic modification of human genes utilizing the CRISPR nuclease polypeptides of **Table 1 Table 14** constructed in **Examples 1**, **3**, and **8**. Specifically, the polypeptides are used to introduce indels into the human target genes when combined with RNA guides.

RNA guides are either those shown in **Table 2** of **Example 1**, **Table 8** of **Example 4**, or shown in **Table 23** below. The RNA guide sequences shown in **Table 23** are cloned into plasmids as described in **Examples 1** above. The RNA guide sequences shown in **Table 23** below exclude a designed +1 G nucleotide that is added by the U6 PolIII promoter at the start of the transcript (*i.e.*, the 5' end of the RNA) for more efficient transcription. Editing template RNAs are combined with the nuclease polypeptides, transfected into HEK293T cells, and the genomic DNA is recovered approximately 72 hours post transfection as described in **Example 1**. Samples are prepared for NGS and analyzed as described in **Example 1**.

Table 23. Full-length Editing Template RNA Sequences

RNA guide	Target	Scaffold ID	Sequence
	EMX1-T3	Reference	UUUAUGUGAUUGAUGCCCAAguuacaguuaaggcucuuuggaaacaaagaagccuuaauuguaaaacgccuauaugguaaagugauguacguuuggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauuuaaaaaacuuucugcggcuugcuuacuuuaguaagcuuuguggcugaggcagaauuccuu (SEQ ID NO: 91)
	EMX1-T3	Truncated Reference	UUUAUGUGAUUGAUGCCCAAguuacaguuaagggaaaccuuaauuguaaaacgccuauaugguaaagugauguacguuuggguauauaucgccagccugaaccucuacgccagaaauggcagcuuuaucaugggguuaggacgauauu (SEQ ID NO: 92)
	AAVS1-T3	Scaffold 2	UUGUGAGAAUGGUGCGUCCUguuacaguuaagggaaaccuuaauuguaaaacgccuauacggugaagggauauacgcuuggguuuguccagccugagccucuaugccagaaauggcgccuucaucgugggguuaggaca (SEQ ID NO: 93)
	EMX1-T3	Scaffold 2	UUUAUGUGAUUGAUGCCCAAguuacaguuaagggaaaccuuaauuguaaaacgccuauacggugaagggauauacgcuuggguuuguccagccugagccucuaugccagaaauggcgccuucaucgugggguuaggaca (SEQ ID NO: 94)
	EMX1-T7	Scaffold 2	CUUUAGCAAGGGACUAUUCAguuacaguuaagggaaaccuuaauuguaaaacgccuauacggugaagggauauacgcuuggguuuguccagccugagccucuaugccagaaauggcgccuucaucgugggguuaggaca (SEQ ID NO: 95)

** The spacer sequence is shown in uppercase and the scaffold sequence is show in lowercase*

It is anticipated that the CRISPR nuclease polypeptides described in this Example are capable of introducing indels at target sites encoded by RNA guides, such as those described in this Example, into human genes.

Example 10: RNA-Templated Editing of Human Genes in HEK293T Cells Using CRISPR Nuclease-Reverse Transcriptase and CRISPR Nickase-Reverse Transcriptase Fusion Polypeptides

This Example shows genetic modification of human genes utilizing the CRISPR nuclease-reverse transcriptase polypeptides of **Tables 16** and **21** and the CRISPR nickase-reverse transcriptase polypeptides of **Tables 16** and **22** constructed in **Examples 6 and 8**. Specifically, the fusion polypeptides are used to install sequence substitutions of 6 nucleotides into the human target genes when combined with editing template RNAs.

Editing template RNAs are either those shown in **Table 18** of **Example 7** or shown in **Table 24** below. The editing template RNA sequences shown in **Table 24** are cloned into plasmids as described in **Example 7** above. The editing template RNA sequences shown in **Table 24** below exclude a designed +1 G nucleotide that is added by the U6 PolIII promoter at the start of the transcript (*i.e.*, the 5' end of the RNA) for more efficient transcription. Editing template RNAs are combined with the fusion polypeptides, transfected into HEK293T cells, and the genomic DNA is recovered approximately 72 hours post transfection as described in **Example 7**. Samples are prepared for NGS and analyzed as described in **Example 1**.

Table 24. Full-length Editing Template RNA Sequences

Editing template RNA	Target	Scaffold ID	Sequence
	EMX1-T3	Truncated Reference	UUUAUGUGAUUGAUGCCCAA**GUUACAGUUAAGGGAA ACCUUAAUUGUAAAACGCCUAUAUGGUAAAGUGAUG UACGUUUGGGUAUAUAUCGCCAGCCUGAACCUCUAC GCCAGAAAUGGCAGCUUUAUCAUGGGUUAGGACGA** UAUUCUAUCAUCUGAggaaacGGCAUCAAUCACAUUCUC UCUCUUGACGCGGUUCUAUCUAGUUACGCGUUAAACC AACUAGAAA (SEQ ID NO: 96)
	EMX1-T3	Scaffold 2	UUUAUGUGAUUGAUGCCCAA**GUUACAGUUAAGGGAA ACCUUAAUUGUAAAACGCCUAUACGGUGAAGGGAUA UACGCUUGGGUUUGUCCAGCCUGAGCCUCUAUGCCA GAAAUGGCGCCUUCAUCGUGGGUUAGGACA**CUAUCA UCUGAggaaacGGCAUCAAUCACAUUCUCUCUCUUGACG CGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAAA (SEQ ID NO: 97)

* *The Scaffold sequence is shown in bold and the 6-nucleotide substitution in lowercase*

It is anticipated that the CRISPR nuclease- and CRISPR nickase-reverse transcriptase fusion polypeptides described in this Example are capable of incorporating substitutions encoded by editing template RNAs, such as those described in this Example, into human genes.

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.

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.

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.

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.”

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 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.

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 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.

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.

5

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 SEQ ID NO: 160, or is a variant of SEQ ID NO: 1, the variant comprising an amino acid sequence at least 90% identical to SEQ ID NO: 1;

(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 (i) a scaffold sequence recognizable by the CRISPR nuclease polypeptide and (ii) a spacer sequence specific to a target sequence within a genomic site of interest; the target sequence being adjacent to a protospacer adjacent motif (PAM) specific to the CRISPR nuclease polypeptide; 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 comprises one or more nuclear localization signal (NLS) upstream or downstream to the CRISPR nuclease polypeptide, the RT polypeptide, or both; a peptide linker between the CRISPR nuclease polypeptide and the RT polypeptide, or a combination thereof.

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 the peptide linker between the CRISPR nuclease polypeptide and the RT polypeptide.

4. The gene editing system of claim 2, wherein the fusion polypeptide comprises the peptide linker located between the CRISPR nuclease polypeptide and the RT polypeptide, and wherein the fusion polypeptide comprises a first NLS located at the N-terminus of the fusion polypeptide and a second NLS located at the C-terminus of the fusion polypeptide.

5. The gene editing system of any one of claims 1-4, wherein the CRISPR nuclease polypeptide is the variant of SEQ ID NO: 1.

6. The gene editing system of claim 5, wherein the variant of SEQ ID NO: 1 comprises, relatively to SEQ ID NO: 1:

(i) one or more mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1;

(ii) one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions;

(iii) an N-terminal truncation, or

(iv) a combination of (i), (ii) and/or (iii).

7. The gene editing system of claim 6, wherein the variant comprises (i), and wherein the one or more nickase mutations are at positions H397, D396, N420, D24, E337, and/or D524 of SEQ ID NO: 1, optionally at positions D396, H397, and/or N420 relative to SEQ ID NO: 1.

8. The gene editing system of claim 7, wherein:

(a) the mutation at D396 is an amino acid substitution of D396A, D396G, D396L, or D396S;

(b) the mutation at H397 is an amino acid substitution of H397A, H397G, H397L, or H397S; and

(c) the mutation at N420 is an amino acid substitution of N420A, N420G, N420L, or N420S,

optionally wherein the mutation at H397 is amino acid substitution of H397A or H397L.

9. The gene editing system of claim 8, wherein the variant comprises a nickase mutation at position H397 of SEQ ID NO: 1, optionally wherein the nickase mutation is H397A.

10. The gene editing system of any one of claims 5-9, wherein 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 the variant comprises the one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions, which 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.

11. The gene editing system of claim 10, wherein the one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions, are located at one or more of positions D56, E59, G60, E63, I67, G71, S564, D568, W593, T605, E655, E694, and E706 in SEQ ID NO: 1.

12. The gene editing system of any one of claims 5-11, wherein the CRISPR nuclease polypeptide contains up to 20 arginine and/or lysine substitutions, optionally up to 20 arginine substitutions, relative to SEQ ID NO: 1; optionally wherein the CRISPR nuclease polypeptide contains up to 15 arginine and/or lysine substitutions, optionally up to 15 arginine substitutions, relative to SEQ ID NO: 1.

13. The gene editing system of claim 11, wherein the one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions, are located at the following positions relative to SEQ ID NO: 1:

a) I67, D568, and E706;

b) I67, D56, and D568;

c) I67, T605, and D568;

d) I67, D56, and E706;

e) I67, T605, and E706;

f) I67, D568, and E694;

g) I67, D56, E694, and E706;

h) I67, D56, T605, and E706; or

i) I67, D568, W593, and E706.

14. The gene editing system of claim 13, wherein the CRISPR nuclease polypeptide comprises the following arginine substitutions relative to the reference CRISPR nuclease:

a) I67R, D568R, and E706R;

b) I67R, D56R, and D568R;

c) I67R, T605R, and D568R;

d) I67R, D56R, and E706R;

e) I67R, T605R, and E706R;

f) I67R, D568R, and E694R;

g) I67R, D56R, E694R, and E706R;

h) I67R, D56R, T605R, and E706R; or

i) I67R, D568R, W593R, and E706R;

optionally wherein the CRISPR nuclease polypeptide comprises the arginine substitutions of a) or i).

15. The gene editing system of any one of claims 6-14, wherein the CRISPR nuclease polypeptide comprises the N-terminal truncation, which is a deletion within residues 1-15 of SEQ ID NO: 1; optionally wherein the N-terminal truncation is the deletion of residues 1-14 or 1-15 of SEQ ID NO: 1.

16. The gene editing system of any one of claims 5-15, wherein the CRISPR nuclease polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 1.

17. The gene editing system of claim 16, wherein the CRISPR nuclease polypeptide comprises an amino acid sequence at least 98% identical to SEQ ID NO: 1.

18. The gene editing system of claim 5, wherein the variant comprises (a) the one or more nickase mutations in the HNH nuclease domain, which optionally are at positions D396, H397, and/or N420 relative to SEQ ID NO: 1; optionally wherein the nickase mutation is at position H397; and (b) the one or more arginine and/or lysine substitutions, which optionally are at positions I67, D568, W593, E706, or a combination thereof.

19. The gene editing system of claim 6, wherein the variant comprises:

(a) the one or more nickase mutations in the HNH nuclease domain, which optionally are at positions D396, H397, and/or N420 relative to SEQ ID NO: 1;

(b) the one or more arginine and/or lysine substitutions, which optionally are at positions I67, D568, W593, and/or E706; and

(c) the N-terminal truncation within residues 1-15 of SEQ ID NO: 1, optionally wherein the N-terminal truncation is the deletion of residues 1-14 or 1-15 of SEQ ID NO: 1.

20. The gene editing system of claim 19, wherein the variant comprises:

(a) the nickase mutation of H397A;

(b) the arginine substitutions of I67R, D568R, and E706R; or the arginine substitutions of I67R, D568R, W593R, and E706R; and

(c) the N-terminal truncation comprising deletion of residues 1-15 of SEQ ID NO: 1.

21. The gene editing system of claim 5, wherein the CRISPR nuclease variant comprises the mutations in the CRISPR nuclease polypeptides listed in **Table 1**, **Table 11**, or **Table 14**.

22. The gene editing system of any one of claims 1-21, wherein the CRISPR nuclease polypeptide in the fusion polypeptide does not include its native N-terminus methionine residue.

23. The gene editing system of claim 1, wherein the CRISPR nuclease polypeptide is provided in **Table 1**, **Table 11**, or **Table 14**.

24. The gene editing system of any one of claims 1-23, wherein the RT polypeptide is Moloney Murine Leukemia Virus (MMLV)-RT, optionally wherein the MMLV-RT comprises the amino acid sequence of SEQ ID NO: 66 or SEQ ID NO: 176.

25. The gene editing system of claim 1, wherein the fusion polypeptide is set forth in **Tables 16**, **21**, and **22**.

26. The gene editing system of any one of claims 1-25, wherein the system comprises the fusion polypeptide.

27. The gene editing system of any one of claims 1-26, wherein the system comprises the first nucleic acid encoding the fusion polypeptide.

28. The gene editing system of claim 27, wherein the first nucleic acid is located on a vector, which optionally is a viral vector.

29. The gene editing system of claim 27, wherein the first nucleic acid is a messenger RNA (mRNA).

30. The gene editing system of any one of claims 1-29, wherein the spacer sequence in the gRNA is 15-30-nucleotide in length; optionally 15-20-nucleotide in length.

31. The gene editing system of any one of claims 1-30, wherein the PAM is 5'-NGG-3', in which N represents any nucleotide.

32. The gene editing system of any one of claims 1-31, wherein the scaffold sequence comprises a nucleotide sequence at least 85% identical to SEQ ID NO: 2.

33. The gene editing system of claim 32, wherein the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2.

34. The gene editing system of claim 32, wherein the scaffold is a truncated variant of SEQ ID NO: 2, which has a 3' truncation relative to SEQ ID NO: 2, and wherein the truncated variant is about 110-140-nt in length; optionally wherein the 3' truncation comprises deletions within residues 143-202 of SEQ ID NO: 2.

35. The gene editing system of claim 34, wherein the truncated variant further comprises one or more deletions within residues 10-40 of SEQ ID NO: 2; optionally wherein the deletions comprise residues 14-20 and/or 25-32 of SEQ ID NO: 2.

36. The gene editing system of claim 34 or claim 35, wherein the truncated variant further comprises one or more mutations relative to SEQ ID NO: 2; optionally wherein the one or more mutations comprise a deletion or substitution within residues 81-85 of SEQ ID NO: 2.

37. The gene editing system of claim 34, wherein the scaffold comprises the nucleotide sequence of SEQ ID NO: 27, or SEQ ID NO: 28; and wherein the scaffold is about 115-130 -nt in length.

38. The gene editing system of claim 37, wherein

(a) the CRISPR nuclease polypeptide comprises the amino acid sequence of SEQ ID

NO: 1; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 27;

(b) the CRISPR nuclease polypeptide is a variant of SEQ ID NO: 1 comprising mutations at positions I67, D568, and E706, which optionally are arginine substitutions I67R, D568R, and E706R; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 28; or

(c) the CRISPR nuclease polypeptide is a variant of SEQ ID NO: 1 comprising a deletion within residues 1-15 of SEQ ID NO: 1; optionally a deletion of residues 1-14 or 1-15 of SEQ ID NO: 1; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 28.

39. The gene editing system of any one of claims 1-38, wherein the PBS in the RT donor RNA is 5-50-nucleotide in length; optionally 5-20-nucleotide in length.

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

41. The gene editing system of claim 40, 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.

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

43. The gene editing system of any one of claims 1-42, 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.

44. The gene editing system of claim 43, 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.

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

46. The gene editing system of any one of claims 1-45, wherein the RNA molecule 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.

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

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

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

(iii) the spacer sequence, the scaffold sequence, the template sequence, the PBS, a linker, and the 3' extension.

48. The gene editing system of any one of claims 1-47, wherein the system comprises the RNA molecule.

49. The gene editing system of any one of claims 1-47, wherein the system comprises the second nucleic acid encoding the RNA molecule.

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

51. The gene editing system of any one of claims 1-50, wherein the system comprises one or more lipid nanoparticles (LNPs) associated with the fusion polypeptide or the encoding nucleic acid, and/or the RNA molecule and/or the encoding nucleic acid.

52. 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 the fusion polypeptide or the encoding nucleic acid, and the RNA molecule and/or the encoding nucleic acid.

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

54. A kit comprising (a) the fusion polypeptide or the encoding nucleic acid, and (b) the RNA molecule and/or the encoding nucleic acid of the gene editing system set forth in any one of claims 1-52.

55. A gene editing method, comprising delivering the gene editing system of any one of claims 1-52 to a host cell to edit a genomic site targeted by the gRNA of the gene editing system.

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

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

58. A fusion polypeptide, comprising a CRISPR nuclease polypeptide set forth in any one of claims 1-21 and a reverse transcriptase polypeptide set forth in claim 1 or claim 22.

59. The fusion polypeptide of claim 58, which comprises the amino acid sequence of SEQ ID NO: 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, or 90.

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

61. The nucleic acid of claim 60, which comprises the nucleotide sequence of SEQ ID NO: 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, or 89.

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